Orientation Behavior Using Registered Topographic Maps

Cynthia Ferrell *
Massachusetts Institute of Technology
Artificial Intelligence Laboratory
545 Technology Square, Room 819
Cambridge, MA 02139 USA
voice: (617) 253-7007
fax: (617) 253-0039
email: ferrell@ai.mit.edu

Abstract

The ability to orient toward visual, auditory, or tactile stimuli is an important skill for systems intended to interact with and explore their environment. In the brain of mammalian vertebrates, the Superior Colliculus is specialized for integrating multi-modal sensory information, and for using this information to orient the animal to the source sensory stimuli, such as noisy, moving objects. Within the Superior Colliculus, this ability appears to be implemented using layers of registered, multi-modal, topographic maps. Inspired by the structure, function, and plasticity of the Superior Colliculus, we are in the process of implementing multi-modal orientation behaviors on our humanoid robot using registered topographic maps.

In this paper, we explore integrating visual motion and oculomotor maps to study experience-based map registration mechanisms. Continuing work includes incorporating self-organizing feature maps, including more sensory modalities such as auditory and somatorsensory maps, and extending the motor repertoire to include the neck and body degrees of freedom for full-body orientation.

1 Introduction

The ability to orient to sensory stimuli is an important skill for autonomous agents that operate in complex, dynamic environments. In animals, orientation behavior serves to direct the the animal's eyes, ears, nose, and other sensory organs to the source of sensory stimulation. By doing so, the animal is poised to assess and explore the nature of the stimulus with complementary sensory systems, which in turn affects and guides ensuing behavior. Hence, orientation behavior is performed frequently and repeatedly by agents that are tightly coupled with their environment, where perception guides action and behavior assists in more effective perception.

Certainly, orientation behavior is a basic skill we would like to implement on Cog, our humanoid robot (Brooks & Stein 1994). We would like Cog to perform a variety of tasks, many of which fall under two broad behavioral themes: exploratory behavior and social skills. For example, orienting the body to an object of interest assists manipulation tasks by putting the object where it is most accessible to sensory and motor systems. Eventually the work presented in this paper will be integrated with the ability to reach for visual targets (Marjanovic, Scassellati, & Williamson 1996). The same is true for social skills where the robot should position itself so that the person is easy to interact with.

Our approach to implementing orientation behavior on Cog is heavily inspired by relevant work in neuroscience (Stein & Meredith 1993), (Brainard & Knudsen 1993). In the brain of mammalian vertebrates, the Superior Colliculus is an organ specialized for producing orientation behavior. In non-mammalian vertebrates (birds, amphibians, etc.), the optic tectum is the analogous organ. The structure of the Superior Colliculus is characterized by layers of topographically organized maps. Collectively, they represent the sensorimotor space of the animal in ego-centered coordinates. These maps are interconnected and interact is such a way that the animal performs orientation movements in response to sensory stimuli.

Topographically organized maps have been discovered throughout the brain of mammalian vertebrates. In addition to the Superior Colliculus, they have been identified in various perceptual areas of the neocortex (the visual, auditory and somatosensory corticies, for instance). It is widely recognized that the organization of these maps are plastic and can be shaped through experience.

*Support for this research was provided by a MURI grant under the Office of Naval Research contract N00014–95–1–0600.

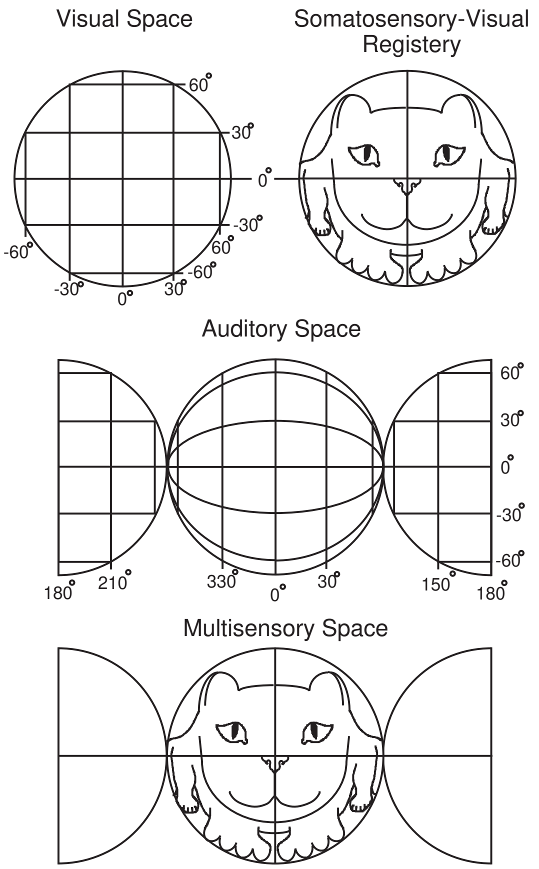

Figure 1: The Superior Colliculus is organized into layers of topographic maps. A variety of sensory maps, motor maps, and multi-modal maps have been discovered. These maps are registered with one another to share a common multisensory coordinate system. This figure illustrates registered visual, auditory, and somatorsensory spatial representations. Adapted from (Stein & Meredith 1993).

Subsequently, cortical maps have garnered a lot of attention, and a variety of work has explored the phenomena of self-organizing feature maps, (Kohonen 1982), (Ritter & Schulten 1988), (Obermayer, Ritter, & Schulten 1990).

Through implementing something like the Superior Colliculus on Cog, this paper explores how topographically organized maps develop and interact to produce a unified observable behavior. There are a variety of topics this paper explores in relation to this endeavor. From a behavioral perspective, we explore how dynamic spatio-temporal representations of sensory and motor space can be used to integrate multi-modal information and produce coherent behavior. From a developmental perspective, we investigate experience dependent mechanisms by which these maps self-organize and interconnect with one another. Given that behavioral experience affects both the connectivity within and between maps, and that the current state of connectivity dictates behavioral performance, we are dealing with a coupled system where the dynamics of forming connections affects and is effected by behavioral performance. Eventually, we would like to explore the dynamics of development where we expect to observe different developmental time scales of map self-organization and inter-map integration as the orientation behavior performance improves.

The rest of this paper is organized as follows. First we will briefly cover the organization, structure, and function of the Superior Colliculus, as our implementation is strongly inspired by what is understood about this organ. Next, because the physical architecture of the robot and the computer places heavy constraints on our implementation, we describe Cog, the experimental platform used in our experiments. After this, we present the state of our implementation at the time this paper was written, as well as extensions currently under development. Finally, we present tests and results of our system to date, and conclude with a brief description of ongoing work and future directions.

2 The Superior Colliculus

The Superior Colliculus is a midbrain structure composed of seven laminar layers. The deep layers are those believed to play a role in orientation behavior. Its physiology reflects its primary role as integrating different modalities to evoke motor responses. For example, among its many different afferent and efferent connections, it receives inputs from several sensory modalities such as the visual, auditory, and somatosensory corticies, and sends outputs to brain stem and spinal cord.

An important function of the Superior Colliculus is to pool sensory inputs from different modalities and redirect the corresponding sensory organs (eyes, ears, nose) to fixate on the source of the signal. Through the convergence of sensory inputs, the Superior Colliculus gives different sensory modalities access to the same motor circuitry so that any sensory modality can be used to direct the other sensory modalities to the source of the stimulus. For instance, by doing so, the animal can hear a sound emanating from an object outside its visual range (perhaps coming from behind the animal) and quickly turn its head and eyes to foveate on whatever is making the noise.

2.1 Organization of the Superior Colliculus

Localized regions of the Superior Colliculus (or Optic Tectum) consist of neurons with receptive fields that form topologically organized maps. A variety of topological maps have been identified in several species (cats, monkeys, owls, electric eels, and frogs to name a few). Each map corresponds to either a single modality or a combination of modalities. The modalities represented by the maps varies between species, depending on those sensory or motor systems used in orientation behavior.

In the cat, there are visuotopic maps representing motion in visual space, somatotopic maps yielding a body representation of tactile inputs, and spatiotopic maps of auditory space encoding inter-aural time differences (ITD) and inter-aural intensity differences. Hence, a sensory stimulus originating from a given direction will elicit activity in the corresponding region of the appropriate sensory map. For example, an object moving in the visual field causes the corresponding region in the visuotopic map to become active. There are also motor movement maps consisting of pre-motor neurons whose movement fields are topologically organized. In the cat, these exist for the eyes, head, neck, body, ears. For example, stimulating a specific region in the oculomotor map elicits a movement to fixate on the corresponding area.

2.2 The Role of Map Registration

These multi-modal maps overlap and are aligned with each other so that they share a common multisensory spatial coordinate system. The maps are said to be *registered* with one another when this is the case. Arranging multi-modal information into a common representational framework within each map and aligning them allows the information to interact and influence each other. There are several advantages to this organizational strategy. First, it is an economical way of specifying the location of peripheral stimuli, and for organizing and activating the motor program required to orient towards it; thereby allowing any sensory modality to orient the other sensory organs to the source of stimulation. Second, it supports enhancement of simultaneous sensory cues. Stimuli that occur in the same place at the same time are likely to be interrelated by common causality. For instance, a bird rustling in the bushes will provide both visual motion and auditory cues. During enhancement, certain combinations of meaningful stimulus become more salient because their neuronal responses are spatio-temporally related. Once the multi-modal maps are aligned, neuronal enhancement (or depression) is a function of the temporal and spatial relationships of neural activity among the maps.

2.3 Development and Experience Dependent Plasticity

During development, the organization of the topographic maps is plastic. For each map, its representation of space is use dependent. In monkeys, it has been found that the organization of somatosensory maps for the hand can be changed by varying the amount and location of stimulation(s). Experiments have shown that the size of the map region corresponding to a particular cutaneous region on the hand is correlated to how much stimulation that part of the hand receives over time. Furthermore, adjacent regions of the map correspond to regions on the hand that are temporally adjacent (Stein & Meredith 1993). This phenomena has been seen in other perceptual areas of the cortex and is typical of self-organizing feature maps (SOFMs). A number of people have modeled this phenomena using neural networks (Kohonen 1982), (Bauer & Pawelzik 1992), (Durbin & Mitchison 1990).

It has also been found that the registration between maps is malleable over the developmental period. This phenomena has been studied in the inferior and superior colliculus of young barnyard owls, where the registration of the auditory map to the visual map shifts according to experience. (Brainard & Knudsen 1993), (Brainard & Knudsen 1995) found that the visual map is used to train the auditory map so that the auditory map shares the same coordinates as the visual map. Even if the visual map is artificially shifted by mounting distortion spectacles on the owls, the auditory map also shifts to keep in register with the visual map.

3 Cog: The Experimental Platform

Figure 2: Cog is the humanoid robot used in our experiments, shown on the right. The "brains" of cog is a MIMD computer shown in the center of this image. The contents of DPRAMs (images, processed images, maps, etc.) can be displayed on a bank of twenty displays shown on the left. The Macintosh Quadra is the front end to the MIMD computer.

This section presents an overview of Cog, the humanoid robot used in our experiments. We briefly describe the implementation of the robotic platform, the perceptual systems, the computational system, and the software systems relevant to this paper. All the systems described below were designed and constructed by members of the Cog Project (see acknowledgements).

3.1 The Robotic Platform

A fundamental design goal for the robot was to make it anthropomorphic as possible so that 1) the robot could move in a human-like manner, and 2) encourage humans to interact with it in a natural way. The most important human characteristics to emulate are size, speed, and range of motion. Hence, Cog resembles a human from the waist up and is shown in figure 2.

Cog's body has six degrees of freedom (DOF): the waist bends side-to-side and front-to-back, the "spine" can twist, and the neck tilts side-to-side, front-to-back, and twists left-to-right. Mechanical stops on the body and neck give a human-like range of motion. In addition, each degree of freedom has current sensing in the motor controller to provide some force feedback, temperature sensing to determine a longer term time-average of how hard the motors are working, and joint encoders to provide a proprioceptive sense.

Cog's head is equipped with a compact, binocular, active vision system. To maintain an anthropomorphic appearance, the "eyes" were mounted about 3 inches apart. To mimic human eye movements, each "eye" can rotate about a vertical axis (pan DOF) and a horizontal axis (tilt DOF). To approximate the range of motion of human eyes, mechanical stops were included on each eye to permit a 120° pan rotation and a 60° tilt rotation. In addition, each eye performs fine motor control and high-speed positioning so that we may emulate human visual behaviors.

3.2 The Perceptual Systems

To give the robot both a wide field of view and a high resolution foveal area, each eye consists of two black and white CCD cameras. We could have simplified our design by using a single camera per eye. However, by using two cameras per eye we have a much higher resolution fovea than the single camera eye. The lower camera of each eye gives Cog a wide peripheral field of view ($88.6°(V) \times 115.8°(H)$ FOV), and the upper camera of each eye gives Cog a high resolution fovea ($18.4°(V) \times 24.4°(H)$ FOV).

In addition to the visual system described above, Cog has several other perceptual systems under concurrent development. To date, we have developed an auditory system (Irie 1995), a vestibular system, and a variety of force resistive sensors to give Cog a tactile sense. These systems are in the process of being ported to the Cog platform.

3.3 The Computational Platform

This section summarizes the computational system we developed to meet Cog's requirements. First, the system must allow for real-time control of the robot since the robot operates in a dynamic environment full of people and objects we would like the robot to interact with. Furthermore, the system must be robust, scalable, concurrent, and support learning and development processes.

Cog's "brain" is a scalable MIMD computer consisting of up to 256 processor nodes. Currently the nodes are based on the Motorolla 68332, but Texas Instruments C32 DSP nodes are under development which will be responsible for the bulk of perceptual processing. Processors can communicate through eight ports to other processor nodes or to other parts of the video capture/display system. All components of the processing system communicate through dual-ported RAM (8K by 16 bits) connections, so altering the topology is relatively simple. During operation, the brain is a fixed topology network which can be changed manually and scaled by physically adding additional nodes and dual-ported RAM connections. The entire brain is connected through a serial line with a Macintosh Quadra, which is used for communication and input, but not for any actual processing. Each node also uses standard Motorola SPI (serial peripheral interface) to communicate with up to 16 motor controller boards.

The video capture/display system consists of custom designed frame grabbers and display boards. A frame grabber takes the NTSC signal from one of the eye cameras, digitizes the input, subsamples it to 128 by 128 pixels, and writes the resulting grayscale values to six ports. For this design, we have chosen to use only 128 by 128 grayscale in order to reduce the amount of data to be processed and to increase the speed of the visual processing. The frame grabbers operate at approximately 30 frames per second, and can write simultaneously to six processors. The display boards produce NTSC video output for display on a bank of 20 display monitors. Direct camera output and digitized output of the frame grabbers can also be routed directly to the monitor bank.

A network of special purpose motor controller boards mounted on the robot act as Cog's "spinal cord", connecting the robot's "brain" to the rest of the body. Each motor has a dedicated motor controller board that reads the encoder (and other sensors), performs servo calculations, and drives the motor. The motor controller boards have hardware that generates a 32KHz PWM waveform. The duty cycle is updated at 2KHz by an on-board MC6811E2 microcontroller. Currently the microcontroller implements a PID control law for position and velocity control. Position and velocity commands are sent to the motor controller boards from the MIMD computer described above.

3.4 The Software Environment

Each processor has an operating system; *L*, a compact, downwardly compatible version of Common Lisp that supports real-time multi-processing (Brooks 1994*a*); and *MARS*, which is a front end to *L* that supports communication between multiple processes on a single processor as well as communication between processes running on separate processors (Brooks 1994*b*). *MARS*, like the *Behavior Language* (Brooks 1990), is a language for building networks of concurrently running processes. The processes can communicate either locally by passing messages over virtual wires, or globally through a process inspired by hormonal mechanisms. *MARS*, unlike the *Behavior Language*, supports on-line learning mechanisms by allowing the network morphology to change dynamically, i.e. spawning or killing processes or changing network connectivity during run-time.

4 A Developmental Approach to Orientation Behavior

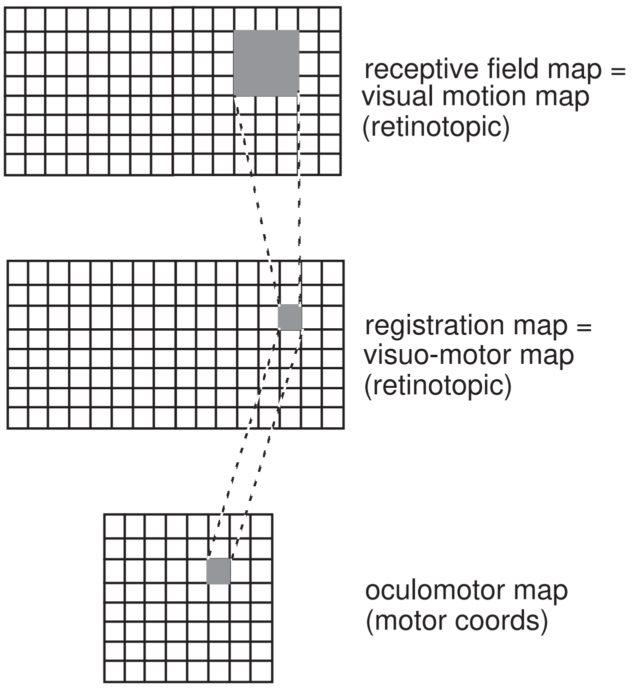

Figure 3: Example of layered topographic maps implemented on Cog. The multi-modal registration map acts to relay activity in the receptive field map (the visual motion map) to activate the oculomotor map such that a visual stimuli is foveated. See section 4.1 for further explanation.

Registered topographic maps form a substrate upon which multi-modal information can be integrated to produce coherent behavior. How are these topographic maps formed? How do they become registered with one another? How is the organization of the ensemble guided by experience?

4.1 The Framework

In our framework, a map is a two dimensional array of elements where each element corresponds to a site in the map. The maps are arranged into interconnected layers, where a given map can be interfaced to more than one map. Each connection is uni-directional, so recurrent connections between maps require both a feedforward connection and a feedback connection. The activity level of sites on one map is passed to another map thorough these connections, hence the input to a given map is a function of the spatio-temporal activity of the maps feeding into it and the connectivity between these maps. Currently, all connections have equal weights, although this could change in the future. The output of a given map is its spatio-temporal activity pattern. What this pattern of activity represents depends upon the map: if it is a visuotopic map, it could represent motion coming from a particular direction in the visual field; if it is an oculomotor map, it could encode a motor command to move the eyes, and so forth.

The smallest map ensemble capable of producing an observable behavior consists of a sensory input map, a motor output map, and an established set of connections between them. The input map could have a fairly rigid structure consisting simply of time-differenced intensity images. Because visual information already contains a spatial component, this simple map is topographic without any additional tuning. The motor map could also be fixed where a given site on the map corresponds to a given motor command. If the motor commands vary linearly with motor space, for instance, this map is also topographically organized. Assuming the cameras are motionless, a moving object occupies a localized region in the visual field, and correspondingly causes a localized intensity difference (an active region) in the time-differenced image map. If there exists connections from this region of the time-difference map to the appropriate region of the oculomotor map, then a motion stimulus in the visual field activates the corresponding region of the time-difference map, which in turn excites the connected region of the oculomotor map, which evokes the necessary camera motion to foveate the stimuli.

4.2 Developmental Mechanisms

Plasticity can be introduced into the simple system above in two ways: 1) the map organization could change so that a given map site could correspond to different locations in space. 2) The connections between maps could change so that a given site could change which site(s) it connects to on the other map.

In animals, as described in section 2.3, the organization of the maps and the registration between maps is tuned during the critical period of development. Several mechanisms and models have been proposed to account for this organizational process. The mechanisms we use for map organization and alignment on Cog are inspired by similar mechanisms (Kohonen 1982). However, different combinations of mechanisms are used depending on what is being learned: i.e. tuning the organization within a map, registering different sensory maps, or registering sensory maps and motor maps.

A variety of mechanisms determine how map connections are established. Guided by sensori-motor experience, these mechanisms govern how connections are modified to improve behavioral performance.

- *Competition*: There is competition between concurrently active sites where only the most active site is modified per trial. In our system, the most active is currently approximated as the centroid of activity of the active region. Furthermore, each site of a given map can only form a limited number of connections to the other map. So, candidate map sites compete to determine those that can connect to a given site on the other map.

- *Locality, neighborhood influences*: The neighboring sites around the most active site are also updated each trial. The amount a neighboring site is adjusted decays with distance from the maximally active site. This mechanism penalizes long connections and encourages topographic organization. The size of the neighborhood can vary over time. Typically, it starts off fairly large until the map displays some rough topographic organization, then it decreases as the map undergoes fine tuning adjustments.

- *Error correction*: It is not sufficient that the maps are topographically organized and aligned – they must be organized and interfaced so that the agent performs well in its environment. For tight feedback loop sensorimotor tasks (such as saccading to a visual stimulus), an appropriate error signal is very important and useful for tuning the behavior of the system. Naturally, the error signal must be a good measure of performance and obtainable at a fast enough rate to enable on-line learning. Connections are modified to reduce the discrepancy between current performance and desired performance. The magnitude of the correction is proportional to the size of the error on that trial.

- *Correlated temporal activity*: Hebbian mechanisms are often used for self-organizing processes. By strengthening connections between simultaneously active sites, they are useful for relating information between different sensory maps.

- *Learning rate*: The magnitude of the adjustment for each trial is also proportional to the learning rate. The learning rate can vary over time, where it starts of relatively large for course tuning, and then decreases for finer adjustments.

4.3 A Simple Behavior

In this section we look at an example to see how these mechanisms are applied to forming and organizing these multiple maps to perform a task. A simple orientation task is the ability to saccade to noisy, moving stimuli (clapping hands, shaking a rattle, etc.). We say that a good saccade centers the stimulus in the fovea camera's field of view, whether the stimulus is seen in the wide field of view or the fovea field of view. We assume that the system favors information from the foveal view because it is of higher resolution than the peripheral view and thereby can be used to perform a more accurate saccade.

Experience dependent plasticity could play a role in several ways. It could be used to guide the representational organization of the auditory and visual motion maps, guide the registration between the auditory and visual motion maps, or guide the registration of the the sensory maps to the oculomotor map. Below, these three types of organization are explored in turn:

- Self organizing feature maps
- Registration of sensory maps
- Registration of sensory and motor maps.

Each mapping process can be viewed as learning a multi-modal map that registers the information from two other modality maps. We call the multi-modal map the *registration map*, and the other maps could be either sensory maps, motor maps, or both. One of the modality maps provides the rough spatial organization of the multi-modal map. We call this map the *receptive field map*. Typically it has a topographic representation of space. Often a retinotopic map is used, for instance. The second modality map, which may or may not be topographic, is registered to the first map through the multi-modal map. This process is illustrated in figure 3. Note that each site of a modality map connects to only one site on the registration map, but the same site on the registration map could connect to mulitple sites on the modality maps.

4.3.1 Self Organizing Feature Maps

One example of a self organizing feature map is learning the visual motion map for the peripheral field of view. In this case, the receptive field map is a time-difference map of consecutive intensity images (in retinotopic coordinates) and the registration map is the visual motion map. Initially, the receptive field map contains broad, overlapping receptive fields for the corresponding sites of the registration map. Those connections that are spatio-temporally correlated are strengthened over time, whereas those that are not are weakened and eventually die off. Hence, the primary developmental mechanisms used for this case are competition, neighborhood updates, and hebbian learning.

Recall that the organization of the registration map will reflect how the map is used. The resulting visual motion map should represent that motion in peripheral view that is relevant to behavior. This is not necessarily a direct mapping from the intensity time-difference map. For example, if motion is present in the fovea region, then the system should favor this information over the information coming from the peripheral view. Over time, we would not expect to see the center $20° \times 20°$ of the peripheral field of view represented in the peripheral motion map because this information is not used to perform saccades.

4.3.2 Registration of sensory-sensory maps

An example of aligning sensory maps is registering the auditory map to the visual motion map. In this case, the receptive field map is the visual motion map (in retinotopic coordinates), the registration map is a visuo-auditory map (also in retinotopic coordinates), and the third map is the auditory ITD map. The auditory ITD map could be tonotopically organized, instead of topographically organized, since auditory signals do not contain inherent spatial information. Initially, the visual map and the auditory map are interfaced to the registration map. The registration map contains broad, overlapping receptive fields. Registration of the auditory map with the visual map entails mapping the correct ITD values to the corresponding regions in the registration map that are in turn connected to the visual map where specific locations in visual coordinates are represented by spatial location. During the developmental process, those visual and auditory signals that are spatio-temporally correlated are strengthened, and those that are not eventually die off. Hence there is a significant amount of weeding out of inappropriate connections. Over time, the ITD receptive fields in the registration map become refined and properly located in retinotopic coordinates. Here, the primary developmental mechanisms are competition, neighborhood updates, and hebbian learning.

4.3.3 Registration of Sensory-Motor Maps

An example of aligning sensor and motor maps is registering the oculomotor map with the visual motion map. In this case, the receptive field map is the visual motion map (in retinotopic coordinates), the registration map is a visuo-motor map (also in retinotopic coordinates), and the third map is the oculomotor map (in eye motor coordinates). Regions in the motor map correspond to motor movements that could foveate a stimulus. Initially the visual map and the oculomotor map are connected to the registration map with broad, overlapping receptive fields. When the motion map is stimulated and the site of maximal activity is determined (typically the centroid of the stimulated region), the corresponding region of the oculomotor map is stimulated. The site of maximal response of the motor map is taken as the motor command, and the corresponding motor movement is evoked. This movement orients the eye to the stimulus. Once oriented, the motion stimulus stimulates a different region in the visual motion map. The visual error is computed as the difference from centroid of motion to the center of the field of view. This error is used to update the connections responsible for the orientation movement to reduce the error in the future. Hence, the primary developmental mechanisms are competition, neighborhood updates, and error correction.

5 Architectural Organization

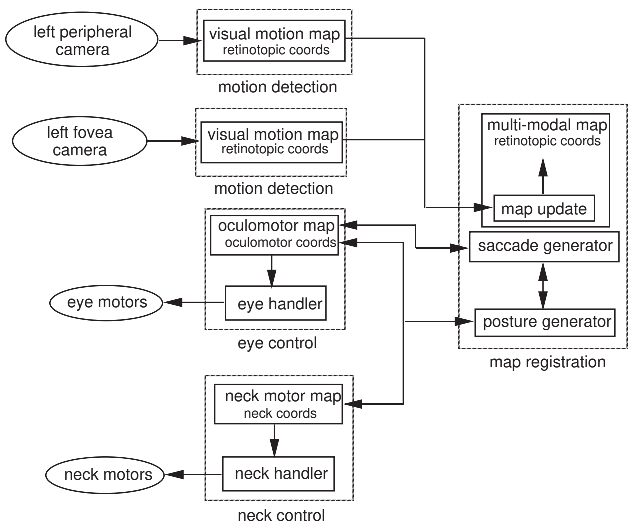

Figure 4: This diagram shows how the multi-modal topographic maps are arranged on Cog's computational hardware. Currently five processors are used: two visual processors, two motor control processors, and one processor which performs the developmental mechanisms. See text for further explanation.

To date, the sensory-motor map registration task has been implemented on Cog's hardware and is shown in figure 4. The diagram shows how the processes are arranged on Cog's MIMD computer. Currently, five processing nodes are used:

- *Peripheral motion processor*: Contains the peripheral visual motion map. It computes the difference between consecutive left peripheral camera images at 15 frames/s. It also determines the most active site (the centroid of motion) and a visual error signal.

- *Fovea motion processor*: Contains the fovea visual motion map. It computes the difference between consecutive left fovea camera images at 15 frames/s. It also determines the most active site (the centroid of motion) and a visual error signal.

- *Registration processor*: Contains the visuo-motor map and carries out the developmental process. It receives motion information from the vision processor and determines which motion information to use. If fovea motion is present, it ignores the information from the peripheral camera. It also translates the most active site on the visual map to the region of activity on the motor map, and passes this information to the oculomotor processor. After the motion is performed, it uses the error signal from the vision processors to update the registration map connections according to developmental mechanisms.

- *Oculomotor processor*: Contains the oculomotor map. Upon receiving the site of activity, it commands the motors to perform the movement. It also sends an "efferent copy" to the registration processor, so the registration processor can ignore visual motion information while the cameras are moving.

- *Neckmotor processor*: Contains the neckmotor map. It is commanded by the registration processor to move the neck around so the motion stimulus is seen from many different places in the visual field. Currently, the neck is primarily used for the training processes. However, soon it will incorporated into the orientation behavior.

6 Tests

To date we have run experiments to test whether the implementation we have described learns the registration between the retinotopic visual motion map and the oculomotor map. A sampling of our results are shown in figures 5, 6, and 7.

So far, experiments have been performed using the left eye only. The system will be extended to handle both eyes when stereopsis and vergence capabilities are implemented. Motion information from both eyes will be fused and used to excite the visuotopic motion map. Conflicts between the eyes will be resolved during this fusion stage. The simplest approach would be to resolve conflicts via a dominant eye mechanism. Another method could involve exciting the visuotopic map with the stonger of the two excitations coming from each eye. These two methods along with other possibilities need to be explored. Most likely, a combination of methods will be implemented.

To learn the registration between the peripheral motion map with the oculomotor map, we trained Cog over a number of trials while it looked at a continuously moving stimulus. At the beginning of each trial, the robot

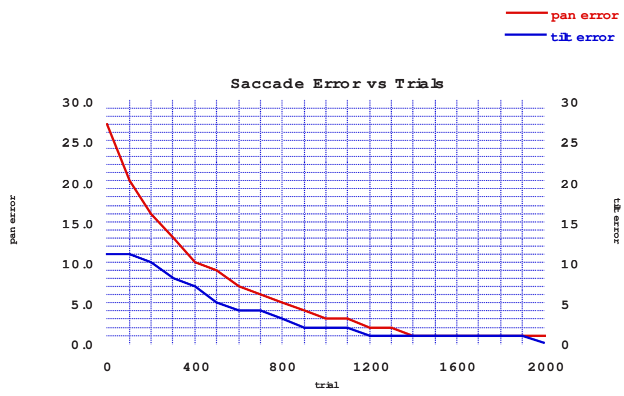

Figure 5: Registration data for aligning the visual motion map and the oculomotor map. The data is derived from the registration map, converting sites in visuotopic coordinates to activated sites in the motor movement map (pan and tilt) required to foveate the stimulus. Initially the maping is random, with a neighborhood radius size of 2 and a learning rate of .25. The neighborhood size remained fixed for all trials in this experiment. At $trial = 0$, the average error over the 20×20 region was approximately $(11°, 26°)$ for pan and tilt DOFs respectively. By $trial = 400$, the average error is reduced to $(10°, 7°)$; it is the slowest to converge to an average error $\leq (1°, 1°)$ of the three experiments shown in this section. By the 1400th trial, the average error was close to 1° for pan and tilt DOFs.

changes its posture (centers its eyes and moves its neck to a random location). This places the motion stimulus in a different location in its visual field. Currently Cog explores the center $20° \times 20°$ of the peripheral visual field, which corresponds to a 20×20 region of the registration map. The robot uses the visual information to stimulate the oculomotor map and perform the saccade. The visual error is then acquired, and the registration between the maps is updated according to the rule:

$$\Delta m(x, y) = \rho \times \epsilon(x, y) \times N(x, y) \qquad (1)$$

where:

- $m(x, y)$ is the value of site (x, y) of registration map m. Recall that this value represents the connection from the visual motion map site to the corresponding oculomotor map site. The learning process involves updating these inter-map connections.

- (x^*, y^*) is the site of maximal activity of the motion map. For this application, it corresponds to the site of maximal activity of the registration map as well.

- ρ is the learning rate.

- $\epsilon(x, y) = target(x, y) - m(x, y)$. It is an error distance measure between the motion map site $m(x, y)$ and

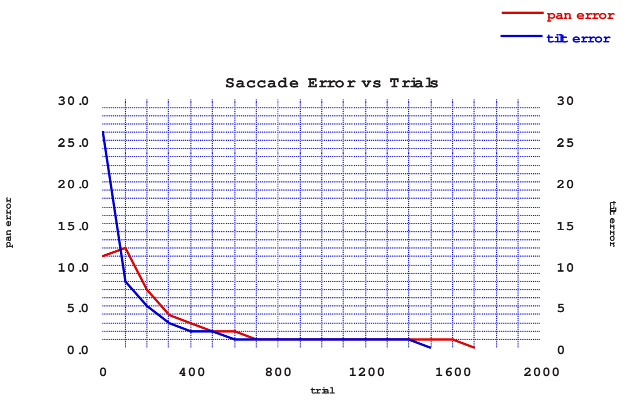

Figure 6: Same experiment as shown in figure 5, except the neighborhood size is manually decreased over time. Initially the maping is random, with a neighborhood radius size of 4 and a learning rate of .25. By $trial = 400$, the average error is reduced to $(3°, 2°)$ and the neighborhood size was set equal to 2. This experiment is the second fastest to converge to an average error $\leq (1°, 1°)$ of the three shown in this section. By the 800th trial, the average error was close to $1°$ for the pan and tilt DOFs.

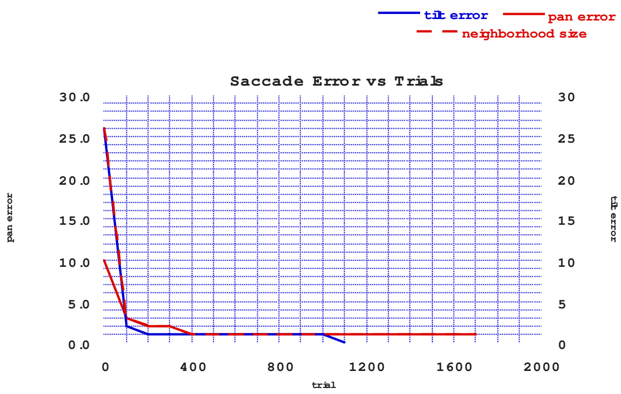

Figure 7: Same experiment as shown in figure 5 and figure 6, except the neighborhood size is automatically decreased over time. Each trial, the neighborhood size is set equal to the larger value of the average error measures (pan or tilt). Initially the maping is random, with a neighborhood size of 26 and a learning rate of .25. By $trial = 100$, the average error is reduced to $(3°, 2°)$, and by $trial = 400$ it is reduced to $(1°, 1°)$. Of the three experiments shown, it is the fastest to converge to an average error $\leq (1°, 1°)$.

the target site. This measurement is made after the saccade motion finishes. Note that $target(x, y)$ is the center of the field of view for the saccade learning task.

- r is the neighborhood radius.
- $N(x, y) = f(1 - \frac{|(x^*, y^*) - (x, y)|}{r})$. It is the neighborhood update function that decays linearly with distance from the site of maximal activity (x^*, y^*). Threshold function, f, sets the result equal to zero if its argument is negative. So, for site locations outside radius r of (x^*, y^*) , $N(x^*, y^*) = 0$.

7 Continued Work

Currently, we are extending these tests to include the full visual field, and continuing our experiments with dynamically varying neighborhood size and learning rate parameters. Soon we will explore the self organizing properties of the representation of visual information by implementing a SOFM for visual motion. We will also begin efforts to register the visual motion with an auditory ITD map, as well as investigate the dynamics of development when self-organization and registration mechanisms are run simultaneously. We expect to see evidence for different developmental time scales as the robot learns the orientation task.

With the above work in place, we will extend the system to include the neck and body degrees of freedom so that the robot can perform full body orientation behavior. This will complicate the current task by adding additional degrees of freedom that must complement each other. We will continue to investigate the issue of developmental time scales since more complicated behaviors will have to develop incrementally and bootstrap off of existing behaviors. We would like to integrate the full orientation behavior with reaching and manipulation tasks currently under parallel development by other members of the group (Marjanovic et al. 1996).

8 Conclusions

This paper describes an implementation of orientation behavior on Cog using registered topographic maps. We have presented biological evidence that this is an effective method for orienting to multi-modal stimuli in animals. We have also presented a series of mechanisms and methods for developing this behavior on Cog over time. This biologically inspired framework gives us the opportunity to explore several interesting issues. It allows us to investigate using dynamic spatio-temporal representations of sensory-motor space to integrate multi-modal information and produce a unified behavior. It also allows us to investigate the dynamics of development using mechanisms of experience dependent plasticity. Ongoing work is promising.

Acknowledgements

A number of people have made this work and the Cog Project possible. Among those who have contributed to the ongoing development of Cog are (in alphabetical order): Mike Binnard, Rod Brooks, Robert Irie, Eleni Kapogannis, Matt Marjanovic, Yoky Matsuoka, Brian Scasselatti, Nick Shectman, Rene Schaad, and Matt Williamson.

References

Bauer, H. & Pawelzik, K. (1992), 'Quantifying the Neighborhood Preservation of Self-Organizing Feature Maps', *IEEE Transactions on Neural Networks* **3**(4), 570–579.

Brainard, M. & Knudsen, E. (1993), 'Experience-dependent Plasticity in the Inferior Colliculus: A Site for Visual Calibration of the Neural Representation of Auditory Space in the Barn Owl', *The Journal of Neuroscience* **13**(11), 4589–4608.

Brainard, M. & Knudsen, E. (1995), 'Dynamics of Visually Guided Auditory Plasticity in the Optic Tectum of the Barn Owl', *Journal of Neurophysiology* **73**(2), 595–614.

Brooks, R. (1990), AIM 1227: The Behavior Language User's Guide, Technical report, MIT Artificial Intelligence Lab Internal Document.

Brooks, R. (1994*a*), L, Technical report, IS Robotics Internal Document.

Brooks, R. (1994*b*), MARS, Technical report, IS Robotics Internal Document.

Brooks, R. & Stein, L. A. (1994), 'Building Brains for Bodies', *Autonomous Robots* **1:1**, 7–25.

Durbin, R. & Mitchison, G. (1990), 'A Dimension Reduction Framework for Understanding Cortical Maps', *Nature* **343**(6259), 644–647.

Irie, R. (1995), Robust Sound Localization: an Application of an Auditory System for a Humanoid Robot, Master's thesis, MIT.

Kohonen, T. (1982), 'Self-Organized Formation of Topologically Correct Feature Maps', *Biological Cybernetics* **43**, 59–69.

Marjanovic, M., Scassellati, B., & Williamson, M. (1996), Self-Taught Visually-Guided Pointing for a Humanoid Robot, *in* 'Proceedings of the 4th Intl. Conference on Simulation of Adatpive Behavior', Cape Cod, MA.

Obermayer, K., Ritter, H., & Schulten, K. (1990), 'A Principle for the Formation of the Spatial Structure of Cortical Feature Maps', *Proceedings of the National Academy of Science USA* **87**, 8345–8349.

Ritter, H. & Schulten, K. (1988), Kohonen's Self-Organizing Maps: Exploring their Computational Capabilites, *in* 'Proceedings of the IEEE International Conference on Neural Networks', Vol. 1, San Diego, CA, pp. 109–116.

Stein, B. & Meredith, M. (1993), *The Merging of the Senses*, A Bradford Book, Cambridge, MA.