PART II
RESEARCH PAPERS

5. MODELING THE BIOLOGICAL PROTOTYPES

SOME CONSTRAINTS ON INTELLIGENT SYSTEMS

Autonomous Computation in a Changing World

Stephen Grossberg

Department of Cognitive and Neural Systems, Boston University, 677 Beacon Street, Boston, MA 02215

steve@bu.edu, http://www.cns.bu.edu/Profiles/Grossberg

Among the many possible topics concerning how autonomous intelligent systems should be designed, I will focus on one that is close to work with which I am familiar. A core problem on which much more work needs to be done is how to design systems that can autonomously learn, recognize, and perform complex tasks in a rapidly changing environment. Such self-organizing systems should also be able to interact effectively with humans and other self-organizing systems in order to achieve goals cooperatively.

In order to make the interface between human and system as seamless as possible, biological designs, notably designed inspired by and even emulating brain architectures, will be helpful. The list of possible applications is incredibly long, ranging from autonomous search and data base management tools on the world wide web, medical data base prediction to help doctors and other health professionals, classifiers of complex imagery of multiple types, new approaches to speech perception in noisy multi-speaker environments, and controllers of autonomous mobile robots, to models of normal brain and behavior, and predictions about how different brain lesions can generate the behavioral symptoms of mental disorders.

Available results have already suggested that the brain designs for sensory and cognitive processes differ from, and are even computationally complementary to, the designs for spatial navigation and action. This complementarity can be noticed by observing that cognitive knowledge needs to accumulate in a stable way over a period of years, with new knowledge not accidentally erasing previously learned, but still useful, knowledge. This is the familiar problem of "catastrophic forgetting". In contrast, the parameters that control action need to be continually updated in order to adapt to changes, including damage, to motor effectors. Here, catastrophic forgetting is a useful property. Thus, these systems will need to incorporate new ideas about parallel processing between information subsystems that compute complementary properties.

The design of increasingly autonomous intelligent agents will also require an end-to-end approach, in which all the aspects of perception, cognition, emotion, and action are realized in a single system. Feedback cycles of information processing need to be designed from perception through action and then back to perception again, mediated by feedback through the environment. Such cycles of information processing can evaluate the effects of system performance on the environment, and modify the system where needed to achieve better environmental control. It has also become clear that, in addition to these externally mediated cycles of information processing with the environment, internally mediated feedback is needed to achieve autonomous system properties. Such internal feedback realizes properties of intentionality and attention that are characteristic of biological intelligence. The design of self-organizing feedback systems will require a deeper analysis of nonlinear systems, since various types of nonlinearity are needed to achieve key system properties that depend on feedback, such as the stability of fast learning in a changing environment.

One example of such an autonomous system is the primate cerebral cortex. All sensory and cognitive neocortex is organized into laminar circuits, wherein bottom-up, top-down, and horizontal connections are synthesized into a unified design. Recent modeling (Grossberg, 1999; Grossberg, Mingolla, and Ross, 1997; Grossberg and Raizada, 2000; Grossberg and Williamson, 2000) has clarified how these laminar circuits are designed (Figure 1) to simultaneously achieve at least three properties: (1) stable development and learning of circuit connections and adaptive weights in response to a changing world, thereby providing a solution of the *stability-*

plasticity dilemma; (2) a seamless fusion of bottom-up data-driven processing and top-down intentional processing whereby high-level constraints can selectively focus attention upon important information; and (3) the coherent grouping or binding of spatially distributed information into representations of objects and events, while suppressing noise and weaker groupings, without a loss of analog sensitivity to input values, the so-called property of *analog coherence*.

The design of more subtle decision making processes in an autonomous agent will require more sophisticated cognitive-emotional interactions, whereby the information acquired through cognitive processing is evaluated and selected in terms of internal system values and goals. Such interactions help to direct attention selectively to those subsets of information that predict future success in achieving system goals. Designs for such systems need to be able to use unsupervised learning when no evaluative feedback is available, but to be able to switch to supervised learning whenever such feedback is available. In a self-organizing autonomous learning system, both unsupervised and supervised learning need to be able to operate without a change of system design. Taken together, these constraints point to the development of new types of self-organizing parallel processing systems wherein nonlinear feedback within the system and between system and world help the system to rapidly adapt to a changing world, and thereby to better represent, predict and control it.

Illustrative References

Brown, J., Bullock, D. and 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.

Fiala, J., Grossberg, S. and Bullock, D. (1996). Metabotropic glutamate receptor activation in cerebellar Purkinje cells as substrate for adaptive timing of the classically conditioned eye-blink response. *Journal of Neuroscience*, **16**, 3760-3774.

Grossberg, S. (1999). How does the cerebral cortex work? Learning, attention, and grouping within the laminar circuits of visual cortex. *Spatial Vision*, **12**, 163-187.

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

Grossberg, S. (2000). The complementary brain: Unifying brain dynamics and modularity. *Trends in Cognitive Sciences*, **4**, 233-246.

Grossberg, S. (2000). The imbalanced brain: From normal behavior to schizophrenia. *Biological Psychiatry*, in press.

Grossberg, S., Boardman, I. and Cohen, M.A. (1997). Neural dynamics of variable-rate speech categorization. Journal of Experimental Psychology: *Human Perception and Performance*, **23**, 481-503.

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. and Ross, W. (1997). Visual brain and visual perception: how does the cortex do perceptual grouping? *Trends in Neurosciences*, **20**, 106-111.

Grossberg, S. and Myers, C.W. (2000). The resonant dynamics of speech perception: Interword integration and duration-dependent backward effects, *Psychological Review*, in press.

Grossberg, S. and Raizada, R.D.S. (2000). Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex. *Vision Research*, **40**, 1413-1432.

Grossberg, S. and Williamson, J.R. (2000). A neural model of how horizontal and interlaiminar connections of visual cortex develop into adult circuits that carry out perceptual grouping and learning. *Cerebral Cortex*, in press.

Figure Caption

Figure 1. Some model cell interactions between the lateral geniculate nucleus (LGN) and cortical areas V1 and V2 for perceptual grouping and attention: Excitatory connections are shown with open symbols. Inhibitory interneurons are shown filled-in black. (a): The LGN provides bottom-up activation to layer 4 via two routes. Firstly, it makes a strong connection directly into layer 4. Secondly, LGN axons send collaterals into layer 6, and thereby also activate layer 4 via the 6 → 4 on-center off-surround path. Thus, the combined effect of the bottom-up LGN pathways is to stimulate layer 4 via an on-center off-surround, which provides divisive contrast normalization of layer 4 cell responses. (b): *Folded feedback* carries attentional signals from higher cortex into layer 4 of V1, via the modulatory 6 → 4 path. Corticocortical feedback axons tend preferentially to originate in layer 6 of the higher area and to terminate in the lower cortex's layer 1, where they can excite the apical dendrites of layer 5 pyramidal cells whose axons send collaterals into layer 6. Several other routes through which feedback can pass into V1 layer 6 exist. Having arrived in layer 6, the feedback is then "folded" back up into the feedforward stream by passing through the 6 → 4 on-center off-surround path. (c): Connecting the 6 → 4 on-center off-surround to the layer 2/3 grouping circuit: like-oriented layer 4 simple cells with opposite contrast polarities compete (not shown) before generating half-wave rectified outputs that converge onto layer 2/3 complex cells in the column above them. Like attentional signals from higher cortex, groupings which form within layer 2/3 also send activation into the *folded feedback* path, to enhance their own positions in layer 4 beneath them via the 6 → 4 on-center, and to suppress input to other groupings via the 6 → 4 off-surround. There exist direct layer 2/3 → 6 connections in macaque V1, as well as indirect routes via layer 5. (d): Top-down corticogeniculate feedback from V1 layer 6 to LGN also has an on-center off-surround anatomy, similar to the 6 → 4 path. The on-center feedback selectively enhances LGN cells that are consistent with the activation that they cause, and the off-surround contributes to length-sensitive (endstopped) responses that facilitate grouping perpendicular to line ends. (e): The entire V1/V2 circuit: V2 repeats the laminar pattern of V1 circuitry, but at a larger spatial scale. In particular, the horizontal layer 2/3 connections have a longer range in V2, allowing above-threshold perceptual groupings between more widely spaced inducing stimuli to form. V1 layer 2/3 projects up to V2 layers 6 and 4, just as LGN projects to layers 6 an 4 of V1. Higher cortical areas send feedback into V2 which ultimately reaches layer 6, just as V2 feedback acts on layer 6 of V1. Feedback paths from higher cortical areas straight into V1 (not shown) can complement and enhance feedback from V2 into V1.

The Neurodynamics of Intentionality in Animal Brains May Provide a Basis for Constructing Devices that are Capable of Intelligent Behavior

Walter J Freeman

Department of Molecular and Cell Biology
University of California, Berkeley CA 94720-3200 USA

ABSTRACT

Intelligent behavior is characterized by flexible and creative pursuit of endogenously defined goals. It has emerged in humans through the stages of evolution that are manifested in the brains and behaviors of other animals. Intentionality is a key concept by which to link brain dynamics to goal-directed behavior. The archetypal form of intentional behavior is an act of observation through time and space, by which information is sought for the guidance of future action. Sequences of such acts constitute the key desired property of free-roving, semi-autonomous devices capable of exploring remote environments that are inhospitable for humans. Intentionality consists of the neurodynamics by which images are created of future states as goals, of command sequences by which to act in pursuit of goals, of predicted changes in sensory input resulting from intended actions (reafference) by which to evaluate performance, and modification of the device by itself for learning from the consequences of its intended actions. These principles are well known among psychologists and philosophers. What is new is the development of nonlinear mesoscopic brain dynamics, by which using chaos theory to understand and simulate the construction of meaningful patterns of neural activity that implement the perceptual process of observation. The prototypic hardware realization of intelligent behavior is already apparent in certain classes of robots. The chaotic neurodynamics of sensory cortices in pattern recognition is ready for hardware embodiments, which are needed to provide the eyes, noses and ears of devices for survival and autonomous operation in complex and unpredictable environments.

Key Words: *Chaos theory, Intentionality, Mesoscopic Brain dynamics, Perception, Reafference*

1.0 Neurodynamics of intentionality in the behavioral act of observation

1.1 The properties of intentionality

The first step in pursuit of an understanding of intentionality is to ask, what happens in brains during an act of observation? This is not a passive receipt of information from the world. It is a purposive action by which an observer directs the sense organs toward a selected aspect of the world and interprets the resulting barrage of sensory stimuli. The concept of intentionality has been used to describe this process in different contexts, since its first use by Aquinas in 1272 [1]. The three salient characteristics of intentionality as it was developed by him are (a) intent or directedness toward some future state or goal, (b) wholeness, and (c) unity [12]. These three aspects correspond to current use of the term in psychology [with the meaning of purpose], in medicine [with the meaning of mode of healing and integration of the body], and in analytic philosophy [with the meaning of the way in which beliefs and thoughts are connected with ("about") objects and events in the world, also known as the symbol-grounding problem].

Intent comprises the endogenous initiation, construction, and direction of behavior into the world. It emerges from brains. Humans, animals and autonomous robots select their own goals, plan their own tactics, and choose when to begin, modify, and stop sequences of action. Humans at least are subjectively aware of themselves acting, but consciousness is not a necessary property of intention. Unity appears in the combining of input from all sensory modalities into *Gestalts*, in the coordination of all parts of the body, both musculoskeletal and autonomic, into adaptive, flexible, yet focused movements. Subjectively, unity appears in the awareness of self and emotion, but again this is not intrinsic to intention. Wholeness is revealed by the orderly changes in the self and its behavior that constitute the development, maturation and adaptation of the self, within the constraints of its genes or design principles, and its material, social and industrial environments. Subjectively, wholeness is revealed in the remembrance of self through a lifetime of change, although the influences of accumulated and integrated experience on current behavior are not dependent on recollection and recognition. In brief, simulation of intentionality should be directed toward replicating the mechanisms by which goal states are constructed, approached and evaluated, and not toward emulating processes of consciousness, awareness, emotion, etc. in machines.

1.2 The limbic system is the chief organ of intentional behavior

Brain scientists have known for over a century that the necessary and sufficient part of the vertebrate brain to sustain minimal intentional behavior is the ventral forebrain, including those components that comprise the external shell of the phylogenetically oldest part of the forebrain, the paleocortex, and the deeper lying nuclei with which the cortex is connected. These components suffice to support remarkably adept patterns of intentional behavior, in dogs after all the newer parts of the forebrain have been surgically removed [17], and in rats with neocortex chemically inactivated by spreading depression [3]. Intentional behavior is severely altered or absent after major damage to the medial temporal lobe of the basal forebrain, as manifested most widely in Alzheimer's disease.

Phylogenetic evidence comes from observing intentional behavior in salamanders, which have the simplest of the existing vertebrate forebrains [21, 28]. The three main parts are sensory (which, as in small mammals, is predominantly olfactory), motor, and associational (Figure 1). These parts can be judged to comprise the limbic system in all vertebrates, but in the salamander they have virtually none of the "add-ons" found in brains of higher vertebrates, hence the simplicity. The associational part contains the primordial hippocampus with its interconnected septum and amygdaloid nuclei, striatal nuclei, which are identified in higher vertebrates as the locus of the functions of spatial orientation (the "cognitive map") and temporal integration in learning (the organization of long and short term memory). These processes are essential, inasmuch as intentional action takes place into the world, and even the simplest action, such as searching for food or evading predators, requires an animal to know where it is with respect to its world, where its prey or refuge is, and what its spatial and temporal progress is during sequences of attack or escape. The feedback loops that support the flow of neural activity in the neurodynamics of intentionality are schematized in Figure 2.

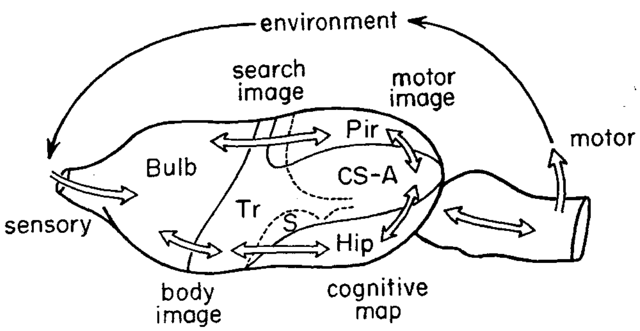


Figure 1. This schematic illustrates the sensory, motor, and associational components of the right hemisphere (seen from above) of the simplest extant vertebrate brain in the salamander. The bidirectional connections between these 3 major subdivisions of the forebrain provide for the macroscopic interactions that support the neurodynamics of the process of intentionality: goal formation, action, perception, and learning from the sensory consequences of the action taken into the environment. These components are form the prototype of the limbic system, which is found in all vertebrate brains, typically buried within exuberant growth of other "add-on" structures that operate in concert with the limbic system.

1.3 Neurodynamic manifestations of intentionality in brain activity of the primary sensory cortices: the EEG (electroencephalogram , 'local field potential')

The crucial question for neuroscientists is, how are the patterns of neural activity that sustain intentional behavior constructed in brains prior to perception? An answer is provided by studies of electrical activity of the primary sensory cortices of animals that trained to respond to conditioned stimuli [2, 8, 10-12, 14, 22, 23]. The construction is not by recall of stored patterns but by pattern formation in distributed nonlinear systems with connections that have been modified cumulatively through learning. The manner in which this take place involves hierarchical ordering of neural activity between microscopic, mesoscopic and macroscopic levels having differing time and space scales. Cortical neurons are selectively activated by sensory receptors and made to generate microscopic activity in the form of trains of action potentials (pulses) on their axons.. These and neighboring neurons by their synaptic interactions form a population forms that "binds" their activity into mesoscopic patterns 14, 18, 19, 29, 30]. These mesoscopic brain activity patterns are revealed by electrical fields of potential (EEGs) generated by interactive masses of neurons are induced by the arrival of stimuli, which trigger sequences of 1st order state transitions. These sequential states in turn converge into integrated macroscopic patterns that occupy the entirety of each cerebral hemisphere and give rise to the global patterns of brain activity, that may be related to the patterns of metabolic activity that are revealed by non-invasive brain imaging (fMRI, PET, SPECT, etc.).

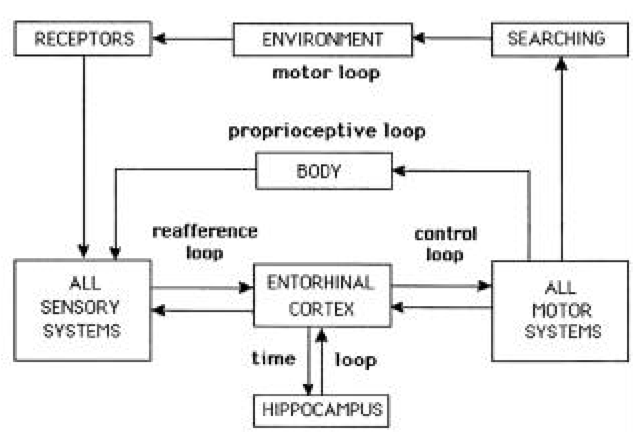


Figure 2 This diagram of brain state space maps the multiple feedback loops that support the intentional arc. Flow of neural activity inside the brain is in two directions. Forward flow from the sensory systems to the entorhinal cortex and on to the motor systems is by spatial AM patterns of action potentials at the **microscopic** level, by which transmitting cortices drive the neurons in their targets. Feedback flow from the motor systems to the entorhinal cortex by control loops, and from the entorhinal cortex to the sensory systems inside the brain, is by spatial AM patterns of action potentials at the **mesoscopic** level. This feedback constrains and modulates the microscopic activity in the forwardly transmitting populations. The mesoscopic feedback messages are order parameters that bias the attractor landscapes of the sensory cortices in preafference. Forward flow supports motor output and provides the content of percepts. Feedback flow supports integrative processes in learning that lead to the wholeness of intentionality. They enable the formation of a **macroscopic** AM pattern that reflects the integration of the activity of an entire hemisphere.

Owing to the nonlinear state transitions by which they form, these mesoscopic brain states are not representations of stimuli, nor are they simple effects caused by stimuli. Each learned stimulus serves to elicit the construction of a pattern that is shaped by the synaptic modifications between cortical neurons from prior learning, which vastly outnumber the synapses formed by incoming sensory axons, and also by the brain stem nuclei that bathe the forebrain in neuromodulatory chemicals. Each cortical activity pattern is a dynamic operator that creates and carries the meanings of stimuli for the recipient animal. It reflects the individual history, present context, and expectancy, corresponding to the unity and the wholeness of the intentionality. The patterns created in each cortex are unique to each animal. All sensory cortices transmit their signals into the limbic system, where they are integrated with each other over time, and the resultant integrated meaning is transmitted back to the cortices in the processes of selective attending, expectancy, and the prediction of future inputs, which together comprise the neural process of "reafference".

The same kinds of EEG activity as those found in the sensory and motor cortices are found in various parts of the limbic system. This discovery indicates that the limbic system also has the capacity to create its own spatiotemporal patterns of neural activity. They are related to past experience and convergent multisensory input, but they are self-organized. The limbic system provides a neural matrix of interconnections, that serves to generate continually the neural activity that forms goals and directs behavior toward them. EEG evidence shows that the process occurs in discontinuous steps, like frames in a motion picture. Each step follows a dynamic state transition, in which a complex assembly of neuron populations jumps suddenly from one spatiotemporal pattern to the next, as the behavior evolves. Being intrinsically unstable, the limbic system continually transits across states that emerge, spread into other parts of the brain, and then dissolve to give rise to new ones, a process that Japanese mathematicians have described as "chaotic itinerancy" between "attractor ruins" [34]. Its output controls the brain stem nuclei that serve to regulate its own excitability levels, implying that it regulates its own neurohumoral context, enabling it to respond with equal facility to changes that call for arousal and adaptation or rest and recreation, both in the body and the environment. It may be said that the neurodynamics of the limbic system, assisted by other parts of the forebrain such as the frontal lobes, initiates the novel and creative behavior seen in search by trial and error.

The limbic activity patterns of directed arousal and search are sent into the motor systems of the brain stem and spinal cord. Simultaneously, patterns are transmitted to the primary sensory cortices, preparing them for the consequences of motor actions. This process has been called "reafference" [12, 35], "corollary discharge" [32], "focused arousal" [29] and "preafference" [22, 23]. It sensitizes sensory systems to anticipated stimuli prior to their expected times of arrival Sensory cortical constructs consist of brief staccato messages to the limbic system, which convey what is sought and the result of the search. After multisensory convergence, the spatiotemporal activity pattern in the limbic system is up-dated through temporal integration in the hippocampus. Between sensory messages there are return up-dates from the limbic system to the sensory cortices, whereby each cortex receives input that has been integrated with the output of the others, reflecting the unity of intentionality. Everything that a human or an animal knows comes from this iterative circular process of action, reafference, perception, and up-date. It is done by successive frames that involve repeated state transitions and self-organized constructs in the sensory and limbic cortices. This neurodynamic system is defined here as the "limbic self" in the brain of an individual, where intentional behavior is created, with help from other parts of the forebrain.

An act of observation comprises Aquinas' intentional action of "stretching forth" and learning from the consequences. It embodies the existential "action-perception cycle" of Merleau-Ponty [26]. It corresponds to Piaget's [27]cycle of "action, assimilation, and adaptation" in the sensorimotor stage of childhood development. His postulated sequences of equilibrium, disequilibrium, and re-equilibration conform to state transitions in brain dynamics, which initiate and sustain action, construct dynamic patterns in the sensory cortices, and up-date the limbic patterns by modifying synapses in the learning that follows the sensory consequences of intended actions. For Piaget, cause and effect are chains of events that have the appearance of linkage corresponding to the unfolding experience of that exploration, by which a child is trying to make sense of its world by manipulating objects in it. The origin of causal inference is buried deeply in the pre-linguistic exploratory experience of each of us. It is not easily accessed by cognitive analysis or introspection.

We are all aware of our acts of observation. It is partly by expectation of what we are looking for through reafference, partly by perceiving the changes that our actions make in the dispositions of our bodies through proprioception, and partly by our selection of stimuli from the environment through exteroception. We perceive our intentional acts as the "causes" of changes in our perceptions, and the subsequent changes in our bodies as "effects" [12]. If this hypothesis of limbic dynamics is correct, then everything that we know we have learned through the action-perception cycle, including the iterative state changes by which it is produced in brains of animals and humans. It is this cycle, in prototypic form without need for appeal to consciousness, that must be simulated in our attempts to devise intelligent machines.

2.0 Characteristics of brain states as they are revealed by EEGs

The "state" of the brain is a description of what it is doing in some specified time period. A state transition occurs when the brain changes and does something else. For example, locomotion is a state, within which walking is a rhythmic pattern of activity that involves large parts of the brain, spinal cord, muscles and bones. The entire neuromuscular system changes almost instantly with the transition to a pattern of jogging or running. Similarly, a sleeping state can be taken as a whole, or divided into a sequence of slow wave and REM stages. Transit to a waking state can occur in a fraction of a second, whereby the entire brain and body shift gears, so to speak. The state of a neuron can be described as active and firing or as silent, with sudden changes in the firing manifesting state transitions. Populations of neurons also have a range of states, such as slow wave, fast activity, seizure, or silence. The mathematics of nonlinear dynamics is designed to study these states and the transitions by which they are accessed and abandoned.

2.1 The problem of stability of cortical states

The most critical question to ask about a state is its degree of stability or resistance to change. Evaluation is done by perturbing an object or a system [8]. For example, an object like an egg on a flat surface is unstable, but a coffee mug is stable. A person standing on a moving bus and holding on to a railing is stable, but someone walking in the aisle is not. If a person regains his chosen posture after each perturbation, no matter in which direction the displacement occurred, that state is regarded as stable, and it is said to be governed by an attractor. This is a metaphor to say that the system goes ("is attracted") to the state through an interim state of transience. The range of displacement from which recovery can occur defines the basin of attraction, in analogy to a ball rolling to the bottom of a bowl. If the perturbation is so strong that it causes concussion or a broken leg, and the person cannot stand up again, then the system has been placed outside the basin of attraction, and a new state supervenes with its own attractor and basin.

Stability is always relative to the time duration of observation and the criteria for what is chosen to be observed. In the perspective of a lifetime, brains appear to be highly stable, in their numbers of neurons, their architectures and major patterns of connection, and in the patterns of behavior they produce, including the character and identity of the individual that can be recognized and followed for many years. Brains undergo repeated transitions from waking to sleeping and back again, coming up refreshed with a good night or irritable with insomnia, but still, giving the same persons as the night before. Personal identity is usually quite stable. But in the perspective of the short term, brains are highly unstable. Thoughts go fleeting through awareness, and the face and body twitch with the passing of emotions. Glimpses of their internal states of neural activity reveal patterns that are more like hurricanes than the orderly march of symbols in a computer. Brain states and the states of populations of neurons that interact to give brain function, are highly irregular in spatial form and time course. They emerge, persist for a small fraction of a second, then disappear and are replaced by other states. It is the flexibility and creativeness of this process that makes it so successful in animals for their adaptation to rapidly changing and unpredictable environments, and that makes it the desired platform on which to base the design of intelligent machines.

2.2 Three types of stable cortical states

In using dynamics we approach the problem by defining three kinds of stable state, each with its type of attractor. The simplest is the point attractor. The system is at rest unless perturbed, and it returns to rest when allowed to do so. As it relaxes to rest, it has the history of what happened, but that history is lost after convergence to rest. Examples of point attractors are silent neurons or neural populations that have been isolated from the brain, and also the brain that is depressed into inactivity by injury or a strong anesthetic, to the point where the EEG has gone flat (Figure 3, bottom trace). A special case of a point attractor is noise. This state is observed in populations of neurons in the brain of a subject at rest, with no evidence of overt behavior. The neurons fire continually but not in concert with each other. Their pulses occur in long trains at irregular times. Knowledge about the prior pulse trains from each neuron and those of its neighbors up to the present fails to support the prediction of when the next pulse will occur. The state of noise has continual activity with no history of how it started, and it gives only the expectation that its amplitude and other statistical properties will persist unchanged.

A system that gives periodic behavior is said to have a limit cycle attractor. The classic example is the clock. When it is viewed in terms of its ceaseless motion, it is regarded as unstable until it winds down, runs out of power, and goes to a point attractor. If it resumes its regular beat after it is re-set or otherwise perturbed, it is stable as long as its power lasts. Its history is limited to one cycle, after which there is no retention of its transient approach in its basin to its attractor. Neurons in populations rarely fire periodically, and when they appear to do so, close inspection shows that the activities are in fact irregular and unpredictable in detail, and when periodic activity does occur, it is either intentional, as in rhythmic drumming, clapping and dancing, or it is pathological, as in the periodic oscillations of the eyes in nystagmus, or of the limbs during Parkinsonian tremor, or of the cortex during the hypersynchrony of partial complex seizures that are revealed by near-periodic spike trains (Figure 3, top trace).

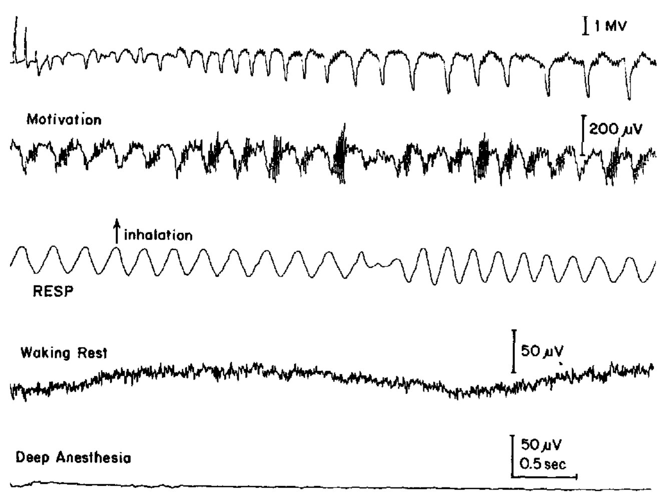


Figure 3. Four levels of function of the olfactory system are revealed by EEG recording. The lowest is the non-interactive 'open loop' state imposed by deep anesthesia, which suppresses brain activity. The next is the resting steady state with broad spectrum $1/f^2$ aperiodic waves. The aroused level in which behavior is generated is shown by the repeated state transitions, by which bursts are formed that reveal spatial patterns of AM (amplitude modulation) relating to odorant recognition with inhalation. The upper trace shows the pattern of high-amplitude spikes when an epileptic seizure has been triggered by powerful electrical stimulation. This state is likewise chaotic, but with a reduced correlation dimension. This state also occurs during recovery from deep anesthesia on the way to the resting state [9, 31].

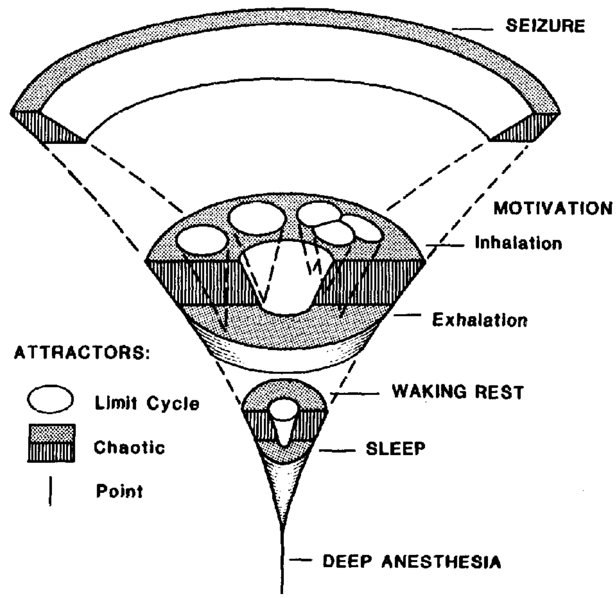


Figure 4. A bifurcation diagram of the olfactory system state space is constructed from the EEGs in Figure 3.

The third type of attractor gives aperiodic oscillation of the kind that is observed in recordings of EEGs. There is no one or small number of frequencies at which the system oscillates. The system behavior is therefore unpredictable, because performance can only be projected far into the future for periodic behavior. This type is now widely known as "chaotic". The existence of this type of oscillation was known to Poincaré a century ago, but systematic study was possible only recently after the full development of digital computers. The best known systems with chaotic attractors have a small number of components and a few degrees of freedom, as for example, the double-hinged pendulum, the dripping faucet, and the Lorenz, Chua, and Rössler attractors [13]. These simple models are stationary, autonomous, and noise-free, forming the class of "deterministic chaos". Large and complex real-world systems, which include neurons and neural populations are noisy, infinite-dimensional, nonstationary, non-autonomous, yet capable of chaotic behavior which has been called "stochastic chaos" [14]. The source is postulated to be the synaptic interaction of millions of neurons, which create fields of microscopic noise in cortex, but which are constrained by their own interactions to generate mesoscopic order parameters that regulate the spatiotemporal patterns of cortical activity revealed by the EEG. These spatiotemporal patterns are revealed by spatial patterns of amplitude modulation ("AM patterns") of a spatially coherent aperiodic carrier wave in the gamma range of the EEG. They appear in time series as bursts of oscillation (Figure 3), and their spatial patterning indicates the existence of an attractor landscape, which is actualized in the olfactory system with each inhalation (Figures 4 and 5 during intentional behavior.

The discovery that brain dynamics operates in chaotic domains has profound implications for the study of higher brain function [31]. A chaotic system has the capacity to create novel and unexpected patterns of activity. It can jump instantly from one mode of behavior to another, which manifests the facts that it has a collection of attractors, each with its basin, and that it can move from one to another in an itinerant trajectory [34]. It retains in its pathway across its basins its history, which fades into its past, just as its predictability into its future decreases. Transitions between chaotic states constitute the dynamics that we need to understand how brains perform such remarkable feats as abstraction of the essentials of figures from complex, unknown and unpredictable backgrounds, generalization over examples of recurring objects never twice appearing the same, reliable assignment to classes that lead to appropriate actions, and constant up-dating by learning.

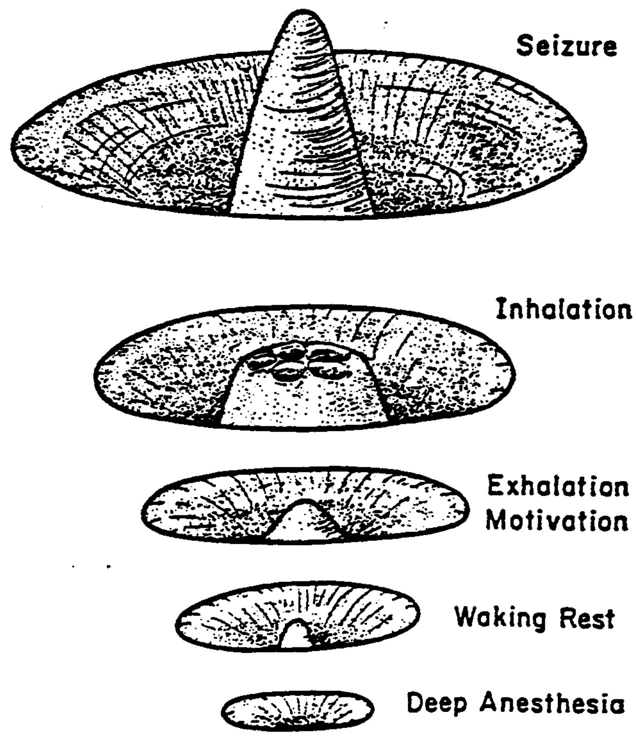


Figure 5. This perspective drawing of a projection from an infinite dimensional brain state space into 3-space offers a view of how an attractor landscape of learned basins of attraction is created with each inhalation. The selection is made by the input odorant. If the stimulus is novel or unknown, the system goes into the chaotic well, which provides the aperiodic umpatterned activity that drives Hebbian learning for new basin formation.

2.3 The 1st order cortical state transition is an elemental step in intention

Systems such as neurons and brains that have multiple chaotic attractors also have point and limit attractors, each with its basin of attraction, which serves to provide the generalization gradient required for perception of recurring stimuli that are never twice the same. If the basin is that of a point or a limit cycle attractor, the system can proceed predictably to an identical end state. If the basin leads to a chaotic attractor, the system goes into ceaseless fluctuation, as long as its energy lasts. If the starting point is identical on repeated trials, which can only be assured by simulation of the dynamics on a digital computer, the same aperiodic behavior appears. If the starting point is changed by an arbitrarily small amount, although the system is still in the same basin, the trajectory is not identical. A deterministic chaotic system that is in the basin of one of its chaotic attractors is legendary for its sensitivity to the initial conditions. If the difference in starting conditions is too small to be originally detected, it can be inferred from the unfolding behavior of the system, as the difference in trajectories becomes apparent. This observation shows that a chaotic system has the capacity to create novel patterns constituting endogenous increases in information in the course of continually constructing its own trajectory into the future.

Our EEG evidence indicates that every primary sensory cortex maintains multiple basins corresponding to previously learned classes of stimuli, as well as to the unstimulated state, which together form an attractor landscape. They all show evidence that the vehicle they use for transmission of their output is an aperiodic carrier wave that is amplitude-modulated in the two spatial dimensions of cortical coding, and that is gated by extra-cortical forcing functions in the theta range (2-7 Hz). We note that we predicted a common code for all sensory systems, on the basis that the signals from all sensory cortices must be combined in the limbic system to form gestalts. We postulate that preafferent input from the limbic system can serve to bias the landscapes in such a way as to facilitate the capture of the multiple sensory systems by basins of the attractors corresponding to the goal of the intended observation, perhaps in the manner of the variable tiling in a Voronoi diagram. This chaotic prestimulus state of expectancy establishes the sensitivities of the cortices, so that the very small number of sensory action potentials evoked by the expected stimuli can simultaneously carry the cortical trajectories into the basins of the appropriate attractors as they are created by the forcing function, in the case of olfaction by inhalation (Figure 5), irrespective of which equivalent receptors actually receive the expected stimuli in the different sensory modalities. In the absence of the stimulus, the cortices continue to transmit their outputs to the limbic system, confirming the continuing absence. The stimuli are also selected by the limbic system through orientation of the sensory receptors in space by sniffing, looking, and listening. We believe that the basins of attraction in each of the sensory cortices are shaped by limbic input to sensitize them for receiving and processing the desired class of stimuli in every modality, whatever may be the goal at the moment of choice.

3.0 Problems in use of chaotic dynamics in the development of advanced machine intelligence

Chaotic dynamics has proved to be extremely difficult to harness in the service of intelligent machines. Most studies that purport to control chaos either find ways to suppress it and replace it with periodic or quasiperiodic fluctuations, or to lock two or more oscillators into synchrony sharing a common aperiodic wave form, often as an optimal means for encryptation and secure transmission. Our aim is to employ chaotic dynamics as the means for creating novel and endogenous space-time patterns, which must be the means to achieve any significant degree of autonomy in devices that must operate far from human guidance, where in order to function they must make up their courses of action as they go along. We know of no other way to approach a solution to the problem of how to introduce creative processes into machines, other than to simulate the dynamics we have found in animal brains. To be sure, there are major unsolved problems in this approach, chief among them that we know too little about the dynamics of the limbic system. Hence we find it necessary to restrict the development of hardware models to the stage of brain-world interaction that we know best, which is the field of perception. In brief, what are the problems in giving eyes, ears and a nose to a robot, so that it might learn about its environment in something like the way that even the simpler animals do - by creating hypotheses and testing them through their own actions?

3.1 Noise stabilization of chaotic dynamics, opening the way to analog-digital hybrid embodiments

The operations in the olfactory system by which the state transitions and pattern constructions for pattern classification are simulated in software and hardware embodiments have been described in a series of publications [9-12, 14]. Our simulations are done with a set of approximately 920 interconnected first-order nonlinear ordinary differential equations, forming what we have named the KIII model [8]. The basic element, the KO set, is a 2-stage linear integrator simulated in hardware [6, 7] by 2 operational amplifiers, whose output is passed through an asymmetric sigmoid function modeled by 2 diodes back-to-back. Connections between 64 elements are time multiplexed (Figure 6) through a MUX, an amplifier with voltage-controlled gain, and a DMUX [10]. Switching is controlled by a digital computer at a clock rate suitable for the pass band of the carrier wave. For each connected pair the gain is stored in memory, so that the connection strengths are easily modified during learning. With this device the connectivity grows by 2-N instead of by N^{2}›. In digital embodiment the equations have been solved by numerical integration on Unix, Macintosh, and PC platforms, and by vector programming on the Cray M/X.

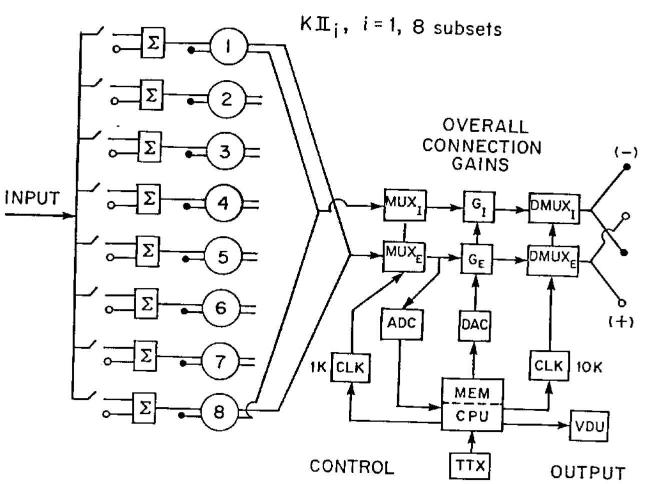


Figure 6. Schematic for connecting KII sets by multiplexing.

Interaction of KO sets of like kind (excitatory or inhibitory) giving point attractors is modeled by KI sets; interaction of KIe and KIi sets giving limit cycle attractors is modeled by KII sets. Three serial KII sets in layers that correspond to the olfactory bulb, prepyriform cortex, and an intervening control nucleus called the AON is modeled by the KIII set; if the 3 characteristic frequencies are incommensurate, and the feedback delays between the 3 layers are distributed to act as low-pass filters, the solutions of the equations give the aperiodic waveforms and broad $1/f^2$ spectra (Figure 7) of EEGs trom the 3 layers. The asymmetric sigmoid endows the system with the property of nonlinear state transitions on step inputs, owing to the amplitude-dependent gain of the KO elements.

In the course of digital simulation it has become apparent that a minimum of 64 elements will suffice for 2-D pattern classification under Hebbian and non-Hebbian reinforcement learning [16, 24, 25, 37, 38]. The large number of equations leads to attractor crowding [15], in which the basins of attraction shrink close to the size of the digitizing step in using rational numbers for computation, so that sooner or later the system jumps out of its designated chaotic basin into a neighboring basin that is most likely to be that of a point or limit cycle attractor, which kills the system. This problem has been solved by use of additive noise on the order of 15% of the amplitude of the aperiodic state variables [4, 5, 13, 15], giving robust attractor landscapes for learning and pattern classification [24, 25]. The lesson learned is that deterministic chaos, in which the system is low-dimensional, stationary, strictly autonomous, and noise-free, is inappropriate for modeling biological and machine intelligence. Brains operate with what we call 'stochastic chaos' [13], which is high-dimensional, nonstationary with regularly repeated state transitions, engaged with its surround, and deeply embedded in noise created by KIe sets and manifested in high densities of action potentials. The noise in digital models is simulated with random number generators, either rectified to simulate KIe sets or off-set with d.c. bias to simulate the noise in KIIei sets.

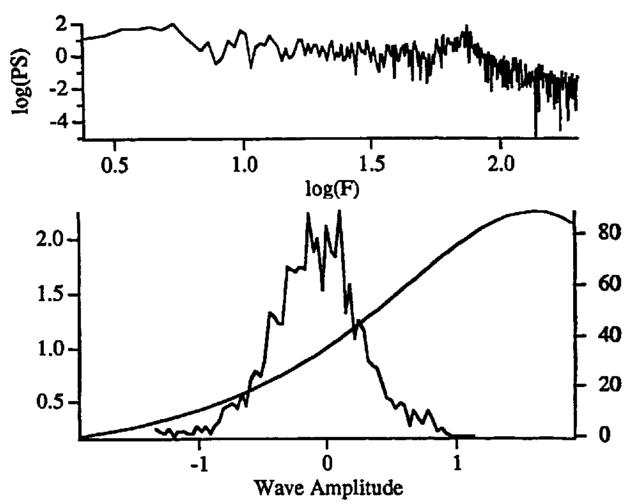


Figure 7. The power spectrum and amplitude histogram for a simulated EEG trace from the KIII model, with a section of the asymmetric nonlinear gain curve, showing the nature of the nonlinearity that provides for destabilization by the input. The interactive gain increases with excitatory input.

The finding in digital embodiments that noise is not only unavoidable but is necessary for stable high-dimensional chaotic dynamics opens the way to analog embodiments [7], in which noisy components resemble the characteristics of local pools in nerve cell assemblies, but which offer much higher rate of temporal and spatial integration, the use of continuous variables in place of rational numbers, and the feasibility of implementing the dynamics on chips suitable for incorporation into mobile devices.

3.2 Embedding devices for perception into autonomous cognitive machines

The KII sets have multiple robust limit cycle attractors, which become embedded as chaotic attractors when coupled in serial layers with distributed delayed feedback. The KIII model is offered as the prototype for constructing devices in hardware and software to implement the elementary steps of perception, thus providing robots with the sensory ports that they need to guide them through their environments. These steps are the interpretive operations necessary to normalize, compress, abstract and generalize over successive inputs preparatory to classification [5, 16, 25, 33, 36]. These cognitive operations are done by the nonlinear operations in the input stage and by the basins of attraction in the landscape formed by learning in each of the sensory systems. They are required in each of the ports providing information to the mobile device about its visual, auditory, tactile, and chemical environments. Our tests of the KIII model have shown that it can learn a new class in half a dozen trials instead of the thousands of trials required by MLPs, and that new learning occurs without degradation of previous attractors, although, as in the case of the olfactory system, the attractors are modified through attractor crowding. The superior level of 'intelligence' is demonstrated by the capacity of the KIII model to separate items in 64-space that belong to identifiable classes but are not linearly separable. The classes are, in fact, constructed by the model and are not imposed from outside, constituting an aspect of autonomy. In other words, the system creates its own features from its own experience of the constancy of relations between channels in the 8x8 64-channel input array.

Formation of a world-view by which the device can guide its explorations for the means to reach its goals depends on the integration of the outputs of the several sensory systems, in order to form a multisensory percept known as a gestalt. This integration is easily done when all of the ports have their outputs in the same form: a vector consisting of a 2-D spatial pattern of amplitude modulation of a 1-D aperiodic wave form in the gamma range (nominally 30-60 Hz), which is segmented in time at a frame rate of nominally 2-7 Hz and frame durations on the order of 0.1 sec. Precise clocking and synchronization are not prerequisite.

The sequential frames deriving from sampling the environment must then be integrated over time and oriented in space. An example of how these higher operations might be done was provided by W. Gray Walter [36] with his electronic tortoises, which had the capacity for autonomous goal-directed search involving the adjudication of conflicting needs in an uncertain environment.

The performances of these devices set a challenging level of 'intelligence' to which to aspire, and they also serve to highlight some of the difficulties in using the descriptive term "autonomous". As with animals the devices were untethered, and they learned to avoid obstacles without need for instruction or intervention, if within their limited capacities for locomotion. However, they were programmed to satisfy their own needs without regard for or comprehension of anything else's, perhaps in analogy to house pets, whose sole purpose, however inadvertent, is to provide enjoyment to their owners, and seldom to do useful work or bend their talents to the benefits of the owners, or, in the case of the machines, the designers and builders.

It is already apparent that fully autonomous vehicles are not in the best interest of researchers and the general public, except as demonstrations of what might emerge as major problems from this line of study. It is also clear that such devices can and will be built, and that the proper path of future management will not be by techniques of training and aversive conditioning, but by education, with inculcation of desired values determined by the manufacturers that will govern the choices that must by definition be made by the newly autonomous mechanical devices.

Acknowledgments

This work was supported by grants from the National Institute of Mental Health MH06686, the Office of Naval Research N00014-90-J-4054, and from ARO MURI DAAH04-96-1-0341.

References

1] Aquinas, St. Thomas. (1272). *Treatise on Man*. In: Summa Theologica. Translated by Fathers of the English Dominican Province. Revised by Daniel J Sullivan. Published by William Benton as Volume 19 in the Great Books Series. Chicago: Encyclopedia Britannica, Inc., 1952.

2} Barrie JM, Freeman WJ, Lenhart M (1996) Modulation by discriminative training of spatial patterns of gamma EEG amplitude and phase in neocortex of rabbits. Journal of Neurophysiology 76: 520-539.

3] Bures J, Buresová O, Krivánek J (1974) The Mechanism and Applications of Leão's Spreading Depression of Electroencephalographic Activity. New York: Academic Press.

4] Chang H-J, Freeman WJ (1998) Biologically modeled noise stabilizing neurodynamics for pattern recognition. International Journal of Bifurcation & Chaos 8 (2), 321-345.

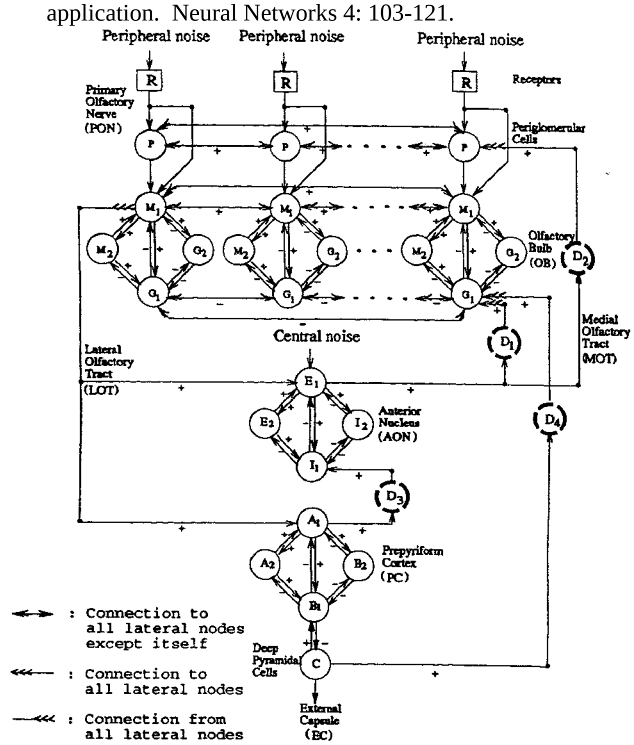
5] Chang H-J, Freeman WJ (1999) Local homeostasis stabilizes a model of the olfactory system globally in respect to perturbations by input during pattern classification. International Journal of Bifurcation and Chaos 8: 2107-2123.

6] Edelman JA, Freeman WJ (1990) Simulation and analysis of a model of mitral-granule cell population interactions in the mammalian olfactory bulb. Proceedings IJCNN I: 62-65.

7] Eisenberg J, Freeman WJ, Burke B (1989) Hardware architecture of a neural network model simulating pattern recognition by the olfactory bulb. Neural Networks 2: 315-325.

8] Freeman WJ (1975) Mass Action in the Nervous System. New York: Academic Press.

9] Freeman WJ (1987) Simulation of chaotic EEG patterns with a dynamic model of the olfactory system. Biological Cybernetics 56: 139-150.
10] Freeman WJ (1988) Pattern learning and recognition device. United States Patent # 4, 748, 674, May 31, 1988.
11] Freeman WJ (1992) Neurons to Brain Chaos. International Journal of Bifurcation and Chaos 2: 451-482.
12] Freeman WJ (1995) Societies of Brains. Mahwah NJ, Lawrence Erlbaum Associates.
13] Freeman WJ (1999) Noise-induced first-order phase transitions in chaotic brain activity. International Journal of Bifurcation and Chaos 9: 2215-2218.
14] Freeman WJ [2000] Neurodynamics. An Exploration of Mesoscopic Brain Dynamics. London UK: Springer-Verlag.
15] Freeman WJ, Chang H-J, Burke BC, Rose PA, Badler J (1997) Taming chaos: Stabilization of aperiodic attractors by noise. IEEE Transactions on Circuits and Systems 44: 989-996.
16] Freeman WJ, Yao Y, Burke B. (1988) Central pattern generating and recognizing in olfactory bulb: A correlation learning rule. Neural Networks 1: 277-288.
17] Goltz FL (1892) Der Hund ohne Grosshirn. Siebente Abhandlung über die Verrichtungen des Grosshirns. Pflügers Archiv 51: 570-614.
18] Gray CM (1994) Synchronous oscillations in neuronal systems: mechanisms and functions. Journal of Comparative Neuroscience 1: 11-38.
19] Haken H (1983) Synergetics: An Introduction. Berlin: Springer.-Verlag
20] Hardcastle VG (1994) Psychology's binding problem and possible neurobiological solutions. Journal of Consciousness Studies 1: 66-90.
21] Herrick CJ (1948) The Brain of the Tiger Salamander. Chicago IL: University of Chicago Press.
22] Kay LM, Freeman WJ (1998) Bidirectional processing in the olfactory-limbic axis during olfactory behavior. Behavioral Neuroscience 112: 541-553.
23] Kay LM, Lancaster L, Freeman WJ (1996) Reafference and attractors in the olfactory system during odor recognition. International Journal of Neural Systems 7: 489-496.
24] Kozma R, Freeman WJ (1999) A possible mechanism for intermittent oscillations in the KIII model of dynamic memories - the case study of olfaction. Proceedings, IJCNN'1999, Washington DC.
25] Kozma R, Freeman WJ (2000) Encoding and recall of noisy data as chaotic spatio-temporal memory patterns in the style of the brains. Proceedings, ICJNN'2000. Como, Italy
26] Merleau-Ponty M (1942) The Structure of Behavior (AL Fischer, Trans.). Boston: Beacon Press (1963).
27] Piaget J (1930) The child's conception of physical causality. New York: Harcourt, Brace. p. 269
28] Roth G (1987) Visual Behavior in Salamanders. Berlin: Springer-Verlag
29] Sheer DE (1989) Sensory and cognitive 40-Hz event-related potentials: Behavioral correlates, brain function, and clinical application. Brain Dynamics. Basar E, Bullock TH (eds.) Berlin: Springer-Verlag.
30] Singer W, Gray CM (1995) Visual feature integration and the temporal correlation hypothesis. Annual Review of Neuroscience 18: 555-586.
31] Skarda CA, Freeman WJ (1987) How brains make chaos in order to make sense of the world. Behavioral and Brain Sciences 10: 161-195.
32] Sperry RW (1950) Neural basis of the spontaneous optokinetic response. Journal of Comparative Physiology 43: 482-489.
33] Storm C, Freeman WJ (1999) A novel dynamical invariant measure addresses the stability of the chaotic KIII neural network. Proceedings, IJCNN'1999, Washington DC.
34] Tsuda I (1996) A new type of self-organization associated with chaotic dynamics in neural networks. International Journal of Neural Systems 7: 451-459.
35] von Holst E & Mittelstaedt H (1950) Das Reafferenzprinzip Naturwissenschaften 37: 464-476.
36] Walter WG (1963) The Living Brain. New York: Norton.
37] Yao Y, Freeman WJ (1990) Model of biological pattern recognition with spatially chaotic dynamics. Neural Networks 3: 153-170.
38] Yao, Y., Freeman WJ, Burke, B., Yang, Q. (1991) Pattern recognition by a distributed neural network: An industrial application. Neural Networks 4: 103-121.



Measuring intelligence: a neuromorphic perspective

Marwan A. Jabri
Electrical and Computer Engineering
Oregon Graduate Institute, 20000 NW Walker Rd,
Beaverton, OR 97006 USA
and
School of Electrical and Information Engineering
The University of Sydney, NSW 2006 Australia
marwan@ece.ogi.edu

ABSTRACT

Neuromorphic engineering is about the development of biologically inspired roving machines that can exhibit intelligent behaviour, learn on-line and in real-time. The question of how to assess and measure the intelligence of such machines is essential if progress in neuromorphic engineering is to be assessed. However, it is awkward to talk about measuring intelligence without a clear understanding of the capabilities that researchers aim or dream to equip neuromorphic systems with. In this communication we promote the position that metrics for measuring of the intelligence of neuromorphic systems should be task-based, should factor in the computational resources, the on-line learning efficiency, the capability to learn from intermittent reward that can vary in frequency and importance to the task at hand, the capability to anticipate events and to modify decision making processes based on anticipated events, the capability to balance exploration and exploitation as to discover new methods or to fine-tune existing methods, and the ability to optimize the utilization of its resources using ground rules that maximizes it success. To factor in all these aspects requires a fundamental assessment of what such machines achieve as goals and at what cost. We propose that a simple achievement rule, energy and resource oriented metric be used.

Keywords: *neuromorphic engineering, on-line learning, reward-based learning, anticipation, exploration and exploitation, regularity and modularity.*

1 Introduction

Neuromorphic engineering was a term coined by Carver Mead and described the process of building systems based on biological models and embedding them in roving machines. In the last 20 years, neuromorphic engineering addressed the development of various biological like processing system such as retinas, cochleas (Schaik 2000), legged robots and creatures (Tilden 1994) (Lewis, Etienne-Cummings et al. 2000), sensorimotor control (Horiuchi and Koch 1999) (Etienne-Cummings, Spiegel et al. 2000) and integration systems (Jabri, Coenen et al. 1997).

Although analog microelectronics was initially promoted (and continue to some extent) as the ideal substrate for neuromorphic information processing systems (Mead 1989), current works tend to use many implementation technologies, hardware and software.

The aim of many neuromorphic engineering groups is to develop active perception systems, systems that interact with the environment in a closed loop fashion[1].

Neuromorphic engineering is a synergy between neuroscience and engineering. The common neuromorphic methodology is to identify a task or a function, to explore and identify brain areas from neuroscientific knowledge (anatomy, physiology,

[1] The Telluride Neuromorphic Engineering workshop is a yearly meeting where research groups meet and collaborate. See http://zig.ini.unizh.ch/telluride2000.

psychophysics, …), to develop computational models that encapsulate the information processing at some level of abstraction, and to develop implementations of the models. The determination of an acceptable level of abstraction of the biological systems during the computational model development is a challenging task, and is typically done as to preserve some essence of the biological information or mechanical processes.

In assessing the intelligence of engineering machines, and because of its close relationship to the neuroscientific community, neuromorphic engineering has traditionally relied on several levels of metrics. Not all metrics are necessarily directly related to the behaviour of the machines and the classification of the intelligence of such behaviour. The common levels are:

- Device/circuits
- Representation
- Organization
- Behaviour with and without artificial lesions
- Learning & behaviour adaptation

In these assessments, tasks have commonly been related to the biological systems being modeled: specific brain areas, the central nervous system, and the mechanical apparatus. We elaborate on these tasks in the next section.

2 Neuromorphic Tasks – Present and Future

2.1 *Peripherals systems*

Biologically based or inspired peripheral systems are probably the most researched neuromorphic systems. The development of silicon-based implementations of retinas and cochleas has been pioneered in Carver Mead's laboratory in the eighties. Artificial olfactory and somatosensory systems have also been researched and developed.

It is clear why most early neuromorphic research focused on peripheral systems: They are the sensors and they drive the motor responses of biological systems, and they are the most understood, in particular in the case of primates.

The research and development of neuromorphic peripheral systems has also contributed to better understanding of the biological devices and they incorporation in systems.

2.2 *Sensorimotor Systems*

Over the last decade sensorimotor systems have been developed. Sensorimotor system are broad in their definition but are supposed to implement forms of sensory (visual, auditory, infrared, sonar) to motor mapping, where the motor are actions that aim at performing forms of active perception, navigation and tracking, or orienting to stimuli in the environment.

Experimental sensorimotor systems have incorporated abilities by incorporating simplified models of the superior colliculus, goal reaching and simple navigation abilities by incorporating computational models of the basal ganglia and ventral tegmental area, predictive control abilities by incorporating computation models of the cerebellum, and spatial representation learning by incorporating computational models of the hippocampal formation.

Sensorimotor systems have so far included sophisticated adaptive learning abilities implemented in software. The measuring of the intelligence of such systems has largely been a matter of retrieving behavioral properties that resemble those of animal, when the systems are implementing sensorimotor tasks, or by observing the behaviour of the system when software lesions are performed. In that case the deficit of the systems are typically compared to those of animal that have had specific brain areas severed or temporarily disabled.

Analog Very Large Scale Integration (aVLSI) systems with adaptive abilities have also been implemented and typically mimic to some extent their biological counter parts. The assessment of the intelligence of such system has largely been a matter of comparing the signal processing or collective computation of the devices to the biological counterparts. These could also be seen as task-oriented comparison. An example is an implementation of the retina with adaptive intensity saturation control. Another example is a silicon cochlea that implements adaptive gain control.

2.3 Cognitive Systems

If one defines cognitive systems as being capable of performing higher order processing by utilizing first order information and generating higher order knowledge, some sensorimotor systems would qualify of being “cognitive”.

An example is a sensorimotor system similar to that of Fig 1 which implements abstract computational of the cerebellum, basal ganglia and ventral tegmental area.

In this system the cerebellum performs sensory prediction and coordinate transformation from world coordinate to robot centered coordinate. The predicted sensory signal (visual target position) is then used by the basal ganglia to associate a motor command with the visual target as to keep it as much as possible in front of it. If the basal ganglia are lesioned, the robot looses its motor ability. If the cerebellum is lesioned, the tracking lags the object.

In survival terms, and such a sensorimotor system is controlling the hunting abilities of an animal, a lesion of the cerebellum would most likely lead to the animal death, although it can track its prey, though not predictively to the point that it can catch it (assuming a mobile prey), or it cannot escape a predator by anticipating potential contact points. Interestingly, the hypothetical animal would be able to anticipate the position and perform all desired coordinate transformation, however a lesion of the basal ganglia will also lead to its death.

2.4 Future of Neuromorphic Systems

With the rapid development in neuroscience research brought by phenomenal growth in computation, sensing, signal processing and imaging technologies, neuromorphic engineering will increasingly focus on the implementation of complex motor, sensory and cognitive processing. The development of computational models of sub-cortical and cortical will permit the development of sophisticated real-time systems, that will go beyond present sensorimotor loops and will integrate aspects such as planning, object recognition, motion and auditory analysis, and perception. This will put additional pressure in comparing the performance of such systems, and hence on the issue of measurement metrics.

3 Computational Resources

Computational resources in neuromorphic systems, in particular aVLSI systems tend to be a central criterion of design. One attraction of aVLSI neuromorphic systems is the low power requirements (Mead 1989; Jabri, Coggins et al. 1996). However, beside the elegance of the implementation, and specific application requirements, it is becoming more difficult to promote analog as a preferred design methodology, except in some fairly narrow areas such as world interfaces. This is not to say that analog asynchronous parallel computation does not provide any conceptual computational advantages. Only that the inspirations for such advantages have not been met with clear theoretical support over digital computation as yet.

Computational resources have also been considered from the point of view of compactness, efficiency of representing basic computational elements such as sensors and signal processors. Here applications that have specific requirements such as ultra low power and high fault tolerance capability could benefit more from analog than digital representation. This is particularly the case if sparse representation is being used. In a sparse representation of neural networks, neurons within a hierarchy of computation do not fire concurrently. The receptive fields of the neurons are highly tuned/selective and are independent of each other. This translates into data-driven architecture with attractive low power consumption properties.

Given the infancy of neuromorphic systems, autonomous behaviour has not been developed beyond adaptive sensing and signal processing tasks.

4 On-line Learning

Continuous on-line learning with bound resources represents a challenge because of the following problems:

1- Frequency of the associations to be learnt is not sufficiently high to be captured in a distributed representation. Note the tuning pf learning parameters do not necessarily solve this problem as for example, the use of large learning rate can lead to prior information to be forgotten (catastrophic learning effects) if no processes are implemented to move and consolidate information from soft-term memory to long-term memory store.

2- In cases where statistical properties of the sensed signals are to be discovered on-line, sample size effects, and non-stationarity of the signals are very problematic. For example if independent component analysis techniques are being used to discover feature detectors (Bell and Sejnowski 1995), such discovery using information maximization techniques and mutual independence criteria of the features would be more difficult to achieve if performed on-line.

3- Rapid and flexible learning schedules is necessary in situations where autonomous systems requires to learn at various rates and in real-time. This imposes constraints on propagation of information in the system and on its time constants. For example systems doing sequence learning require significant memory resources in the form of analog or digital delay lines and the performance of credit assignment through time over the present and historical information.

The learning issues above represent significant challenges to the incorporation of online and continuous learning. The proposition of metrics for these sort of capabilities is premature, given we do not really know the how, when and where of such capabilities.

5 Anticipation

An important element in neuromorphic systems research is the development of the concept of anticipation within the context of autonomy. The system described earlier in Section 2.3 is an example demonstrating an anticipation property. A roving robot that can anticipate undesirable events would maximize its mission success. Anticipation or prediction of sensory or motor control (predictive control) have been attributed as a role to the cerebellum (Coenen and Sejnowski 1996; Coenen 1998), in addition to the traditional attributed role of motor learning (Marr 1969; Albus 1971).

Present computational models of the cerebellum have addressed individual sensory (or a few) and motor prediction capability. Computational models that demonstrate abilities to adaptively and continuously deal with a large number of sensory modality and motor learning skills are still to be developed. Such skills will be essential to autonomous machines that are expected to perform tasks such as navigation in complex terrains or to perform object manipulation. Anticipation is also important for planning because it affects the performance of the machine and its interaction with the environment and its objects.

Measuring anticipation can be very subjective. However factoring anticipation in the overall goal of a machine will provide easier means for assessment.

6 Curiosity, Exploration and Exploitation

Autonomous machines should possess elements of "curiosity". For instance, it is known that reinforcement based learning algorithms depend on forms of exploration (Sutton and Barto 1981). However, exploration has so far been implemented in terms of probabilistic random actions aimed at exploring the state-space with hope of discovering policies that can be effective in achieving specific goals. The issues of either exploring more effectively or in a directed way, or to explore better policies and solutions are not well understood.

Another important aspect of autonomous systems is that of exploitation of infrequent, but yet important information encountered during machine experiences. The interactions between exploration and exploitation are fundamental in that regard. Reinforcement based learning algorithms have assumed that rewards are specified as end-achievements to the learning machine. The ability to discover and capture sensorimotor associations to yet unspecified goals (and reward) is essential to the rapid learning and the effective exploitation of sensorimotor experiences. To achieve this, the learning machines must be able to recognize unspecified or unscheduled rewards by forms of assessment of its sensory state and its sensory-reward memory.

7 Robustness and fault tolerance

Autonomous systems have to be robust and fault tolerant. We discussed in Section 3 sparse representation and their low power property as well as their potential role in more effective learning by decorrelating features. It is not clear however, without clear redundancy in the underlying resources (e.g. synapses and neurons), that sparse representation alone lead to more fault tolerance. It is also conceivable that other additional encoding representations, such as population-based be a source of fault tolerance (see for instance motor population coding (Georgopoulos 1995)).

Fault tolerance has been attributed to traditional neural network representation because of the distributed representation that develop during learning or that have been hand-crafted. The relationship between pure distributed representation and neural correlate is not trivial, nor automatic. Biological systems have various level of fault-tolerance, some of which is not graceful. Although biological systems survive significant faults, behaviour is commonly degraded or lost. For example in humans or monkeys, the level of behaviour change depends greatly on brain areas that are damaged.

Then, what role does fault tolerance plays in measuring intelligence? From an application point of view, fault tolerance is an important property of designs and system operation. Furthermore, with continuous shrinkage in transistor sizes, the importance of fault-tolerance in highly complex processing system will become increasingly important.

Another more important aspect of fault tolerance requirement is in autonomous system. Here clearly fault tolerance becomes a critical element of endurance and graceful degradation. But is this an important element of intelligence? Although present machine intelligence paradigms only addresses fault tolerance from "an emergent property" perspective, it is possible that fault-tolerance was used a ground-rule for evolutionary development of biological systems, and may lead to yet unknown computational architectures.

Hence, for the short-term, the issue of metrics for fault-tolerance appear to be relevant for autonomous systems in the context of performing tasks in harsh environments and where mechanical and information resource tolerance are important. The tolerance can be graded according to the task and the ability of the machine to complete it in the presence of faults.

8 Consciousness and control

The debate over the neural correlate of consciousness is obviously of most interest to neuromorphic engineering. Our present poor understanding of the underlying neural circuits does not imply that it is not a necessity for autonomous machines. The complex interactions between awareness, planning and survival dictates equipping machines with some level of the "self". The level may be primitive at first. Practical awareness can address sensory representation of the environment and its representations in terms of goals and necessities to survival (e.g. battery charging). The competition of sensory on motor behaviour will need to address priorities and dynamic reward. The representations that emerge from this computation will represent primitive forms of awareness that machines will be capable of processing, but not necessarily of realizing. Realization may emerge as a balanced competition between motor plans, behaviour and reward obtained from behaviour. Hence the development of task-oriented neuromorphic systems will allow the exploration of computational structures and information processing paradigms that can embed such a competition.

In the context of intelligence metrics, the question of consciousness can be stated as that of resource management. The development of a metric framework will have to account for a broad spectrum of sensory, motor and reward situations that could be too complex to represent. One can envision a metric that measures final outcomes based on the essence of task completion measured in terms of energy and survival. That is to be, and to be there in the right time.

9 Complexity, Hierarchy, Regularity and Modularity

The development of design methodologies for highly complex integrated circuits containing tens of million of transistors have taught engineers a number of golden rules in the management of complexity: Hierarchy, regularity and modularity. These human-made engineering rules are similar to the rules that underlie biological systems structures. Representation and learning efficiency (in particular online continuous learning) are the most concerned and affected by the hierarchy, regularity and modularity (HRM) of the underlying structures. The issue of whether HRM issues are relevant to intelligence metrics is similar to those discussed earlier in the context of fault tolerance and representation (e.g. sparseness). HRM of computational structures may affect the optimality of an autonomous system, but may not be critical to its successful operation. Again, the importance of HRM as a ground-rule for autonomy may go beyond optimality and may be critical to the scalability of the architecture and representations. Scalability is relative to the initial conditions and desired bounds. From a practical point of view, it is evident that a HRM-based design will be superior to a design that is flat and that lacks modularity and regularity.

10 Summary and Conclusions

The issues discussed in this position paper converge to the conclusion that in the context of autonomous systems, intelligence metrics should be task oriented and should embed factors such as completion, resources and energy. Completion is easy to assess, with the distinction that it is for practical systems and not for simulations. Resources and energy could be cast to specific implementation, whether software or hardware. Resources will cover aspects of resources used to perform a task, and those available to capture the skills to perform the task. The energy measure will represent the total energy required to perform a task and can easily be measure for software and hardware implementations.

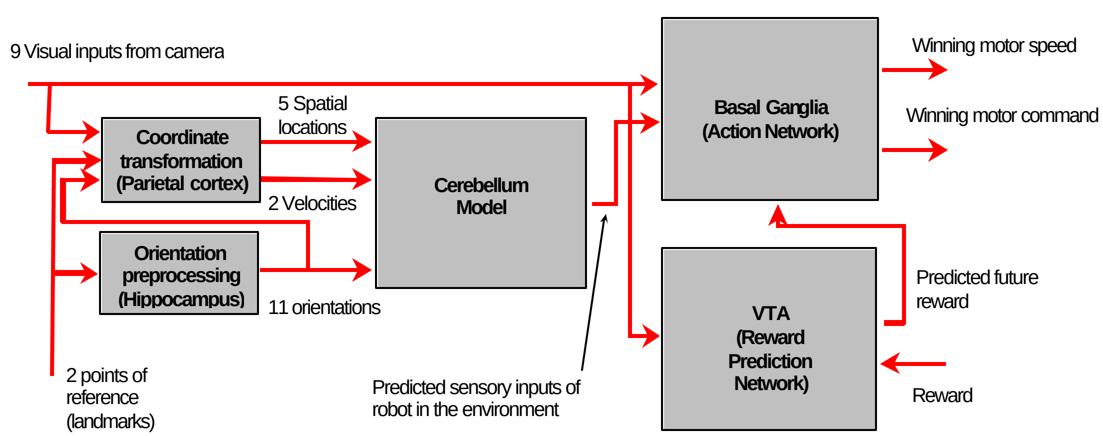


Figure 1. Sensorimotor system implementing anticipation and reinforcement learning allowing a Khepera robot to track (by rotating in place) a moving target.

11 References

Albus, J. S. (1971). “A theory of cerebellar function.” Math. Biosci **10**: 25-61.

Bell, A. J. and T. Sejnowski (1995). “An information-maximisation approach to blind separation and bling deconvolution.” Neural Computation **7**: 1129-1159.

Coenen, O. J.-M. D. (1998). Modeling the Vestibulo-Ocular Reflex and the Cerebellum: Analytical & Computational Approaches. Physics Department, University of California, San Diego.

Coenen, O. J.-M. D. and T. J. Sejnowski (1996). Learning to make predictions in the cerebellum may explain the anticipatory modulation of the vestibulo-ocular reflex (VOR) gain with vergence. Proc. of the 3rd Joint Symposium on Neural Computation, Institute of Neural Computation, University of California, San Diego, and California Institute of Technology.

Etienne-Cummings, R., J. V. d. Spiegel, et al. (2000). “A Foveated Silicon Retina for Two-Dimensional Tracking.” IEEE Trans. Circuits and Systems II **47**(6).

Georgopoulos, A. (1995). Motor Cortex and Cognitive Processing. The Cognitive Neurosciences. M. Gazzaniga, MIT Press**:** 507-518.

Horiuchi, T. and C. Koch (1999). “Analog VLSI-based Modeling of the Primate Oculomotor System.” Neural Computation Journal **11**(1): 243-265.

Jabri, M., O. J.-M. D. Coenen, et al. (1997). Sensorimotor integration and control. Extended Abstracts of the NIPS*97 Workshop: Can Artificial Models Compete to Control Robots?, Denver.

Jabri, M. A., R. Coggins, et al. (1996). Adaptive Analog Neural Systems, Chapman and Hall, UK.

Lewis, T., R. Etienne-Cummings, et al. (2000). Towards

Biomorphic Control Using aVLSI CPG Chips. IEEE ICRA, San Francisco, IEEE Press.

Marr, D. (1969). “A theory of cerebellar cortex.” J. Physiol **202**: 437-470.

Mead, C. (1989). Analog VLSI and Neural Systems. Reading Massachusetts, Addison-Wesley.

Schaik, A. v. (2000). “An Analog VLSI Model of Periodicity Extraction in the Human Auditory System.” *Analog Integrated Circuits and Signal Processing*, Kluwer Academic Publishers **March**.

Sutton, R. S. and A. G. Barto (1981). “Towards a modern theory of adaptive networks: Expectation and prediction.” Psy. Review **88**(2): 135-170.

Tilden, M. W. (1994). “"Living Machines".” European Journal of Autonomous Systems.

Grading Intelligence in Machines: Lessons from Animal Intelligence

Subhash Kak
Department of Electrical & Computer Engineering
Louisiana State University
Baton Rouge, LA 70803-5901; kak@ee.lsu.edu

July 20, 2000

In this note I argue that to find a Vector of Intelligence (VI) for a performance metric for machines, it is helpful to look at animal intelligence, which is clearly defined as a spectrum.

All animals are not equally intelligent at all tasks; here intelligence refes to performance of various tasks, and this performance may depend crucially on the animal's normal behavior. It may be argued that all animals are sufficiently intelligent because they survive in their ecological environment. Nevertheless, even in cognitive tasks of the kind normally associated with human intelligence animals may perform adequately. Thus rats might find their way through a maze, or dolphins may be given logical problems to solve, or the problems might involve some kind of generalization. These performances could, in principle, be used to define a gradation.

If we take the question of AI programs, it may be argued that the objectives of each define a specific problem solving ability, and in this sense AI programs constitute elements in a spectrum. But we think that it would be useful if the question of gradation of intelligence were to be addressed in a systematic fashion. The question is best examined in an ecological context; a similar case for an ecological study of machine vision has been made by Gibson.

The issues that we leave out are those related to defining consciousness and quantum approaches to brain processes and intelligence. Although I have personally worked on these issues, I believe they lie outside the scope of the NIST Conference on Performance Metrics for Intelligent Systems.

On Animal Intelligence

According to Descartes, animal behavior is a series of unthinking mechanical responses. Such behavior is an automatic response to stimuli that originate in the animal's internal or external environments. In this view, complex behavior can always be reduced to a configuration of reflexes where thought plays no role. According to Descartes only humans are capable of thought since only they have the capacity to learn language.

Recent investigations of nonhuman animal intelligence not only contradict Cartesian ideas, but also present fascinating riddles. It had long been thought that the cognitive capacities of the humans were to be credited in part to the mediating role of the inner linguistic discourse. Terrace Te85 claims that animals do think but cannot master language, so the question arises as to how thinking can be done without language:

> Recent attempts to teach apes rudimentary grammatical skills have produced negative results. The basic obstacle appears to be at the level of the individual symbol which, for apes, functions only as a demand. Evidence is lacking that apes can use symbols as names, that is, as a means of simply transmitting information. Even though non-human animals lack linguistic competence, much evidence has recently accumulated that a variety of animals can represent particular features of their environment. What then is the non-verbal nature of animal representations?...[For example] learning to produce a particular sequence of four elements (colours), pigeons also acquire knowledge about a relation between non-adjacent elements and about the ordinal position of a particular element. ([6], page 113)

Clearly the performance of animals points to representation of whole patterns that involves discrimination at a variety of levels. But if conceptualization is seen as a result of evolution, it is not necessary that this would have developed in exactly the same manner for all species. Other animals learn concepts nonverbally, so it is hard for humans, as verbal animals, to determine their concepts. It is for this reason that the pigeon has become a favourite with intelligence tests; like humans, it has a highly developed visual system, and we are therefore likely to employ similar cognitive categories. It is to be noted that pigeons and other animals are made to respond in extremely unnatural conditions in Skinner boxes of various kinds. The abilities elicited in research must be taken to be merely suggestive of the intelligence of the animal, and not the limits of it.

In an ingenious series of experiments Herrnstein and Loveland He64 were able to elicit responses about concept learning from pigeons. In another experiment Herrnstein He85 presented 80 photographic slides of natural scenes to pigeons who were accustomed to pecking at a switch for brief access to feed. The scenes were comparable but half contained trees and the rest did not. The tree photographs had full views of single and multiple trees as well as obscure and distant views of a variety of types. The slides were shown in no particular order and the pigeons were rewarded with food if they pecked at the switch in response to a tree slide; otherwise nothing was done. Even before all the slides had been shown the pigeons were able to discriminate between the tree and the non-tree slides. To confirm that this ability, impossible for any machine to match, was not somehow learnt through the long process of evolution and hardwired into the brain of the pigeons, another experiment was designed to check the discriminating ability of pigeons with respect to fish and non-fish scenes and once again the birds had no problem doing so. Over the years it has been shown that pigeons can also distinguish: (1) oak leaves from leaves of other trees, (ii) scenes with or without bodies of water, (iii) pictures showing a particular person from others with no people or different individuals.

Herrnstein He85 summarizes the evidence thus:

> Pigeons and other animals can categorize photographs or drawings as complex as those encountered in ordinary human experience. The fundamental riddle posed by natural categorization is how organisms devoid of language, and presumably also of the associated higher cognitive capacities, can rapidly extract abstract invariances for some (but not all) stimulus classes containing instances so variable that we cannot physically describe either the class rule or the instances, let alone account for the underlying capacity.

Amongst other examples of animal intelligence are mynah birds who can recognize trees or people in pictures, and signal their identification by vocal utterances—words—instead of pecking at buttons Tu82, and a parrot who can answer, vocally, questions about shapes and colors of objects, even those not seen before Pe83.

Another recent summary of this research is that of Wasserman Wa95:

> [Experiments] support the conclusion that conceptualization is not unique to human beings. Neither having a human brain nor being able to use language is therefore a precondition for cognition... Complete understanding of neural activity and function must encompass the marvelous abilities of brains other than our own. If it is the business of brains to think and to learn, it should be the business of behavioral neuroscience to provide a full account of that thinking and learning in all animals—human and nonhuman alike.

Gradation of Intelligence

An extremely important insight from experiments of animal intelligence is that one can attempt to define different gradations of cognitive function. It is obvious that animals are not as intelligent as humans; likewise, certain animals appear to be more intelligent than others. For example, pigeons did poorly at picking a pattern against two other identical ones, as in picking an A against two B's. This is a very simple task for humans. Herrnstein He85 describes how they seemed to do badly at certain tasks:

- Pigeons did not do well at the categorization of certain man-made and three-dimensional objects.
- Pigeons seem to require more information than humans for constructing a three-dimensional image from a plane representation.
- Pigeons seem to have difficulty in dealing with problems involving classes of classes. Thus they do not do very well with the isolation of a relationship among variables, as against a representation of a set of exemplars.

In a later experiment Herrnstein et al. He89 trained pigeons to follow an abstract relational rule by pecking at patterns in which one object was inside, rather than outside of a closed linear figure. Wasserman Wa93,Wa95 devised an experiment to show that pigeons could be induced to amalgamate two basic categories into one broader category not defined by any obvious perceptual features. The birds were trained to sort slides into two arbitrary categories, such as category of cars and people and the category of chairs and flowers. In the second part of this experiment, the pigeons were trained to reassign one of the stimulus classes in each category to a new response key. Next, they were tested to see whether they would generalize the reassignment to the stimulus class withheld during reassignment training. It was found that the average score was 87 percent in the case of stimuli that had been reassigned and 72 percent in the case of stimuli that had not been reassigned. This performance, exceeding the level of chance, indicated that perceptually disparate stimuli had amalgamated into a new category. A similar experiment was performed on preschool children. The children's score was 99 percent for stimuli that had been reassigned and 80 percent for stimuli that had not been reassigned. In other words, the children's performance was roughly comparable to that of pigeons. Clearly, the performance of adult humans at this task will be superior to that of children or pigeons.

Another interesting experiment related to the abstract concept of sameness. Pigeons were trained to distinguish between arrays composed of a single, repeating icon and arrays composed of 16 different icons chosen out of a library of 32 icons Wa95. During training each bird encountered only 16 of the 32 icons; during testing it was presented with arrays made up of the remaining 16 icons. The average score for training stimuli was 83 percent and the average score for testing stimuli was 71 percent. These figures show that an abstract concept not related to the actual associations learnt during training had been internalized by the pigeon. And the performance of the pigeons was clearly much worse than what one would expect from humans.

Animal intelligence experiments suggest that one can speak of different styles of solving AI problems. Are the cognitive capabilities of pigeons limited because their style has fundamental limitations? Can the relatively low scores on the sameness test for pigeons be explained on the basis of wide variability in performance for individual pigeons and the unnatural conditions in which the experiments are performed? Is the cognitive style of all animals similar and the differences in their cognitive capabilities arise from the differences in the sizes of their mental hardware? And since current machines do not, and cannot, use inner representations, is it right to conclude that their performance can never match that of animals?

Another issue is whether one can define a hierarchy of computational tasks that would lead to varying levels of intelligence. These tasks could be the goals defined in a sequence, or perhaps a lattice, that could be set for AI research. If the simplest of these tasks proved intractable for the most powerful of computers then the verdict would be clear that computers are designed based on principles that are deficient compared to the style at the basis of animal intelligence.

Recursive Nature of Animal Behavior

A useful perspective on animal behavior is its recursive nature, or part-whole hierarchy. Considering this from the bottom up, animal societies have been viewed as "superorganisms". For example, the ants in an ant colony may be compared to cells, their castes to tissues and organs, the queen and her drones to the generative system, and the exchange of liquid food amongst the colony members to the circulation of blood and lymph. Furthermore, corresponding to morphogenesis in organisms the ant colony has sociogenesis, which consists of the processes by which the individuals undergo changes in caste and behavior. Such recursion has been viewed all the way up to the earth itself seen as a living entity. Parenthetically, it may be asked whether the earth itself, as a living but unconscious organism, may not be viewed like the unconscious brain. Paralleling this recursion is the individual who can be viewed as a collection of several "agents" where these agents have sub-agents which are the sensory mechanisms and so on.

Logical tasks are easy for machines whereas AI tasks are hard. It might well be that something fundamental will be gained in building machines that have recursively defined behavior in the manner of life. But how such machines could be designed is not at all clear.

A hierarchy of intelligence levels can be useful also in the classification of animal behavior. There does not appear to be any reason that experiments to check for intelligent behavior at different levels could not be devised. Furthermore, experiments could be conducted to determine the difference in ability for individual animals. That such experiments have not been described until now is merely a reflection of the peculiar history of the field.

Concluding Remarks

Study of animal intelligence provides us with new perspectives that are useful in representing the performance of machines. For example, the fact that pigeons learn the concept of sameness shows that this could not be a result of associative response to certain learnt patterns. If evolution has led to the development of specialized cognitive circuits in the brain to perform such processing, then one might wish to endow AI machines with similar circuits. Other questions arise: Is there a set of abstract processors that would explain animal performance? If such a set can be defined, is it unique, or do different animal species represent collections of different kinds of abstract processing that makes each animal come to achieve a unique set of conceptualizations?

Animal behavior ought to be used as a model to define a hierarchy of intelligence tasks. This hierarchy is likely to be multidimensional. Various kinds of intelligence tasks could define benchmark problems that would represent the various gradations of intelligence.

Should VI reflect the degree of recursion in the organization of the intelligence in the machine? Given that the neural organization of the brain consists of "networks of networks", it appears that this be so. On similar grounds, one may assert that the performance of the machine should span several scales. The relative scale invariance of the performance will be a measure of the "quality" of the intelligence.

References

[1] Gibson, J.J. (1979). *The Ecological Approach to Visual Perception.* Houghton Mifflin, Boston.

[2] Herrnstein, R.J., Loveland, D.H. (1964). "Complex visual concept in the pigeon." *Science* 146, 549-551.

[3] Herrnstein, R.J. (1985). "Riddles of natural categorization." *Phil. Trans. R. Soc. Lond.* B 308, 129-144.

[4] Herrnstein, R.J., W. Vaughan, Jr., D.B. Mumford, and S.M. Kosslyn. (1989). "Teaching pigeons an abstract relational rule: insideness." *Perception and Psychophysics* 46, 56-64.

[5] Kak, S. (1996) "Can we define levels of artificial intelligence?" Journal of Intelligent Systems, vol. 6, 133-144.

[6] Terrace, H.S. (1985). "Animal cognition: thinking without language." *Phil. Trans. R. Soc. Lond.* B 308, 113-128.

[7] Wasserman, E.A. (1993). "Comparative cognition: Beginning the second century of the study of animal intelligence." *Psychological Bulletin*, 113, 211-228.

[8] Wasserman, E.A. (1995). "The conceptual abilities of pigeons." *American Scientist*, 83, 246-255.

Biometric Techniques: The Fundamentals of Evaluation

Thomas A. Chmielewski and Paul R. Kalata
Drexel University
Philadelphia, PA 19104
tchmiele@ cbis.ece.drexel.edu, kalata@cbis.ece.drexel.edu

ABSTRACT

While the term biometrics may connote high technology, it simply stands for the concept of recognizing a human being. Today, we use technology to automate the measurements of physical or behavioral characteristic of an individual, so that these measurements may be compared against previously stored data to authenticate an individual's claimed identify. Current biometric systems use diverse measurements, technology and algorithms making it difficult to compare their performance on an equal basis. In general, a candidate biometric system needs to be accessed against the performance requirements of an application. The biometric community uses statistical measures to define the performance of systems. The objective of this paper is to provide a brief tutorial on biometrics and the current measures used to define performance in order to provide information to new biometric users and to stimulate the research community in this rich problem.

Keywords: *acceptance threshold, biometrics, confidence intervals, false accept rate, false reject rate, FAA, FAR, d', performance, pdf, receiver operating curves, ROC*

1. Biometrics: A Brief Tutorial

Humans identify one another by the way faces look and can sometimes identify individuals at a distance based on their stature and gait. Over the years, inventive humans extended their innate identification capability by applying engineering techniques to allow identification of individuals without having the need for someone who explicitly knew the individual. Specifically, the identification problem was solved by relating a physical entity or "secret" information to a person by using:

- something you "**have**" such as a card or key,
- something you "**know**" such as a password or personal identification number (PIN).

Either alone or in combination, possession and knowledge can enable the use of technology to identify a person. ATMs for example require the use of the ATM card (have) plus a PIN (know) to gain access. Since using possession and knowledge for identification purposes cannot distinguish between the correct person and a potential impostor who acquired the possession/knowledge, there is clearly a need for a higher level of positive personal identification.

It is possible to eliminate the aspect of possession and/or knowledge and rely rather on something that the person "**is**", specifically a physiological or behavioral characteristic that can be easily detected, that is time invariant, and that is significantly different across the population of people who will be identified by it. The term *biometric* is used to describe these characteristics which allow identification of an individual. The key advantage of using biometric data to identify a person is that the biometric cannot be stolen, misplaced or forgotten because it is something that the person "**is**", as contrasted to "possession" and/or "knowledge".

Biometric Systems use technology to automate the measurements of physical or behavioral characteristic of an individual, so that these samples may be compared against *previously stored data* to determine if significant similarities exist in order to confirm the samples sufficiently match the stored data hence confirming or denying the individual's identity. In essence, biometric identification is a pattern recognition problem. In order to allow good decisions to be made, we would like maximum variations across individuals, but minimum variation for any given person across time or environmental conditions.

There are two types of problems that a biometric system must handle namely:

- Verification Problem (authentication): confirming or denying a person's claimed identity (Am I who I claim I am ?); this is a one to one matching process.
- Recognition Problem (identification): establishing a person's identity from a set of stored identities; this is a one to many matching process.

Biometrics currently in commercial use for either identification or recognition include: fingerprints, hand geometry, handwritten signatures, voiceprints, face, iris, retinal patterns and thermograms [1]. Certain, physical characteristics such as fingerprints and iris texture, are considered to be "invariant". Behavioral characteristics such as voice and signature. are considered to be "somewhat variable" since they are influenced by physical and emotional conditions and evolve over time.

The retina, the iris and fingerprints are considered truly unique and provide the greatest precision for biometrics [9]. However, other biometrics should not be dismissed, since each biometric provides unique advantages which can be exploited by selecting the correct biometric for the correct application. For instance, INSPASS (Immigration and Naturalization Service Passenger Accelerated Service System) uses a hand geometry system to quickly verify the identity of arriving passengers in speeding up international arrivals at certain North American international airports. People enrolled in INSPASS are given a magstripe card encoded with appropriate data for their hand geometry. Upon arrival, INSPASS travelers swipe their card (**have**), place their hand in a reader (**are**), and then proceed to the customs gate. Coupling the "have" and "are" makes hand geometry a good solution for this application (even though this biometric is not as unique as others) since it is cost effective, readily accepted by most users and exhibits low failure rates in acquiring data.

2. System Functionality

Before a biometric system can operate, a quality sample(s) of the biometric signal, such as a fingerprint, the image of an iris, or speech from users of the system must be obtained. This is called the *enrollment process*. Enrollment usually involves an operator who coaches the users to provide the best biometric input. These inputs are processed and stored as templates or feature vectors that contain the pertinent information used for later biometric data comparison. Additionally, the person's identity in the form of an "ID number" or some data structure is associated with the template. Enrollment should be done under the best of conditions since the quality of data stored during enrollment effects the performance of the system.

Figure 1 illustrates a verification system block diagram.

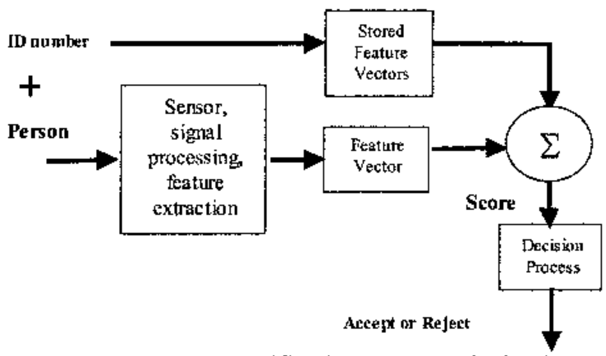


Figure 1. Biometric Verification System Block Diagram

Here it is assumed that a number of individuals have been enrolled in a data base and have been given an " ID number" (such as a bank account). When a person uses the system, their ID number (have) is used to reference a stored feature vector or template in a database. A sensor obtains a biometric sample from the person and then extracts the relevant features from the biometric data into a feature vector. A comparison of the stored feature vector and the computed feature vector is made (*one-to-one process*) generating a matching score. The score is then passed to a decision process. The results of the decision process are acceptance or rejection of the premise that the person at the device is the same as the one who originally generated the feature vector during the enrollment process as referenced by the ID number. Verification acceptance or rejection is based on comparing the matching score to a decision threshold defined by *a priori* statistic of system performance and the application.

The recognition process is significantly different. No ID number is input and the system must compare the feature vector of the person at the device against all stored feature vectors (*a one-to-many process*). If a match is found, then an ID number (and/or other data structure) is retrieved and the person is identified and coupled to this data allowing them access to a building, bank account, etc.

Consider the possible resulting outcomes for a system if a person walks up and attempts to be verified (or identified). There are two possible descriptions of the user: he/she is the correct person (and should be given access) or he/she is an impostor (trying to gain access). Ignoring the case where the biometric system chooses to make no decision, there are two possible outcomes (pass or fail), generating four possible conditions. Table 1 shows these outcomes and conditions: either the correct result occurs, or the system falsely rejects the authentic or falsely accepts the impostor. In the case that the system chooses to make no decision, the user may be given another chance. The performance of a biometric or biometric system is measured by the frequency of false accepts and false rejects.

Table 1. Results of Verification or Identification for a Biometric System User

User	Pass	Fail
Authentic	correct accept-allow access	false reject - refuse access
Impostor	false accept-allow access	Correct reject-refuse access

3. Performance

The quantitative measure of the performance of a biometric verification system is defined by the frequency of false accepts and false rejects [1,12]. These probabilities define how correctly the biometric returns a matching score when a correct individual (authentic) or an incorrect individual (impostor) is presented to the system. While other performance metrics such as the speed of operation, number of templates capable of being stored, and cost are important they are not the focus of this paper.

While we would like to have ideal performance for every biometric (i.e. no False Accepts and no False Rejects), this is unachievable when we consider real world factors such as noise, environmental conditions or the actual discriminating capability of the biometric itself. While not perfect, biometrics are used in many successful applications. Generally, the application dictates the required performance of a biometric. Banks may be willing to accept a certain level of False Accepts but no False Rejects at an ATM in order to keep their customers happy. Alternatively, access to a highly secure facility may not allow any False Accepts but allow False Rejects, since real authentics would be willing to try the multiple times necessary to gain access. A Cost Functional (probability of the decision times a "cost") may also be used to best determine the False Accept versus False Reject trade-off.

3.1 Population Issues: Failure-to-Enroll (FTE)

Even though we would like a biometric to be universally applicable across all users in a given population, there may be some people who cannot use the system due to abnormalities, diseases, injuries, accidents, or degradation of the biometric signal due to their occupation. For instance, masonry workers may wear down their fingerprints so as to make this biometric unreliable. The subset of a population who cannot use a particular biometric are called *outliers*.

Outliers will generally not be able to enroll in a system due to the absence of the biometric or a signal that cannot be converted into a biometric template due to such factors as signal strength, missing characteristics or characteristics present in their biometric not considered by the system. The performance of outliers are categorized by a *failure-to-enroll* (FTE) rate. Numbers to bound the FTE for a biometric can be estimated from the frequency of abnormalities, permanent injuries, or permanent diseases in a given population (or for the world by geographic location) that prevent use of the characteristic. When testing any population, FTE should be accumulated and analyzed.

3.2 System Issues: Failure-to-Acquire (FTA)

Failure-to-Acquire (FTA) is defined as the failure of a biometric system to capture information prior to the extraction of biometric data for the feature vector. This failure may occur during the enrollment process, the verification process or the identification process. It is dependent on the ancilliary processes, human factors and external disturbances that may affect the sensor used to acquire the biometric sample. Factors that cause FTAs include: user distraction, or acute injuries or diseases that prevent acquisition of the biometric signal. During testing, FTAs should be accumulated and included in the false reject rate computation (for persons previously enrolled) especially if the entire system performance is being considered rather than just the raw biometrics' performance.

3.3 False Accept Rates and False Reject Rates (FAR, FRR)

Biometric Systems suppliers typically use FAR (False Accept Rate) together with FRR (False Reject Rate) to describe the capabilities of their system (Table 1). FRR is the error rate at which a true authentic (i.e., an individual claiming to be who they actually are) is rejected by the system. FAR is the error rate at which a false authentic (i.e., an individual claiming to be who they are not, i.e., an impostor) is falsely allowed to use the system. FRR and FAR are interrelated by statistics and are dependent on the acceptance or decision threshold of the biometric system under consideration. Being more liberal in the acceptance criterion (a lower threshold), will generally allow more people (both authentics and impostors) into the system. Conversely, being stricter in the acceptance criterion will reject more people (both authentics and impostors). The setting of the threshold is dictated by the requirements of the application.

The relationship between FAR and FRR is easily understood by plotting a distribution of the matching scores of authentics and impostors on the same graph. Sometimes referred to as Authentics-Impostor Distribution Curve or a Performance Histogram, this graph shows the distribution of the population versus a given score of the biometric. The match (or mismatch) score (how well the template matched the biometric sample) is plotted on the horizontal axis and the population frequency on the vertical axis. For statistical analysis, the curves can be normalized so that the area under each curve is one. When normalized in this way the curves become probability distribution functions and may be used to compute probabilities or error rates. Figure 2 shows a typical set of curves. As can be seen the authentics curve (left) and impostor curve (right) overlap. Note that these curves could be interchanged based on the meaning of the horizontal axis. For Figure 2, the better the match between a user and the stored template, the lower the value, with zero (a perfect match) at the far left. The setting of the decision threshold (shown by the vertical bar) defines both the false accept rate (FAR) and false reject rate (FRR). FAR is the area under the impostor curve to the left of the decision threshold. FRR is the area under the authentics curve to the right of the decision threshold. From Figure 2, it is clear that FAR and FRR are interrelated, one cannot specify each value independently!

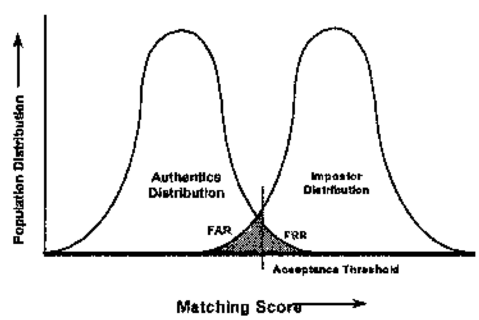


Figure 2. Hypothetical Authentic and Impostor Curves

Impostor and Authentic histograms are different for each biometric. In fact, the actual curves for a given type of biometric may be platform, sensor or algorithm dependent. Obtaining histograms which have sufficient data to perform a reliable curve fit for analysis of their tails requires large amounts of data. The impostor and authentics histograms are usually generated off line by using multiple samples of a closed set of individuals. Hence, for a set of N individuals there will be some number of samples m_N for each individual. By verifying each sample against each other for a given individual an individuals authentics histogram can be generated. The authentics curve for the tested population is obtained by combining all the individual authentics data. Verifying each sample for a given individual against all other individual's samples produces the impostor histogram. Hence, the authentic histogram is a combination ofN people having m_N samples each while the impostor curve is the combination of the verification of each sample for a given person against all other persons' samples.

Performance curves are not necessarily Gaussian. Generally, there is a binning process associated with these performance curves. Binned histograms may show no overlap, incorrectly indicating separation between authentics and impostors. Fitting theoretical curves to the measured data allows computation of FAR/FRR for various thresholds and produces the theoretical tails of the performance curves. Without curve fitting, the granularity of the data bins may not provide sufficient resolution for computation of the FAR/FRR as the decision threshold changes or at extreme points. Fit data may provide more conservative (or more liberal) values than are actually descriptive of the system. It is important to exercise caution with curve fitting since the results may not accurately model the system (especially in the tails) leading to inaccurate probability computations.

Further modifying the shape of performance curves are "goats" and "sheep". Goats are users who consistently return large distance measures when new samples are compared to their enrolled templates. Sheep, generally a larger part of the population, return small distance measures compared to their enrolled data. The sheep (small variance) and goat (large variance) performance can cause the histogram to exhibit bimodal authentics plots with the goats associated with a smaller secondary mode [1].

3.4 An Empirical Bound on FAR

Consider a biometric that may not be capable of clearly distinguishing characteristics of identical twins (possibly a face biometric). One can put the following bound on the FAR due to identical twins [1]. Statistically, 1 in 80 births are twins and about 1/3 of twins are identical (monozygotic). If we consider 240 births, there are 243 individual and one pair of them are identical twins. The chance of selecting a person at random who has an identical twin is roughly 2/243 = 0.82%. With the assumption that a particular biometric cannot distinguish between identical twins, one can define the minimum False Accept rate at 0.82% due just to the birth of identical twins. Besides user cooperation or technical factors, fraternal twins and parent/offspring may also share the same biometrics value (consider that many fraternal twins still look alike). Putting a number on this contribution to FAR is significantly more difficult.

4.0 The Underlying Probability for Decisions

In biometric identification, a decision to the authentic or impostor must be made based on noisy measurements. To this extent, one must understand the probabilistic nature of the measurement, how it is processed to make a decision and the performance of the decision itself.

Consider a noisy biometric random variable, x, characterized by its probability density function (pdf), $f_X(x)$ with properties [6,7]:

$$f_X(x) \geq 0, \quad \int f_X(x)\,dx = 1.$$

Common pdf shapes include: Gaussian, Uniform, Exponential, Binomial and Poisson. Two important parameters which describe the pdf are the mean and standard deviation (the square root of the variance):

$$\text{mean:} \quad \bar{x} = \int x\, f_X(x)\,dx$$

$$\text{variance:} \quad \sigma_X^2 = \int (x-\bar{x})^2 f_X(x)\,dx.$$

The mean, $\bar{x}$, is "where" the pdf mass is concentrated and the standard deviation, σ_X, is the "spread" of the mass about the mean as illustrated by the triangular pdf in Figure 3.

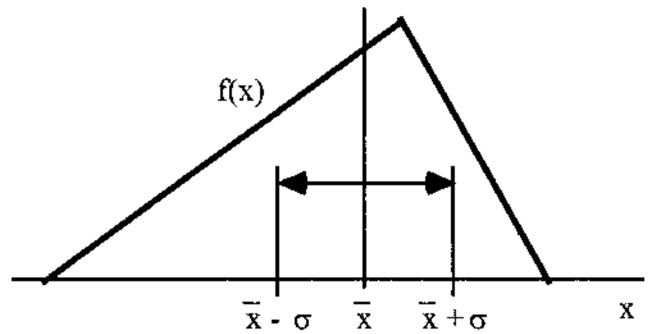


Figure 3. Probability density function

4.1 Biometric Measurement

Consider noisy measurements (or score) of authentic and impostor alternatives $\{m_a, m_i\}$ of $\{x_a, x_i\}$ with zero mean measurement noise $\{e_a, e_i\}$ as illustrated by Figure 4.

authentic measurement: $m_a = x_a + e_a$

impostor measurement: $m_i = x_i + e_i$

The variability of the authentic and impostor measurements invariably overlap each other as illustrated by Figure 4.

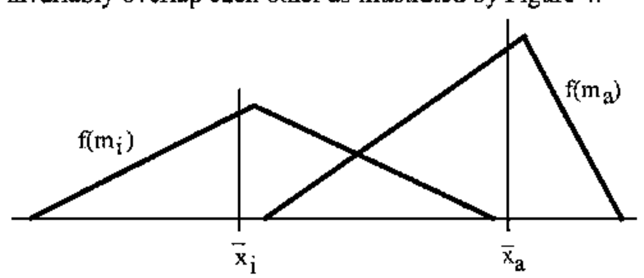


Figure 4. Authentic and impostor measurements

4.2 Biometric Decision

For our analysis, high valued measurements imply an authentic biometric source and low valued measurements imply an impostor source. The decision design [7] is to select a measurement threshold, m_{th}: for measurements exceeding the threshold, decide an authentic source (d_a); alternatively, if the measurement is less than the threshold, decide an impostor source (d_i) as illustrated by Figure 5, i.e.,

$$d_a: \text{ if } m \geq m_{th} \quad d_i: \text{ if } m < m_{th}$$

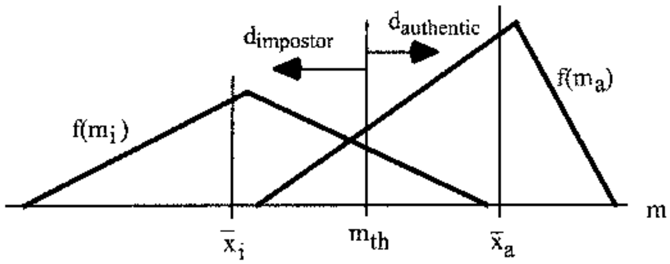


Figure 5. Authentic or impostor threshold decision

For the binary source/decision process there are 4 possible outcomes as illustrated by Table 2 in that:

- an impostor decision with an impostor source results in a correct rejection (dismissal),
- an impostor decision with an authentic source results in a false rejection (mis-detection),
- an authentic decision with an impostor source results in a false acceptance (false alarm), and
- an authentic decision with an authentic source results in a correct acceptance (detection).

The terms {dismissal, mis-detection, false alarm, detection} are common in decision theory and can be interchanged with {correct rejection, false rejection, false acceptance, correct acceptance} respectively as in Table 2.

Table 2. Binary Source/Decision Outcome

Decision	**Source**	
	authentic, x_a	**impostor, x_i**
authentic d_a	correct accept (detect)	false accept (false alarm)
impostor d_i	false reject (mis detect)	correct reject (dismissal)

In virtually all decision cases, no matter where the threshold is set, there will be correct decision and there will be incorrect decisions. The decision design problem is to select the threshold to maximize the (weighted) correct decisions and minimize the (weighted) incorrect decisions.

4.3 Biometric Performance

The decision process performance is determined by the authentic and impostor probabilities which are determined by areas under the pdf's depending. In particular:

Probability of a correct acceptance, $P(d_a|x_a)$, is the area under $f(m_a)$ for $m \geq m_{th}$,

Probability of a false acceptance, $P(d_a|x_i)$, is the area under $f(m_i)$ for $m \geq m_{th}$,

Probability of a correct rejection, $P(d_i|x_i)$, is the area under $f(m_i)$ for $m < m_{th}$, and

Probability of a false rejection, $P(d_i|x_a)$, is the area under $f(m_a)$ for $m < m_{th}$.

Figures 6a, 6b and 6c illustrate these decision probabilities.

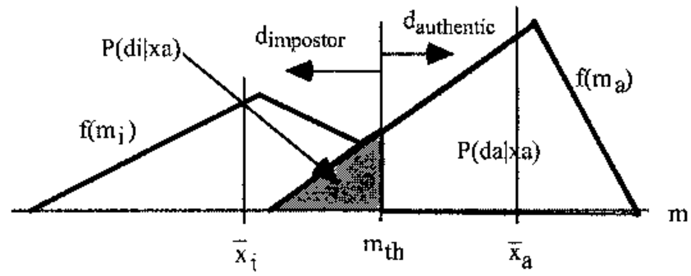


Figure 6a. Probability of Correct Accept and False Reject

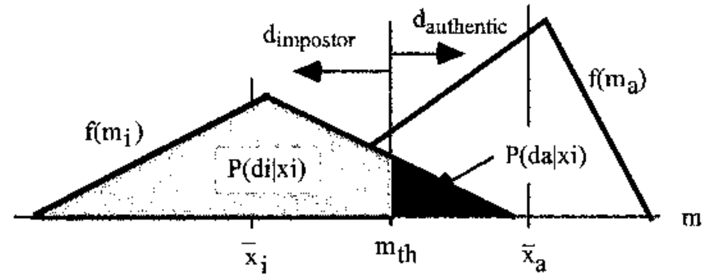


Figure 6b. Probability of Correct Rejection and False Accept

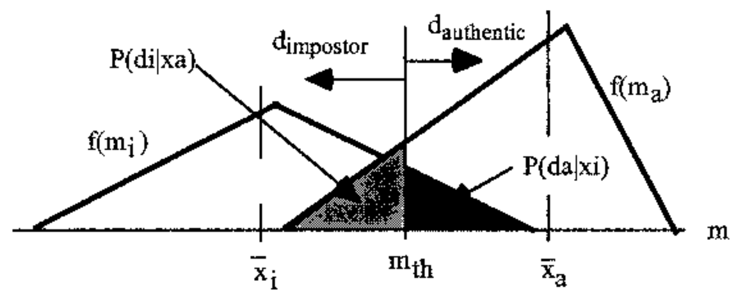


Figure 6c. Probability of False Reject and False Accept

4.4 Cross-Over-Error-Rate (CER)

The value at which the FAR equals the FRR defines the *cross-over-error-rate* (CER) or *equal-error-rate* (EER). The CER may be correlated to the decision threshold that allows this equality. CER provides one method of comparing biometric performance since it is a characteristics of the set of histograms and predefines the threshold setting.

4.5 ROC curves

Another convenient ways to compare the decision process is with "Receiver Operation Characteristic" (ROC) curves which illustrates performance probabilities generated by varying the threshold decision. Figure 7a illustrates the ROC-P(D)/P(FA) performance: Probability of detection (correct acceptance) verses the Probability of false accept. Varying the threshold decision, we can improve in the detection probability, but we also increase the false accept probability. The biometric performance objective is to make this curve as convex to the left as possible and then select which point on this curve is acceptable for biometric identification operation.

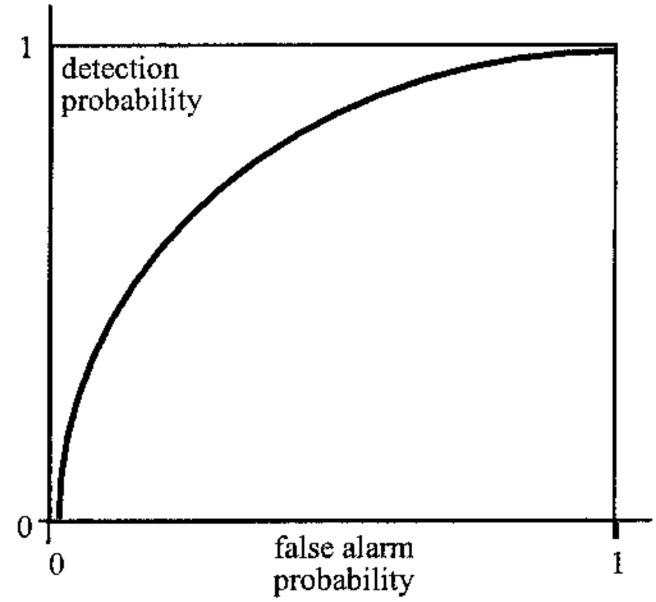


Figure 7a. ROC : P(detection) verses P(false accept)

Another ROC curve is illustrated by Figure 7b, the ROC-P(FR)/P(FA) performance: Probability of false rejects verses the Probability of false accept. This is the ROC curve normally used in biometric literature. Each point on the curve corresponds to a decision threshold and the corresponding FRR and FAR may be easily seen. Varying the threshold decision, we can decrease the false reject probability, but we also increase the false accept probability. The biometric performance objective is to make this curve as concave to the left as possible and then to pick which point on this curve is acceptable for biometric identification operation. As illustrated in Figure 7b the CER (where FAR = FRR) can be easily found from a ROC curve.

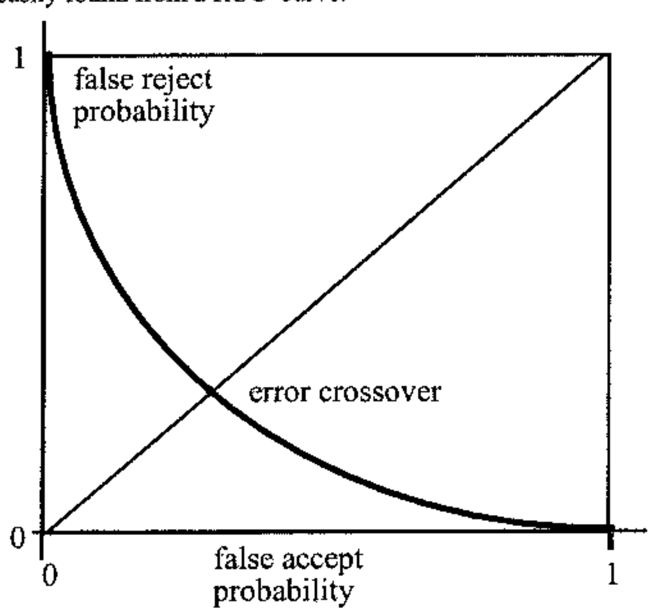


Figure 7b. ROC: P(false reject) verses P(false accept)

4.6 Biometric Confidence

The generation of the biometric pdf's as described above must be made by acquiring test data from the biometric system. Once this is done, the estimates of the means and variance are calculated by the sample means and sample variance:

$$\text{sample mean:} \quad \bar{x}_s = \frac{1}{N} \sum x_n$$

$$\text{sample variance:} \quad s_x^2 = \frac{1}{N-1} \sum (x_n - \bar{x}_s)^2$$

However, these are only estimates of the true mean and variance. Performance as determined by previous sections assumes that we have the true mean and variance. To compensate for only having sample means and sample variances, we have to use "Confidence Levels" and "Confidence Intervals" [8] in describing the biometric performance.

Given N biometric samples, with probability "1-α" confidence level, the sample mean will lie somewhere within a confidence interval between an upper and lower bound which depend on the number of samples, sample mean, the sample variance and a confidence level:

$$\bar{x}_s - \frac{s_x}{\sqrt{N}} t_{\alpha/2,N-1} < \bar{x} < \bar{x}_s + \frac{s_x}{\sqrt{N}} t_{\alpha/2,N-1}$$

where $t_{\alpha/2,N-1}$ is the t-distribution point [8]

$$P(t_{N-1} \geq t_{\alpha/2,N-1}) = \frac{\alpha}{2}$$

Similarly, with probability "1-α" confidence level, the sample variance will lie somewhere within a confidence interval between an upper and lower bound which depend on the number of samples, the sample variance and a confidence level:

$$\frac{(N-1)\, s_x^2}{\chi^2_{\alpha/2;N-1}} < \sigma_x^2 < \frac{(N-1)\, s_x^2}{\chi^2_{1-\alpha/2;N-1}}$$

where $\chi^2_{\alpha/2;N-1}$ is the chi-squared point [8]

$$P(\chi^2_{N-1} \geq \chi^2_{\alpha/2;N-1}) = \frac{\alpha}{2}$$

In biometric system design, we desire to have tight bounds to evaluate/ensure decision performances. Hence, it is obvious from the bound on the sample means, we wish to have a large number of samples to obtain the sample mean with a tight bound. From the "N-1" numerator term in the sample variance bounds, it initially appears that the bound increases with increasing number of samples, but $\chi^2_{\alpha/2;N-1}$ [8] decreases more rapidly than N-1 increases and the net effect is to decrease the sample variance as N-1 increases.

4.7 Decidability Index

The Decidability Index, d' is a measure of the authentic and impostor distributions separation, given by:

$$d' = \frac{|\bar{x}_a - \bar{x}_i|}{\sqrt{\frac{\sigma_a^2 + \sigma_i^2}{2}}}$$

were $\bar{X}_a$ and $\bar{X}_i$ are the means of the authentic and impostor histograms with variances σ^2_a and σ^2_i. Since the design performance is to maximize the probability of correct decisions and minimize the incorrect decisions, there are two biometric separation properties which can influence the decision performance:

i) maximum separation of the authentic and impostor means, i.e. max: $|\bar{x}_a - \bar{x}_i|$, and
ii) minimize both the authentic and impostor variances, i.e. min: $\{\sigma^2_{x_a}, \sigma^2_{x_i}\}$

If either/both of the above biometric separation properties improve, Figure 8 illustrates the pdf's.

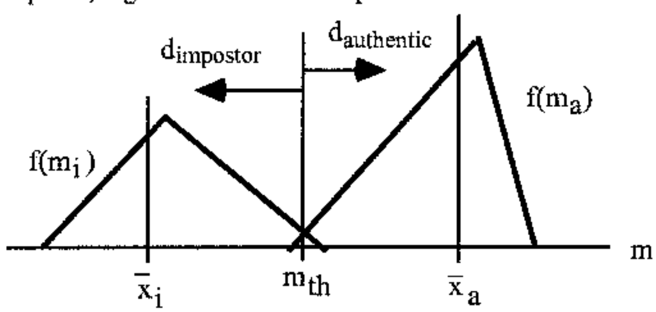


Figure 8. Authentic/Impostor pdf's and threshold decision

Mathematically, d' is independent of the decision threshold and reflects the degree to which any improvement in FAR must be paid for by relaxing of the FRR. As can be seen from the equation the larger d', the better the separation, hence for the best possible discrimination we want a large value of d'. It is obvious that the corresponding Probability of correct/incorrect decisions will improve no matter what the threshold setting is whenever d' is increased.

5. Testing Paradigms

As with using anything new, unfamiliarity can contribute to errors. Biometric systems also exhibit a learning curve. Generally, as users in a test suite become more familiar with the system (habitation) they learn techniques or tricks needed to properly interface to the system and false rejection errors tend to decrease. Proper training can significantly reduce the time to gain system familiarity hence minimizing the initial false reject rate since clues can be given as opposed to being learned by trial and error.

Uncooperative test subjects or unwilling users can skew collected data effecting the measured operative accuracy of a system under test. Thus, it is imperative to design tests that identify or minimize potential human bias during the collection of biometric performance data. Consider the conditions under which data for biometric performance can be collected. Three conditions apply:

- Ideal Data Collection,
- Controlled Real World Data Collection and
- Uncontrolled Real World Data Collection.

In the first case, data is collected and analyzed under the best possible conditions, for instance a user's head may be placed in a chin rest for facial or iris recognition, environmental factors such as ambient temperature and lighting can be controlled. For the second case, data is collected in a real world operational application but under a well controlled experimental situation, for instance having a set of known cooperative subjects use the equipment. The third approach uses data collected at an actual application by actual users, hence it is in an unsupervised, uncontrolled environment.

Each of these scenarios can be used to provide a different perspective on the performance of the system. Ideal data collection allows evaluation of the raw biometric implementation along with its sensors and other components. Controlled real world data introduces the environment as well as an application (but with cooperative users) thereby allowing testing of the biometric system under an actual application. Uncontrolled real world data collection introduces the most variability and is the most difficult. For the last case, it is envisioned that there is an automatic data collection that would report the results of the system's performance over some operational period, there are some issues in this approach as actual data (is it really a false reject or an impostor trying to gain access) would not be known.

5.1 Dual Thresholds

Sometimes "upper and lower bound" acceptance thresholds are sometimes used. This approach allows the user to try a second time if the initial score is between an upper threshold (which would passes the user) and a lower threshold (which would fail the user). Clearly, dual thresholds affects the reported FAR and FRR of the biometric data.

5.2 Attempts and Tries

Collecting statistics for false accepts and false rejects requires a definition of whether the data is collected for a "single try" or is decided after "n tries". The word "try" is used to define a single presentation of the user to the biometric system for measurement (verification). The word "Attempt" defines a cycle of an individual using the biometric system. Most verification devices allow more than one try per attempt and may even internally take multiple samples of the user's biometric for each try. Hence, when collecting and reporting data it is important to define how the counting is done.

Table 3 describes how to count accept and reject data for a "maximum three try" decision process [2]. In this case, accept and reject are defined as whether the user's biometric score is greater or less than a predefined acceptance threshold. If data is collected for multiple users, with multiple attempts per user, the false-reject rate for one-try, two-try or three-try statistics can be computed by using the total of rejects divided by the total attempts, hence for the table shown (assuming each verification results is from a user): FRR = ¾ for one-try statistics, FRR = ½ for two-try statistics and FRR = ¼ for three-try statistics. Other ways of reporting the performance statistics are also possible using this approach for instance one could report the FRR statistics for individual tries.

Table 3 Three Try Decision Counting for a Verification System

Verification Result	One-try Statistics	Two-try Statistics	Three-try Statistics
Accept on 1st try	Accept	Accept	Accept
Accept on 2nd try	Reject	Accept	Accept
Accept on 3rd try	Reject	Reject	Accept
No Accepts in 3 tries	Reject	Reject	Reject

5.3 Some Formal Test Reports

In 1991, Sandia National Laboratories released a report entitled "A Performance Evaluation of Biometric Identification Devices" which describes the testing of a select set of vendors' biometric devices [2]. Specifically, fingerprints, hand geometry, signature, retina and voice biometrics were evaluated and error rate curves generated. A second report titled "Laboratory Evaluation of the IriScan Prototype Biometric Identifier" was issued in 1996 [3]. Both of these reports are a good reference on the procedures and issues that need to be considered for testing of biometric devices. For instance, it is seen that as a user becomes more familiar with a biometric, the false rejection rate decreases, additionally, it was found that retraining and reenrollment of problem users may not always result in higher performance. Sandia's testing attempted to get as many transactions as possible (without too much fatigue or loss of participation from disinterest) from a limited set of users/volunteers.

The FERET (Face-Recognition Technology) program administered by the US Army Research Laboratory [4,5,10] provided a large database of faces collected in a controlled setting (consistent with mug shots or drivers license photos) so that four face recognition algorithms could evaluated under double-blind testing conditions. Results showed dependence on illumination and sensors as well as time between enrollment and verifications (comparison of images taken the same day verses a year apart).

6. Conclusions/Recommendations

Today's automated biometrics technologies are changing due to advances in computer hardware, sensors and algorithms. There are potentially large and diversified uses for biometrics including such applications as: securing Internet credit card based transactions, secure computer logon, ATM access, law enforcement related identification, building access, access to medical records and access to secure facilities. A significant challenge faced in implementing a biometric system is understanding the performance of the system in terms of the tradeoff between FAR and FRR for the particular application. This paper presented an overview of biometrics and statistical measurements currently used to describe the performance of biometric systems. Measuring the performance of a biometric system requires well defined testing and sufficient data to extrapolate performance to the actual user population. Providing quantitative performance to compare different biometrics or even the same biometric using a different platform represents a larger task requiring numerous test subjects, a good definition of the process and a standard data analysis approach. The issue of comparative performance measurements represents a challenge to the research community as well as the biometric industry.

References

1. A. Jain, R. Bolle and S. Pankanti, *Biometrics Personal Identification in Networked Society*. Kluwer Academic Publishers, Boston, July 1999.
2. J. P. Holmes, et al., "A Performance Evaluation of Biometric Identification Devices", *Sandia National Laboratories,* SAND91-0276, June 1991.
3. F. Bouchier, J. Ahrens, G. Wells, "Laboratory Evaluation of the IriScan Prototype Biometirc Identifier", *Sandia National Laboratories*, SAND96-1033, April 1996.
4. P.J. Philips, "FERET (Face-Recognition Technology) Recognition Algorithm Development and Test Results", *Army Research Laboratory*, ARL-TR-995, October 1996.
5. P.J. Philips, et al., "FERET (Face-Recognition Technology) Recognition Algorithms", *Proceedings of the ATRWG Science and Technology Conference*, July 1996.
6. A. Papoulis, *Probability, Random Variables and Stochastic Processes*, 3rd ed, McGraw Hill, New York, 1991.
7. H. Urkowitz, *Signal Theory and Random Processes*, Artech House, 1982.
8. D. Montgomery, *Introduction to Statistical Quality Control*, 3rd ed, John Wiley and Sons, New York, 1996.
9. J.D. Woodward, *Believing In Biometrics*, Information Security Magazine, March 1998.
10. A. Pentland, T. Choudbury, *Face Recogntion for Smart Environments*, IEEE Computer Magazine February 2000, Vol 3, No. 2.
11. M. Negin, T. Chmielewski et al, *An Iris Biometric System for Public and Personal Use,* IEEE Computer Magazine February 2000 Vol 3, No. 2.
12. P. J. Philips, A. Martin et al., *An Introduction to Evaluating Biometric Systems*, IEEE Computer Magazine February 2000 Vol 3, No. 2.

Measuring the quality of visual learning

Giovanni Bianco
Computer Science Serv., University of Verona, Italy
bianco@chiostro.univr.it

ABSTRACT

Biology often offers valuable example of systems both for learning and for controlling motion. Work in robotics has often been inspired by these findings in diverse ways. Nevertheless, the fundamental aspects that involve visual landmark learning has never been approached formally. In this paper we introduce results that explain how the visual learning works. Furthermore, these tools provide bases to measure the quality of visual landmark learning. Basically, the theoretical tools emerge from the navigation vector field produced by the visual navigation strategy. The learning process influence the motion vector field whose features are addressed.

1 INTRODUCTION

Animals are proficient in navigating and diverse methods of biological navigation have been recently studied and categorized as [20]: *guidance, place recognition - triggered response, topological* and *metric* navigations. In order to perform such tasks animals usually deal with identifiable objects in the environment called *landmarks* [21].

The use of landmarks in robotics has been extensively studied [4]. Basically, a landmark needs to possess characteristics such as the *stationarity*, *reliability* in recognition, and *uniqueness*. These properties must be matched with the nature of a landmark: landmarks can be *artificial* or *natural*. Of course it is much easier to deal with artificial landmarks instead of dealing with natural ones, but the latter are more appealing because their use requires no engineering of the environment. However, a general method of dealing with natural landmarks still remains to be introduced. The main problem lies in the selection of the most suitable landmarks [19].

Recently it has been discovered that wasps and bees perform specific flights during the first journey to a new place to learn color, shape and distance of landmarks. Such flights are termed *Turn Back and Look (TBL)* [11]. Once the place has been recognized using landmarks, insects can then accomplish navigation actions accordingly.

Starting from Biological bases, the system described in this paper selects natural landmarks from the surrounding environment adopting the TBL phase. Once landmarks have been selected suitable navigation movements are computed. Iterating the process of computation of the navigation vector over the whole environment, a vector field is produced.

Studying the navigation vector field two main results are provided:

- the *visual potential function* generating the navigation vector field represents the driving principle to perform visual guidance. When proven to be a *Lyapunov* compliant function, we can state the navigation system exhibits convergence to the goal.
- The *conservativeness* of the navigation vector field provides key information about the quality of landmark learning.

Details about the navigation system and the computation of the potential function can be found in [2] and [3].

This paper addresses the learning process and its organization is as follows. In Section 2 aspects both related to findings on biological learning and to biological navigation will be introduced. In addition, this section addresses former work and research in the field of landmark learning in Robotics. In Section 4 the theoretical principles specifically involved with visual learning are detailed. Final remarks conclude the paper.

2 BIOLOGICAL FOUNDATIONS

Over many decades, studies of the visual performance of bees have exploited the fact that bees keep returning to a profitable feeding site once found, even when it is an artificial food source established by an experimenter.

2.1 Landmark learning

As soon as the bee encounters a novel place, she turns by 180 degrees to inspect the place and performs the initial phase of training, termed the *Turn-Back-and-Look (TBL)* phase [10]. A similar behavior was also observed in other insects thus categorizing this phase a typical behavior of an insect when a new visual learning phase is needed [22, 23].

In references [10, 11] and [14] the details and results on the visual parameters learned by TBL are introduced. Basically, findings show that TBL performed on departure serves primarily for acquiring depth information by exploiting image motion, whereas color, shape and size of landmarks are mainly acquired on arrival.

Attempts to understand in detail the geometric significance of learning flights have only recently been made. Essentially, the flights are invariant in certain dynamic and geometric structures thus allowing the insects to artificially produce visual cues in specific areas of the eyes [24]. Perhaps, the main reason is that the precision for the homing mostly depends upon the proximity of chosen landmarks to the goal [6]. In fact, those flights need to be repeated whenever some changes in the goal position occur [12].

2.2 *Landmark guidance*

Landmarks guidance in insects is retinoptically driven and animals tend to reduce the discrepancies between the stored view and the actual one by a matching procedure (reviews in [7] and [21]). The survey work presented in [20] addresses biological navigating behaviors from a robotics point of view.

Referring to landmark guidance in bees, the seminal work is presented in [5]. The authors show how bees learn landmarks by storing an unprocessed two dimensional snapshot of the panorama. The model matches landmarks in the stored snapshot with landmarks in the actual image. If this match is performed far from the goal every matched pair could differ both in angular size and compass bearing. These differences drive a bee toward the right position.

3 RELATED WORK

The guidance model introduced in [5] has some shortcomings and interesting extensions have been addressed in recent works. Basically, a guidance strategy that operates with landmarks strives to reduce the differences between the pre-learnt landmarks at the goal position and the same landmarks viewed from a different place. The extraction of landmarks follow different schemas such as in [18] where visual moments are applied on the panorama image to extract prominent features or as in [8] and [16] where unique (small) portions of the whole image, called templates, are extracted.

Operating with landmarks extracted from the panorama, navigation vectors can be computed. Unfortunately, none of the work previously reported tries to handle the mathematical features of the navigation vector fields thus produced.

A formal interpretation of the visual guidance behavior is firstly presented in [1] where two fundamental principles are extracted from the strategy navigation field: the visual potential function and the measure of conservativeness. The latter has been proved to measure the quality of landmark learning whereas the former is a funnel-shaped function that explain why guidance strategies operate with a gradient process to lead the robot to the goal (the global minimum).

4 THE MOTION FIELD

According to what has been previously expressed, starting with local visual information, a vector needs to be computed by the agent which will be used it to perform the next movement. In our case, the computation of the navigation vector is based on information involving the chosen landmarks. How to get navigation information from landmarks is briefly introduced here for completeness and details can be found in [2, 3].

Basically, once landmarks have been learned, they can provide two kind of information to perform motion:

- their actual size, compared to the size learned at the goal site, reports how far/close the agent is to the goal position
- their actual orientation, compared to the orientation learned at the goal site, speaks about the actual left/right shift of the agent.

This kind of data come from each individual landmark and we need to fuse them in order to get the overall navigation vector. Intriguingly, the fusion procedure has strong biological bases as detailed in [20].

To formalize aspects related to the motion field generated in the environment, we call $\mathbf{p}$ the vector representing the robot's Cartesian position $[x\ y]$ in a world reference $\mathbf{W}$. We also define step k the discrete time k of *robot dynamic state*.

Let $\vec{V}(\mathbf{p}(k)) = [V_x(\mathbf{p}(k))\ \ V_y(\mathbf{p}(k))]$ be the output of the motion strategy at a given step k, i.e. the robot movement at step k. If the robot operates in *position mode*, i.e. at each step it updates its Cartesian position, then

$$\mathbf{p}(k+\mathit{1}) = \mathbf{p}(k) + \vec{V}(\mathbf{p}(k)) \qquad (1)$$

where $\mathbf{p}(k)$ represents the coordinates of robot at step k, and $\mathbf{p}(k+\mathit{1})$ represents the new position at step $k+1$. The goal position is defined as an equilibrium point $\mathbf{p}^*$ for the system.

The computation over the whole environment of vector $\vec{V}$ defines a vector field $\mathbf{V}$. Let us consider a partial set of equivalent statements about a generic vector field $\mathbf{V}$ [15].

- any oriented simple closed curve $\mathbf{c}$: $\oint_{\mathbf{c}} \mathbf{V} \cdot d\mathbf{s} = \mathit{0}$
- $\mathbf{V}$ is the gradient of some function U: $\mathbf{V} = \nabla U$

The former is related to the concept of *conservativeness* of the field. The latter is concerned with the existence of a *potential function* generating an unique field. From a different point of view, conservativeness is a measure of the quality of landmark learning, whereas the existence of a Lyapunov potential function indicates the robot's capability to reach the goal. The following Section addresses the former aspect. Details of the other aspect can be found in [2, 3].

The robot *Nomad200* was used to accomplish the tests. It includes the *Fujitsu Tracking Card (TRV)* which performs real-time tracking of full color templates at a NTSC frame rate (30Hz).

5 PRINCIPLES FOR LANDMARK LEARNING

A landmark must be reliable for accomplishing a task as detailed in Section 2.1. Landmarks that appear to be appropriate for human beings are not necessarily appropriate for other agents (animals, insects or artificial beings) because of the completely different sensor apparatus and matching systems [19]. Therefore we need to state the meaning of landmark reliability in advance for the system in use before to solve the problem of selecting landmarks.

For our system, a template is a region of the grabbed image identified by two parameters m_x and m_y representing the sizes along X and Y axes. The size ranges from 1 to 8, i.e. from *small* (2^1 pixels wide) to *large* (2^8 pixels wide) templates. The TRV can simultaneously track many templates. For each template the card performs a match in a sub-area of the actual video frame adopting the block matching method [9]. This introduces the concept of

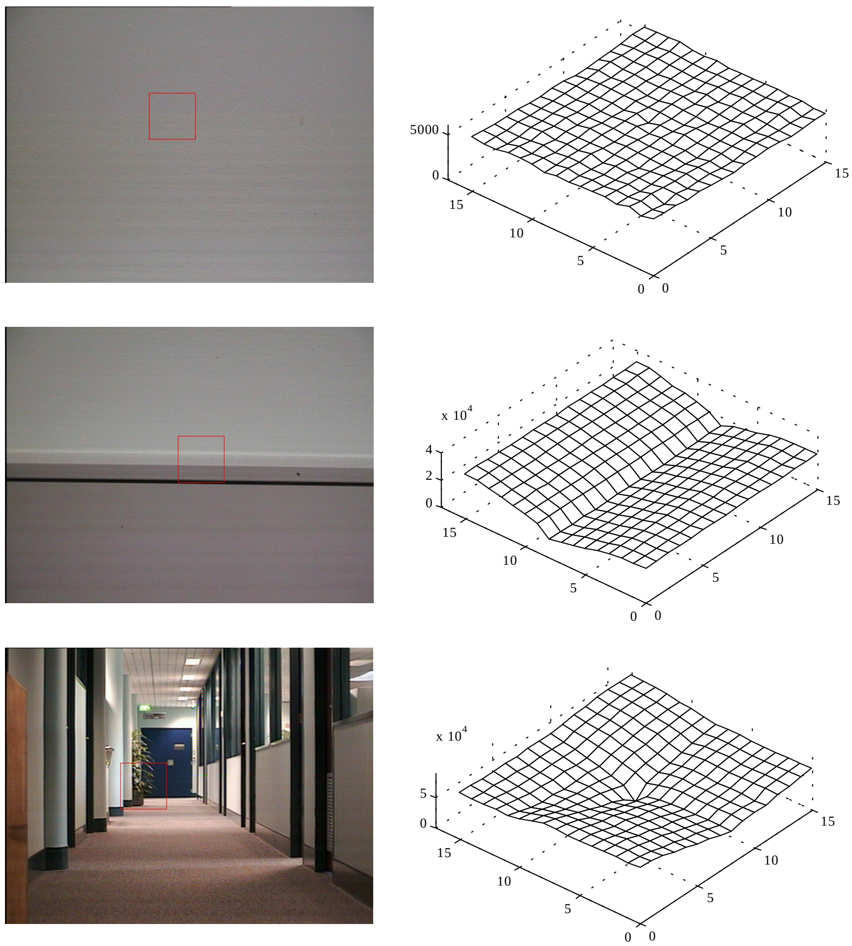


Figure 1: Examples of correlation matrices. These are computed within the local sub-area of the templates (square box in the pictures).

correlation between the template being used and the actual video frame. The sub-area is composed of 16×16 positions in the frame usually taken around the origin (o_x, o_y) of the template (its upper-left corner). The whole set of computed correlation measures forms the *correlation matrix*. Examples of correlation matrices are reported in figure 1.

We can take advantage of the matrices to compute a measure that states upon the reliability of the template under study [17]. As reported in [2, 13] we calculate a figure r, ranging from 0 to 1, which states how deep is the global minimum of the matrix in relation to its neighborhood. Therefore, we define reliable landmarks as *templates which are uniquely identifiable* in their neighborhoods: the greater r the more uniquely identifiable the landmark in its sub-area.

Once that the measure for the reliability of a landmark has been stated, the next step consists of searching the whole panorama for landmarks. There are several degrees of freedom in searching for the best landmarks within a video frame [2], but some simplifications can be introduced: only square templates are used, and the position of a landmark is searched for by maximizing the following:

$$(o_x^*, o_y^*) = \arg \max_{(o_x, o_y) \in \mathrm{grid}} r_l(o_x, o_y) \qquad (2)$$

where $r_l(o_x, o_y)$ identifies the reliability factor for a landmark l whose origin is located in (o_x, o_y) representing a generic place on the grid. The position (o_x^*, o_y^*) represents the cells with the highest r. In order to assure that different landmarks occupy different positions, previously chosen coordinates are not considered. In figure 2, examples of landmarks chosen have been reported. When different sizes are considered, different sets of landmarks are extracted.

The landmarks which have been *statically* chosen are used for navigation tasks. This is done by testing the landmarks to verify that they represent good guides for navi-

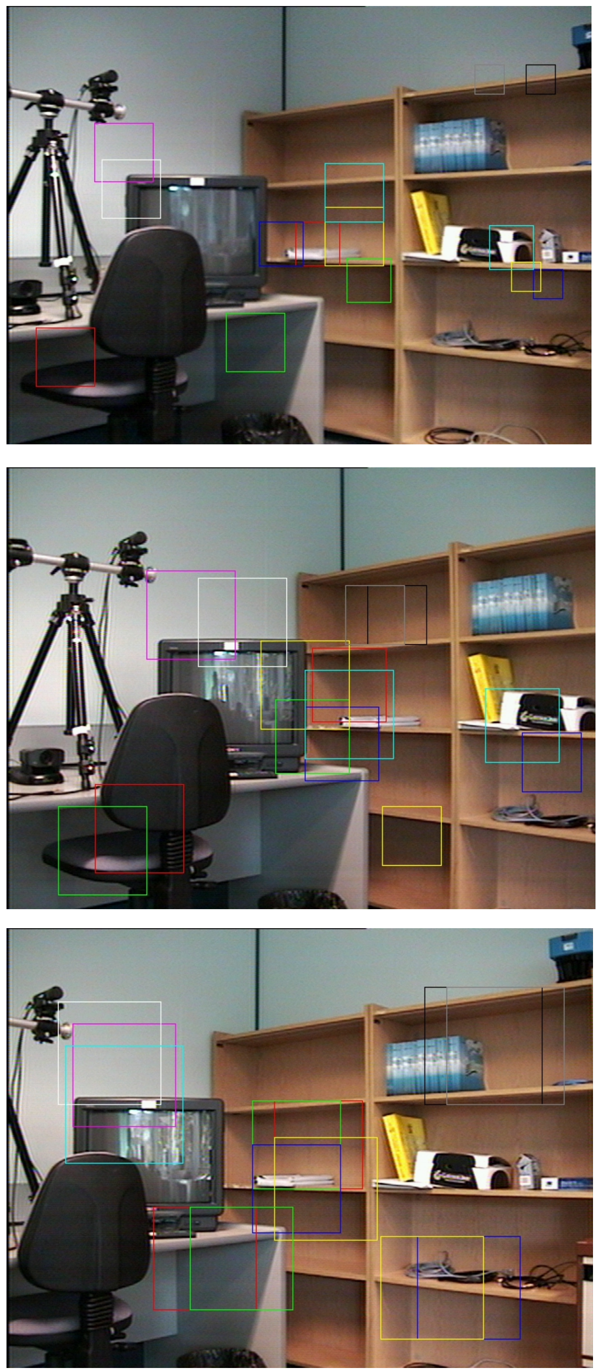


Figure 2: Different choices of landmarks for different landmark sizes. Landmarks are box-shaped.

gation tasks.

TBL helps to verify landmarks by testing whether during the motion the statically chosen landmarks are robustly identifiable. Through a series of stereotyped movements small perturbations (local lighting conditions, changes in camera heading, different perspectives and so on) can influence the reliability of the statically chosen landmarks.

These perturbations to images naturally occur in typical robot journeys thus allowing us to state that the TBL phase represents a *testing framework* for landmarks. In other words, the robot tries to *learn* which landmarks are suitable for use in real navigation tasks by simulating the conditions the robot will encounter along the paths. At the end of the TBL process only those landmarks whose reliability r_l is above a certain threshold ϵ are suitable to be used in navigation tasks.

The reliability factor r_l for landmark l is continuously computed during the TBL phase through the following:

$$r_l = \frac{\sum_{i=1}^{TBL} r_l^i}{TBL} \tag{3}$$

where TBL is the total number of steps exploited till that time, and r_l^i is the reliability of landmark l calculated at time i. In the tests, at the end of the phase, TBL usually consists of 400 steps (it takes about 13 seconds to be performed). The set of landmarks is tracked along the whole TBL phase and r_l is continuously monitored for each landmark (details in [13]).

5.1 The quality of learning

There are strong connections between the learning phase and navigation actions. The conservativeness of the motion field bridges these two aspects.

A vector field $\mathbf{V}$ is said to be *conservative* when the integral computed on any closed path is zero. Conversely, if the field is not conservative then diverse potential functions can be associated with the field. This translates into *non-repeatability* of robot navigation trails in [13].

If the vector field is defined on a connected set in the environments, then the null *circuitation* property is equivalent to [15]:

$$\frac{\partial V_x(x,y)}{\partial y} = \frac{\partial V_y(x,y)}{\partial x} \tag{4}$$

We can measure how this equation differs from the theoretical null value as follows:

$$\frac{\partial V_x(x,y)}{\partial y} - \frac{\partial V_y(x,y)}{\partial x} \tag{5}$$

The property expressed by Equation 5 is referred to as *degree of conservativeness.* The degree of conservativeness of the vector field computed with a threshold set to 0 and landmarks sized 6 is shown in figure 3. Only small regions of the whole area roughly satisfy the constraint.

A small change in the threshold for TBL can dramatically change the situation. In figure 4 the degree of conservativeness for each point is plotted.

A key consideration is concerned with the scale along Z: it is about one order of magnitude less than the one reported in figure 3. A trend toward a conservative field is thus becoming evident.

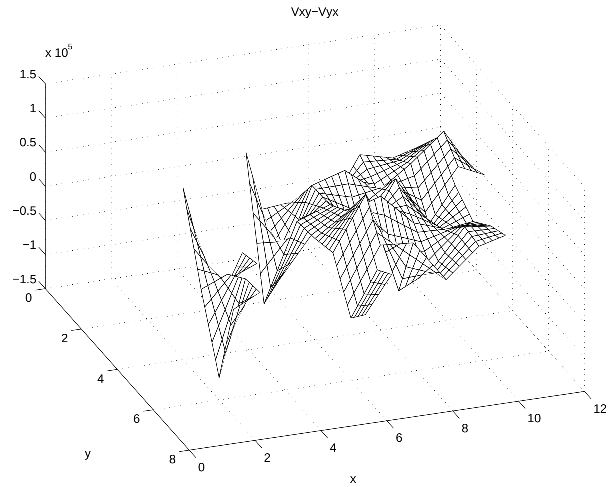


Figure 3: Conservativeness of a vector field computed with a TBL threshold of 0 and landmarks sized 6

The situation obtained with a threshold of TBL set to 0.2 has been reported in figure 5. A large area of the environment has a degree of conservativeness that roughly equals 0.

Similar considerations can be expressed dealing with a different landmark size [1]. The *template* of the graph is the same as before. Therefore, with a good choice of threshold the field becomes conservative regardless of the size of the landmarks.

6 CONCLUSIONS

Landmarks learning for robots can take inspiration from Biology but it needs to be well formalized for its efficient implementation in artificial agents. First, a definition for landmark reliability must be stated. Second, a measure that can assess about the quality of the learning phase needs to be introduced.

In this paper, we have shown how both these aspects can be efficiently addressed. Particularly, we have shown how the learning phase affects the navigation motion field. Further improvements to this study can be achieved by the use of omni directional visual sensors.

References

[1] G. Bianco. *Biologically-inspired visual landmark learning and navigation for mobile robots.* PhD thesis, Department of Engineering for Automation, University of Brescia (Italy), December 1998.

[2] G. Bianco and A. Zelinsky. Biologically-inspired visual landmark navigation for mobile robots. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Kyongju (Korea), October 17-21 1999.

[3] G. Bianco and A. Zelinsky. Dealing with robustness in mobile robot guidance while operating with visual

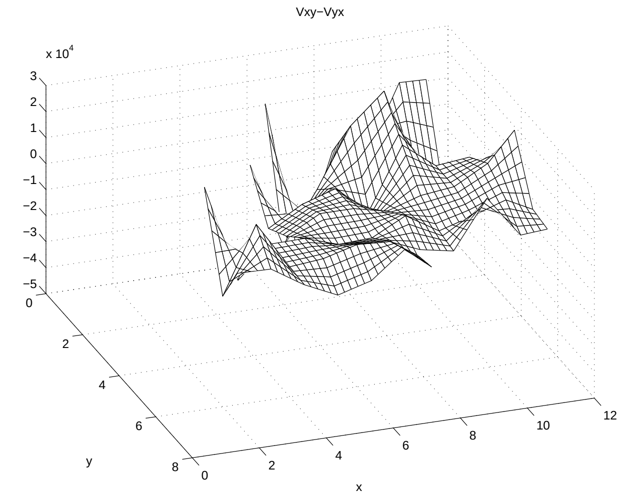


Figure 4: Conservativeness of a vector field computed with a TBL threshold of 0.1 and landmarks sized 6

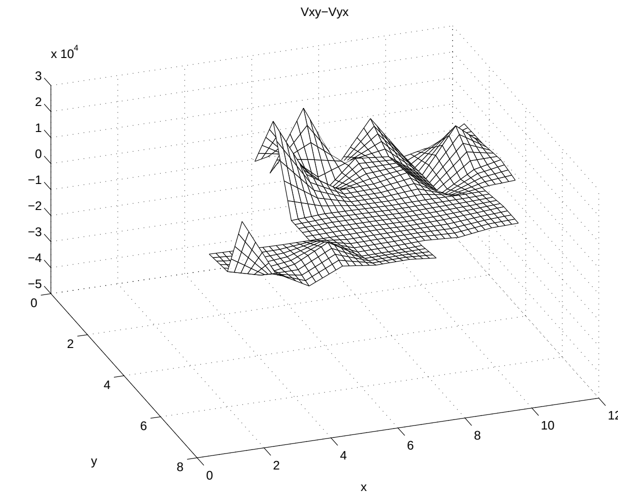


Figure 5: Conservativeness of a vector field computed with a TBL threshold of 0.2 and landmarks sized 6

strategies. In *Proceedings of the IEEE International Conference on Robotics and Automation*, San Francisco (CA), April 24-28 2000.

[4] J. Borenstein, H. Everett, and L. Feng. *Where am I? Sensors and Methods for Mobile Robot Positioning.* The University of Michigan, April 1996.

[5] B. Cartwright and T. Collett. Landmark learning in bees. *Journal of Comparative Physiology*, A(151):521–543, 1983.

[6] K. Cheng, T. Collett, A. Pickhard, and R. Wehner. The use of visual landmarks by honeybees: Bees weight landmarks according to their distance from the goal. *Journal of Comparative Physiology*, A(161):469–475, 1987.

[7] T. Collett. Landmark learning and guidance in insects. *Phil. Trans. R. Soc. London*, B(337):295–303, 1992.

[8] I. Horswill. *Polly: a Vision-Based Artificial Agent.* PhD thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, May 1993.

[9] H. Inoue, T. Tachikawa, and M. Inaba. Robot Vision System with a Correlation Chip for Real-time Tracking, Optical Flow and Depth Map Generation. In *Proceedings of IEEE International Conference on Robotics and Automation*, pages 1621–1626, Nice, France, May 12–14 1992.

[10] M. Lehrer. Bees which turn back and look. *Naturwis*, (78):274–276, 1991.

[11] M. Lehrer. Why do bees turn back and look? *Journal of Comparative Physiology*, A(172):549–563, 1993.

[12] M. Lehrer. Honeybees' visual spatial orientation at the feeding site. In M. Lehrer, editor, *Orientation and communications in arthropods*, pages 115–144. Birkhauser verlag, Basel/Switzerland, 1997.

[13] M. Lehrer and G. Bianco. The turn-back-and-look behaviour: Bee versus robot. *Biological Cybernetics*, 2000.

[14] M. Lehrer and T. Collett. Approaching and departing bees learn different cues to the distance of a landmark. *Journal of Comparative Physiology*, A(175):171–177, 1994.

[15] J. Mardsen and A. Tromba. *Vector calculus.* W.H. Freeman and Company, 1996.

[16] Y. Matsumoto. *View-Based Approach to Mobile Robot Navigation.* PhD thesis, Graduate School of Engineering, The University of Tokyo (Hongo 7-3-1, Bunkyo-Ku), February 1998.

[17] T. Mori, Y. Matsumoto, T. Shibata, M. Inaba, and H. Inoue. Trackable attention point generation based on classification of correlation value distribution. In *JSME Annual Conference on Robotics and Mechatronics (ROBOMEC 95)*, pages 1076–1079, Kavasaki (Japan), 1995.

[18] J. Salas and J. Gordillo. Robot location using vision to recognize artificial landmarks. In *SPIE*, volume 2354, pages 170–180, 1995.

[19] S. Thrun. A bayesian approach to landmark discovery and active perception in mobile robot navigation. Technical report, School of Computer Science Carnegie Mellon University, 1996.

[20] O. Trullier, S. Wiener, A. Berthoz, and J. Meyer. Biologically based artificial navigation systems: Review and prospects. *Progress in Neurobiology*, 51:483–544, 1997.

[21] R. Wehner. Arthropods. In F. Papi, editor, *Animal Homing*, pages 45–144. Chapman and Hall, London, 1992.

[22] J. Zeil. Orientation flights of solitary wasps 1: Description of flights. *Journal of Comparative Physiology*, A(172):189–205, 1993.

[23] J. Zeil. Orientation flights of solitary wasps 2: similarities between orientation and return flights and the use of motion parallax. *Journal of Comparative Physiology*, A(172):207–222, 1993.

[24] J. Zeil, A. Kelber, and R. Voss. Structure and function of learning flights in bees and wasps. *Journal of Experimental Biology*, 199:245–252, 1996.

Autonomous Mental Development and Performance Metrics for Intelligent Systems

Juyang Weng
Department of Computer Science and Engineering
Michigan State University
East Lansing, MI 48824
http://www.cse.msu.edu/~weng/

Abstract

In this paper, some resent advances in neuroscience, psychology, robotics and machine intelligence are briefly reviewed. They prompt us to pay attention to the fundamental difference between the way human intelligence is developed and the traditional engineering paradigm for developing a machine. They make us rethink the issue of intelligence. This position paper proposes that a fundamental criterion for a true intelligent system is not really what it can do in a special setting, but rather, its capability for autonomously and incrementally developing its cognitive and behavioral capability through online real-time interactions with its environment, directly using its sensors and effects, a process called mental development in neuroscience and psychology. The term ``mental'' here includes cognitive, behavioral, sensorimotor and other mental skills that are exhibited by animals and humans. The new direction of autonomous mental development for machines will create a new kind of machines, called developmental robots. With new perspectives from developmental robots, the performance metrics for machine intelligence will undergo a revolution. They will fundamentally change the current fragmented landscape of the AI field by shifting the emphasis of measuring ad hoc capability of performing a task-specific application to a systematic measurement of mental developmental capabilities. Such performance metrics can be adapted from those for humans ¾ a series of tests well developed by a well-established field called psychometrics.

1 Background

Human understanding of the ways our own minds work, the power and limitation of existing machines, as well as the relationship between humans and machines have greatly improved over the last 50 years. It is now clear that a *developed* human mind, that of a normal human adult, is extremely complex. It is also clear that the early optimism in the 60's and the 70's about a quick progress in artificial intelligence such as vision, speech, and language, was not well founded, at least not so with the traditional approaches that have been extensively experimented with so far. However, the past work with the traditional approaches is by no means unimportant. In fact, they are the womb and incubator for the birth and growth of a drastically different approach — autonomous mental development. This new direction is expected to become a revolution in the course of machine intelligence[1]. As Thomas S. Kuhn wrote in his book titled *The Structure of Scientific Revolution* [1]: "Because it demands large-scale paradigm destruction and major shifts in the problems and techniques of normal science, the emergence of new theories is generally preceded by a period of pronounced professional insecurity. As one might expect, that insecurity is generated by the persistent failure of the puzzles of normal science to come out as they should. Failure of existing rules is the prelude to a search for new ones."

The puzzle pieces from recent advances in related fields start to reveal a picture of *mental development*, which is no longer a total myth that is beyond human comprehension, but can be explained in terms of computation. In the following we briefly summarize these new thought-provoking advances.

[1] A more detailed discussion on this issue is available in the proceedings of Workshop on Development and Learning, funded by NSF and DARPA, held at Michigan State University, East Lansing, MI, April 5 – 7, 2000 (http://www.cse.msu.edu/dl/). This workshop was attended by about 30 distinguished researchers in neuroscience, developmental psychology, machine intelligence and robotics who are working on related subjects in their fields. The goal of this workshop was to discuss the state-of-the-art in research on mental development and to discuss, initiate and plan future research on this subject.

1.1 Neuroscience and psychology

A traditional view is that human brain is very much pre-determined by human genes. With this view, the brain unfolds its pre-determined structure during the development, which starts from the time of conception. This structure serves as a placeholder of information that is acquired from the environment. However, recent advances in brain plasticity have begun to reveal a very different picture of brain development. For example, researchers at MIT [2] have discovered that if the optical nerves from the eyes is rewired into the auditory cortex of the primate (ferret) early in life, the primate's auditory cortex gradually takes on representation that is observed in normal visual cortex. Further, the primates have successfully learned to perform vision tasks using the auditory cortex. In other words, the rewired ferrets can see in the sound zone. This discovery seems to suggest that the cortex is governed by self-organizing mechanisms, which derive representation and architecture according to the input signals, either visual or auditory. As another example, studies by researchers at the University of California at San Francisco [3] showed that the finger skin areas from which a neuron in somatic cortex receives sensory signals (called receptive field of the neuron) can change according to sensory experience. If multiple fingers of the adult monkeys receive consistent synchronized pulse stimuli from a cross-finger bar for several days, the receptive field changes drastically, from covering only a single finger in normal cases to covering multiple fingers. This result appears to indicate that the self-organizing program of our brain autonomously selects the source of sensory input within a candidate area according to the statistical properties of the actual sensory signal that is received. These and other related studies on the brain plasticity prompt us to rethink the traditional rigid view about the brain. It appears that the developmental program of the brain does not rigidly determine the brain's architecture and representation. For example, it might determine what statistical properties of the sensory signals should be used and how these properties are used to derive the representation and architecture of the brain.

In recent years, computational modeling of neural development has become a very active subject of study in neural science and psychology. For example, there have been several computational models for the development of response patterns in the retina, the lateral geniculate nucleus, and simple cells in the visual cortex. A subject that is now very actively studied is the mechanisms for developing orientational selectivity in the simple cells of the visual cortex. Although most computational models of developmental mechanisms have been concentrating on early processing (early in the order of processing steps in the brain), such a trend will certainly extend to later processing when global developmental models are increasingly studied for robots. Psychology has begun to move from qualitative descriptive models to more rigorous quantitative models for studying cognitive and behavioral processes. Some recent works in psychology has started to explain the global process of mental development using the computational element of networks [4]. Another new trend in psychology is to use explicit dynamics models to explain some well-known developmental facts about infant behaviors (e.g., the work at Indiana University [5]). These quantitative studies have begun to produce results that are more clearly understandable and verifiable than vague verbal theories and arguments.

1.2 Robotics and Machine Intelligence

Although autonomous mental development in humans is a well-known fact, the counterpart for machines did not receive serious attention until middle 90's. It has long been believed that the approach to machine intelligence does not have to follow what human minds do, just like modern airplanes which do not fly like birds. Gradually, many AI researchers started to realize that machine intelligence requires much more cognitive and behavioral capabilities than most had realized. Flying is a very simple problem in comparison with machine intelligence. Further, many AI researchers have already realized that machine intelligence requires "grounding" — concepts must be grounded on real sensory experience about the physical world, which in turn requires the machine to have a sensor-rich body (i.e., embodiment) that can directly sense stimuli from the physical world and act upon what it senses. However, grounded sensing and action, including learning, has been extensively studied and experimented with in robotics for many years. Why then does the reality of intelligent machines seem so remote? Since 1996, I argued [6] that what has been sorely missing from machines is *autonomous mental* *development*, or simply called *mental development*.

Autonomous mental development requires a true revolution in the way engineering has been done (i.e., paradigm) for thousands of years. The current *manual developmental paradigm* is as follows:

1. *Start with a task:* Given a task to be executed by a machine, it is the human engineer who understands the task (not the machine).
2. *Design task-specific representation:* The human engineer translates his understanding into a representation (e.g., giving some symbols or rules that represent particular concepts for the tasks and the correspondence between the symbols and physical concepts). The representation reflects how the human engineer understands the task.
3. *Task-specific programming:* The human engineer then writes a program (or designs a mechanism) that controls the machine to perform the task using the representation.
4. *Run the program on the machine.* Sensory data may be used to modify the parameters of the task-specific representation. However, since the program is of special purpose for the task, the machine does not even know what it is doing at all. All it does is running the program.

The new paradigm, *autonomous developmental paradigm,* for constructing developmental machines or robots, is as follows:

1. *Design body:* According to the general ecological condition in which the robot will work (e.g., on-land or underwater), human designers determine the sensors, the effectors and the computational resources that the robot needs and the human designs a sensor-rich robot body.
2. *Design developmental program:* Human designer designs the *developmental program* for the robot and starts to run this program.
3. *Birth:* The human operator loads the developmental program onto the computer in the robot body.
4. *Develop mind:* Humans mentally "raise" the developmental robot by interacting with it. The robot develops its mental skills through real-time, online interactions with the environment, including humans (e.g., let them attend special lessons). Human operators teach robots through verbal, gestural or written commands very much like the way parents teach their children. New skills and concepts are learned by the robots daily. The software (brain) can be downloaded from the robots of different mental ages to be run by millions of other computers, e.g., desktop computers.

Mental development has long been mistakenly thought of as being simulated by traditional machine learning techniques (e.g., neural network techniques). In fact, all the traditional machine learning uses the *manual* developmental mode but mental development requires the *autonomous* developmental mode. What is the basic difference? *With autonomous mental development, machines will be able to learn subjects that their programmers do not know, or have not even thought about, just like human children who can learn subjects that their parents do not know*. The essence of autonomous mental development by machines is the capability of learning directly, interactively, and incrementally from the environment using the learner's own sensors and effectors. Therefore, a computer that has only impoverished sensors and effectors cannot do mental development well. A neural network that can only accept human edited offline sensory data does not develop its mind either, even if it can learn incrementally. A *developmental robot* is a robot that runs a developmental program and is allowed to learn and practice autonomously in the real physical world.

Although the concept of developmental program for machines is very new, a very rich set of techniques useful for developmental programs have already been developed in the past several decades in the fields of pattern recognition, robotics and machine intelligence, especially techniques applicable to high-dimensional data. These new techniques are being used in very innovative ways for developmental programs. Several developmental programs have been designed and tested on robots. Running a developmental program, the robots interact with the environment in real time using their sensors and effectors. Internal representation, perceptual capabilities and behavioral capabilities are developed autonomously as a result of interaction of the developmental program with the environment. Humans interact with such robots only through the robot's sensors, as a part of the environment. Just like the nature-nurture interaction for human mental development, the cognitive and behavioral skills of such a robot result from extensive interaction between what is programmed ("innate" developmental program) and what is sensed through real-time online experience. The mind and intelligence emerges gradually from such interactions.

Early examples of such developmental robots include Darwin V at The Neurosciences Institute, San Diego and the SAIL at Michigan State University, developed independently around the same time with very different goals. The goal of Darwin V [7] was to provide a concrete example for how the properties of more complex and realistic neural circuits are determined by the behavioral and environmental interactions of an autonomous device. Darwin V has been tested for the development of generalization behaviors in response to visual stimuli at different positions and orientations (visual invariance learning). It has also been tested for the association of aversive and appetitive stimuli with visual stimuli (value learning). SAIL was designed as an engineering testbed for developmental programs that are meant for scaling up for complex cognitive and behavioral capabilities [8]. SAIL-2 developmental program has been tested for automatic derivation of representation and architecture through development of association between visual stimuli of objects and eye aiming for the objects (object evoked visual attention), between visual stimuli of objects and arm pre-reaching for the object (vision evoked object reaching), between voice stimuli and arm actions (verbal command learning and execution) and between visual stimuli and locomotion effectors (vision-guided navigation). Other studies for online learning directly from sensors are in the direction towards fully autonomous developmental systems. The work at MIT associates video images of objects with synchronized voices (pronounced verbal name of the object) [9]. The work at the University of Massachusetts at Amherst investigated the use of coupling of robot leg joints that have been observed in infants to reduce the search space for a desirable turning gait [10]. Although the history of developmental robots is very short, some experiments by the above studies have demonstrated capabilities that have never been achieved by the traditional methods, such as in visual recognition, verbal communication, hand-eye coordination, autonomous navigation, value acquisition (learning the value of actions), and multimodal association in real time. We are aware that more groups in the US and other countries have already started to investigate this new direction.

2. Some Major Characteristics of Research on Mental Development

2.1 More tractable

It is known that a developed adult human brain is extremely complex, as an epigenetic product of long-term and extensive interactions with the complex human world. The developmental principles for the brain in the complex human world, however, should not be as complex as the human world itself. For example, the visual world is very complex, but the developmental principles that are used by the brain to derive various filters for processing visual signals should not be as complex as the visual world itself. Therefore, computational study of cognitive development could be more tractable than traditional approaches to understanding intelligence and constructing intelligence machines.

2.2 Unified framework

Studies of cognitive development will establish a unified framework for our understanding of a wide variety of cognitive and behavioral capabilities. Discovery of mechanisms responsible for developing cognitive and behavioral capabilities in humans requires more systematic work than an account of a particular individual capability, such as visual recognition in a simplified setting alone or stereotyped walking alone. Sharing of common developmental principles by visual and auditory sensing modalities, as revealed by recent neuroscience studies, will encourage scientists to discover further underlying developmental principles that are shared not only by different sensing and effector modalities, but also by different higher brain functions.

Traditionally, vision and speech have been considered very different, both for humans and for machines. For the same reason, traditional methods for different AI problems are typically very different, resulting in what is well known now as the fragmentation of the AI field. Potentially, AI can be applied to all possible areas of human life and each application area potentially can lead to a fragment of AI if it is treated in an *ad hoc* way. The unified framework of cognitive development will fundamentally change the current fragmented landscape of AI in the years to come, since different applications correspond to different lessons that can be taught to the same developmental robot at different mental ages. We will also see much more interactions and collaborations among scientists and engineers in neuroscience, psychology, robotics, artificial intelligence and other related fields, due to the very similar research issues these fields face under the theme of autonomous cognitive development.

2.3 Task-nonspecific

In contrast to the task-specific nature of the traditional engineering paradigm in AI, developmental programs for machines will be *task-nonspecific*. The power of a developmental program is its general applicability to many different tasks. A developmental program may contain certain pre-processing stages that are specific to some type of sensors or effectors, such as camera or touch sensors. In this sense, it is body-specific (or species-specific). However, it is not task-specific. A developmental program can be run to develop skills for many different tasks, with simpler skills being learned to prepare skills for learning more complex skills. Recently, the scientific community has gained a more complete understanding of human intelligence. As Howard Gardner put it in his book *Multiple Intelligences* [11], human intelligence is multiple, including linguistic, logical-mathematical, musical, bodily-kinesthetic, spatial, interpersonal, and intrapersonal intelligences. This is a rough classification of a very rich ensemble of inter-related cognitive and behavioral capabilities that give rise to human intelligence. The same is true for machine intelligence. Any particular capability that we regard as intelligence in a general setting, such as the visual capability of recognizing various persons on a busy street or the language capability of talking about technology, is not an isolated single thing. It requires the support of many skills developed through extensive real-world experience via sensors and effectors.

2.4 Computational

Further, developmental mechanisms seem to be very much quantitative in nature and thus require clear *computational models*. We will see more complete computational models for mental development that can be simulated on computers and robots for many different environmental conditions and the results can be verified against studies about humans. We will see more efforts on computational modeling of mental development, for humans and machines, that are clearly understandable, implementable on machines and can be subject to rigorous verification and comparison. This will indicate the maturation of the related fields.

2.5 Recursive and active

Development discourages any static or rigid view of the mind. A developed human mind is a snapshot of many years of *recursive and active* mental construction by the developmental program in the human genes, utilizing the sensory and action experience through life time. The term *recursive* means that later mental development relies on the cognitive and behavioral capabilities that have been developed earlier. The term active means that each individual plays an active role in the development of his or her mind --- different actions lead to different experience. The same is true for developmental robots. The recursive and active nature of development discourages the approach of collecting offline data and spoon-feeding them into a machine, which is a prevailing approach in current machine learning studies. Sensory data cannot be pre-specified since what sensory data is sensed depends on online action executed in real time.

2.6 Developmental capabilities as unified metrics for machine intelligence

The *criteria for measuring machine intelligence* will fundamentally change. The metrics that can be used to measure the power of such a new kind of machine is primarily their autonomous interactive learning capabilities in complex human environments. In other words, it is the capability of mental development instead of what the machine can do under a pre-specified setting. Such performance metrics can be adapted from those used by clinical psychologists for testing the cognitive development of human infants (e.g., The Bayley Scales of Infant Development) and children (e.g., The Leither International Performance Scale). The mental age that is used for measuring human intelligence in these tests will be adapted to a scale for measuring machine intelligence. This is a fundamental change from the current metrics that measure what a machine can do under a specific setting. What a machine can do under a specific setting is the intelligence of the machine programmer, not the machine itself. For example, an interactive dictionary stores a lot of human knowledge and it can do remarkable things for humans, but it is not intelligent. Test criteria for machine intelligence may also provide quantitative feedback for improving the intelligence tests for humans.

3 Predicted Impacts

The history of science and technology has shown that impressive technical improvement and persistent cost reduction will follow an important scientific revolution. The amount of technical improvement and cost reduction can be so great that it was difficult to foresee at the time of revolution. Two well known examples are the internal combustion engine technology to today's automobiles and Von Neumann machine idea and the semiconductor technology to today's popular computers. The following predictions may seem to be overly optimistic today, but the history could prove them to be true.

3.1 Human life

This revolution will greatly improve the quality of human life. The introduction of engines greatly relieved humans from hard *manual labors*. The introduction of computers greatly relieved humans from mechanical *computation labors*, especially those that humans cannot do as fast, such as doing calculations, controlling a complex machine or generating synthetic graphics images in real time. The introduction of developmental robots could relieve humans from tedious *thinking labors*. Those are low-level thinking tasks, mainly to execute human high-level commands. The quality of human life could be greatly improved with the arrival of the age of developmental robots. Developmental robots will be used as human assistants, from factories to households. Their developed "brains" are downloaded as software to be run on desk-top computers to do various tasks, from reading emails to helping children to learn. In the past, thinking robots have been only discussed in science fiction because machine thinking has not been sufficiently understood. Thinking seems a collection of internal behaviors of a *developmental being* (animal or machine) and it must be developed through autonomous mental development just like humans and higher animals. Infants think using their simple internal behaviors and adults think using their more developed internal behaviors. A robot that runs a developmental program is like a machine that writes mental program autonomously, when the developmental program interacts with the sensory information from the real world. Its developed internal behaviors represent the true thinking by a machine.

Why did all these advances not occur in the past? This is mainly because the AI field did not pay sufficient attention to, or at least was not serious about, autonomous mental development for machines until just a few years ago. Currently, all the efforts for building AI systems follow the traditional manual development paradigm, with a few recent exceptions mentioned above. With the new paradigm, human programmers are not required to write a particular program for each of the tasks that we want the machines to perform, which has been proved extremely difficult if the task requires what we consider as intelligence. Instead, what the human programmers need to do is to write a developmental program, which is of general purpose. Although developmental programs are by no means easy to design, they are easier to understand and to improve than many special systems designed for specific AI tasks. The practical aspect of developmental robots also rests in the ease of training. The user of a developmental robot does not need to write a program or manually feed data if he wants to teach the robot. He just trains the robot very much like the way he trains a human child, showing it how to do something while talking to it, encouraging or discouraging what the robot does from time to time. Thus, everybody can train a highly improved developmental robot, a child, an elderly, a teacher, a worker — anybody. This is the basic reason why the developmental robots could become popular. Computers would not have been that popular today if they are not as easy to use as today's computers with very intuitive graphical user interfaces.

3.2 Economy and jobs

The economic impact of developmental robots will be enormous. The country that takes the lead in developmental robots will first create a new industry for this new kind of machine. This new industry will take advantage of the advanced automobile industry to develop sensor-rich humanoid robots (Honda in Japan has already started it). It will also take advantage of the fast progress of the computer industry to build computers and memories best suited for the computational need of developmental robots. The cost for large storage will drop consistently when the market grows. For example, the cost of hard-disk storage that is of human brain size in terms of number of bytes has already dropped from about $5M in 1998 to around $250,000 today (June 2000). Real-time speed with large memory is reached through coarse-to-fine memory search schemes. There will be a new industry for humanoid robots, fueled by the need for building bodies for developmental robots. Many different types of bodies, designed for different working conditions and environments will be made to satisfy increased application scope of developmental robots. It is expected that in the next 10 to 20 years, the developmental robot industry will primarily aim at professional applications, such as research institutions, amusement parks, public service areas, and the defense industry. During this period, consumers can benefit from the software that is developed on professional robots. Eventually, developmental humanoid robot may cost the same as a car plus a high-end personal computer. The country that takes the lead in this new endeavor will create an abundance of economical activities and well-paid jobs related to this new industry.

3.3 Understanding of human mind

The impact on the scientific understanding of our mind will be far reaching. This revolution will drastically improve our understanding about one of the most complex subjects that faces mankind today — our own minds. For example, what are the basic mechanisms that govern the ways in which our minds develop? To what degree can the environment change the formation of the mind? What can the environment do to effectively and positively influence the human mind and improve the life of mankind? The answers to these questions require the knowledge about the developmental root of the mind. Without studying the computational models of mental development, these questions cannot be sufficiently and clearly answered.

3.4 Medicine

The knowledge created by this revolution will also improve medical care. It will provide basic knowledge useful for treating learning disabilities, mental disorders, and mental problems associated with aging. For example, what developmental mechanisms are responsible for attention deficiency? What developmental mechanisms are responsible for enabling an individual to establish the value of an event, a behavior, or the social norm? What techniques are effective for teachers to improve the development of certain cognitive and behavioral capabilities? Computationally, which areas of the brain are responsible for certain mental disorders? During aging, which mechanism of the brain is likely to deteriorate first and what remedies are possible?

4. Why now?

As we discussed above, the recent new discoveries about human brain tell us loud and clear that our human brain utilizes the developmental principles that are shared by different sensing and effector modalities. Since higher brain functions appear to be even more plastic than early sensory processing, it is expected that the higher brain functions also use developmental principles that are generally applicable to different subject matters that humans learn. The time is right to study what these developmental principles really are.

Technically, it is now possible to study massively parallel, distributed brain activities and relate them to mental development. The advances in neural imaging techniques, such as EEG, EMG and fMRI, now allow high resolution, concurrent, and real-time measurement of brain activities.

In the machine intelligence and robotics fields, the fundamental difference between the way human mind is developed and the traditional engineering paradigm for machine development was recently identified as the fundamental reason for the difficulties in AI. The studies about the fundamental limitations of the current engineering paradigm have recently started.

Some preliminary computational models for developing the mind by machines were recently proposed the tested. These early efforts have achieved some results that have not been possible using the traditional engineering paradigm. Therefore, computational models of mental development for machines are not beyond human comprehension and they are within the manageable scope for humans to model computationally.

The performance-to-cost ratio of computers has reached a critical level that now it is practical to simulate brain development in real time on a robot, with a storage whose size is equivalent to a considerable fraction of human brain. Further, this can now be done at a very moderate cost. For example, the development of the most computational challenging modality, vision, can now be simulated on real robot in real time by software running on a PC workstation.

Technology for building robots has also been improved significantly. In recent years, research laboratories and related industries in US and Japan gained remarkable experience in actually building robots that resemble human and animal bodies with similar articulate structures, from human-size humanoid robots (e.g., the series of Honda humanoid robots) to advanced consumer toy robots (e.g., Sony AIBO dog robots). The robotic technology is ready for building various humanoid or animal robots as bodies for developmental machines.

5 Research issues

In some sense, the task-nonspecific nature of mental development makes the studies of mental development easier than the traditional task-specific approaches. This is true for both human subject (neuroscience and psychology) and machine subject (AI and robotics). From the computational view of mental development, the research issues are around sensory signals and effector signals with internal autonomously generated numerical states. A developmental program will associate signals that are from different sensors, stored in internal status and sent to effectors, but its programmer does not need to know what those signals actually mean! To put it intuitively, it is easier to model how an interactive program looks up words from its word memory than to model how the meanings of words in The Merriam-Webster's Dictionary relate to one another. The former is like what a developmental program does for many tasks that a developmental being will come across and the later is like what all the traditional programs do for a particular task.

To understand this fact better, we take a complex behavior as an example. Modeling attention selection in a traditional task-specific way requires the researcher to understand the nature of the task (e.g., driving) and then to study the rules of attention selection based on the steps of the task. Such rules are extremely complex (e.g., due to the complex road situation during driving) and the results are ad hoc in the sense that they are not directly applicable to other tasks or even to the same task under different scenarios. In contrast, attention selection by a developmental being is just a part of behaviors that are being developed continuously and constantly. As long as the effectors for attention selection are defined for the body (external effector) and the brain (internal effector), the attention selection principles are developed autonomously by the same developmental program in a way very similar to the behaviors for other effectors, such as arms and legs.

Consequently, a series very interesting and yet manageable new research problems are opened up for study, for fields that have either human or machine as study subjects. Some of the tractable research problems that can be immediately studied are suggested below.

1. Schemes for autonomous derivation of representation from sensory signals (from the environment and the body).
2. Schemes for autonomous derivation of representation from effector signals (from the practice experience)
3. Autonomous derivation of receptive fields, in both the classic and nonclassic sense. That is, how later processing elements in the brain group outputs from earlier processing elements or sensory elements.
4. Long term memory growth, self-organization and retrieval, for high-dimensional neural signal vectors.
5. Working memory formation and self-organization, for high-dimensional neural signal vectors. The working memory may include short term sensory memory and the system states.
6. Developmental mechanisms for mediation of conscious and unconscious behaviors. That is, those for mediation among higher

and lower level behaviors, such as learned behaviors, learned emotional behaviors, innate emotional behaviors and reflexes.

7. Mechanisms for developing internal behaviors — those that operate on internal nervous components, including attention selection. This subject includes both developmental mechanisms and training strategies for humans and robots.
8. Attention-directed time warping from continuous states. This subject deals with time inconsistency between different instances of experience, with the goal of both generalization and discrimination.
9. Autonomous action imitation and self-improving. The developmental mechanisms for a developmental being to derive an improved behavior pattern from individual online instances of related experience.
10. Mechanisms for *communicative learning* and thinking. The developmental mechanisms that allow later learning directly through languages (auditory, visual, tactile, written etc) as children do when they attend classes. These mechanisms enable development of thinking behavior, which is responsible for planning, decision making and problem solving.

6 Performance metrics

The current fragmentation landscape of AI is a reflection on how different AI problems can be measured by very different metrics, if intelligence is measured as the capability of performing a specific task. However, what a machine can do under a specific setting represents the intelligence of the machine programmer, not necessarily the machine's own intelligence. Further, a special purpose machine that can only work for a particular problem cannot deal with complex problems that require true intelligence, such as vision, speech and language capabilities.

The *criteria for measuring machine intelligence* will fundamentally change. The metrics that can be used to measure the power of developmental robots should emphasize the autonomous interactive learning capabilities in complex human environment. In other words, it is the capability of mental development instead of what the machine can do under a pre-specified setting.

This is indeed the case with well-accepted test scales used by clinical psychologists for measuring mental and motor scales of human children. Two such well known scales are The Bayley Scales of Infant Development (for 1 to 42 months old) and The Leither International Performance Scale (for 2 years to 12 years old). These scales have a very systematic methodology for the administration of tests and scoring. The reliability and calibration of these scales have been supported by a series validity studies, including constuct validity, predictive validity, and discriminant validity that cover very large number of test subjects and different age groups across very wide geographic, social, and ethnic populations.

Here let us take a look at an example of tests in the Leither International Performance Scale for a two years old. The name of the test is Matching Color. The test setup is a row of 5 stalls. Above each stall pasted a color card, black, red, yellow, blue, and green, respectively. During the test, color blocks are presented, one at a time in the order: black, red, green, blue, and yellow. The examiner places the black block in the first stall and tries to get the subject to put the red block in place by placing it on the table before him, then in the appropriate stall, then on the table again, nodding to him to do it and at the same time pointing to the second or red stall. As soon as the subject begins to take hold of the test, the final trial can be attempted. In this test, the examiner tries to get the subject to imitate his procedure. The test is scored as passed if the subject is able to place the four colors (the first one is placed by the examiner) in their respective stalls entirely by himself during any one trial, regardless of the number of demonstrations or the amount of help previously given by the examiner. As we can see, the test does not really concern about whether the child has learned the abstract concept of color, but rather the capability of imitating the action of the examiner using visual color information as a cue in coordination of his motor effectors (hand and arm).

The mental age that is used for measuring human intelligence in these tests can be used as a scale for measuring machine intelligence. Currently metrics that have been used for various AI studies mainly measure what a machine can do under a specific setting, instead of the capability of mental development. Such a capability requires online, interactive learning capability as the above test demonstrates. For example, an interactive dictionary stores a lot of human knowledge and it can do remarkable things for humans, but it is not intelligent. If a machine that can pass the systematic tests like the one shown above, it must have already learned many others skills that no traditional machine has. Therefore, although autonomous mental development is a new direction, its impact on the future of machine intelligence and our understanding of human intelligence will be far reaching. The performance metrics for measuring intelligent machines can be adapted from those used by clinical psychologists for testing the mental development of human infants. The Bayley Scales of Infant Development and The Leither International Performance Scale are two such examples.

References

[1] T. S. Kuhn. *The Structure of Scientific Revolution*, University of Chicago Press, third addition, page 68, 1996.

[2] L. von Melchner, S. L. Pallas and M. Sur. Visual behavior mediated by retinal projections directed to the auditory pathway. *Nature*, vol. 404, April 20, pages 871-876, 2000.

[3] X. Wang, M. M. Merzenich and K. Sameshima and W. M. Jenkins. Remodeling of hand representation in adult cortex determined by timing of tactile stimulation. *Nature*, vol. 378, no. 2, pages 13-14, 1995.

[4] J.L. Elman, E. A. Bates, M. H. Johnson, A. Karmiloff-Smith, D. Parisi and K. Plunkett. *Rethinking Innateness: A connectionist perspective on development.* MIT Press, Cambridge, MA, 1997.

[5] E. Thelen, E., G. Schoner, C. Scheier, and L.B. Smith (In press). The dynamics of embodiment: A field theory of infant perseverative reaching. *Behavioral and Brain Sciences*, to appear.

[6] J. Weng. The Living Machine Initiative. Department of Computer Science Technical Report CPS 96-60, Michigan State University, East Lansing, MI, Dec. 1996. A revised version appeared as a chapter: J. Weng. Learning in Image Analysis and Beyond: Towards Automation of Learning, in *Visual Communication and Image Processing*, C. W. Chen and Y. Q. Zhang (eds.), Marcel Dekker, New York, NY, 1998.

[7] N. Almassy, G. M. Edelman and O. Sprons, Behavioral constraints in the development of neural properties: A cortical model embedded in a real-world device. *Cerebral Cortex*, vol. 8, no. 4, pages 346-361, 1988.

[8] J. Weng, W. S. Hwang, Y. Zhang and C. Evans, Developmental robots: Theory, Method and Experimental Results, in Proc. 2nd Int'l Symposium on Humanoid Robots, Tokyo, Japan, pp. 57- 64, Oct. 8- 9, 1999.

[9] D. Roy, B. Schiele, and A. Pentland. Learning Audio-Visual Associations using Mutual Information. In Proc. *International Conference on Computer Vision, Workshop on Integrating Speech and Image Understanding*. Corfu, Greece, 1999.

[10] M. Huber and R. A. Grupen. A feedback control structure for on-line learning tasks. *Robotics and Autonomous Systems*, vol. 22, no. 3-4, pages 303-315, 1997.

[11] H. Gardner. *Multiple intelligences: The theory in practice*. Basic Books, New York, NY, 1993.