Proceedings of the 1998 IEEE
ISIC/CIRA/ISAS Joint Conference
Gaithersburg, MD • September 14-17, 1998

Spatial Localization and the Refinement of Orienting Behavior: What can be Learned from the Barn Owl?

M. Rucci, J. Wray and G.M. Edelman
The Neurosciences Institute
San Diego, CA

ABSTRACT

The barn owl is a nocturnal predator that relies on audition for hunting. In addition to being able to localize sound sources with high accuracy, the barn owl is capable of adjusting its orienting behavior in the presence of changes in the sensorimotor conditions. This is the case, for example, when the head and ears change drastically in shape and size during growth, altering the auditory inputs. In the last two decades, the biological substrate of spatial localization and orienting behavior in the barn owl have been carefully investigated, and much data are now available regarding the anatomy and physiology of the neural structures involved.

In this paper we review our recent work modeling the principal neural structures responsible for the production of orienting behavior in the brain of the barn owl. In order to expose these models to sensorimotor and environmental conditions similar to those experienced by the barn owl, we coupled the simulation of the neural structures with a robot that emulates the head of a barn owl. This system was composed of a robotic head equipped with two lateral microphones and a camera, and was presented with auditory and visual stimulation. A number of interesting results have emerged from this work. In particular, it has allowed a deeper understanding of how the barn owl reliably localizes a sound source, by elucidating some of the mechanisms underlying the rejection of noise. In addition, it has led to the formulation of a learning scheme accounting for a wide range of biological observations on how the barn owl calibrates orienting behavior. The resulting system was able to orient accurately toward visual and auditory targets, while maintaining accurate performance even in the presence of manipulations of the sensory or motor conditions. This work provides a direct example of how an interdisciplinary approach, based on the coupling of computer simulation of brain structures with robotic systems, can lead to the understanding of basic biological problems while producing robust and flexible control of systems that operate in the real world.

KEYWORDS: *autonomous calibration, spatial localization, adaptive control, neural networks.*

1. INTRODUCTION

Autonomous robotic systems must possess a high degree of flexibility. Similar to biological organisms, they need to adapt to the continuously changing conditions of the environment and of their own sensorimotor characteristics. There is little doubt that the understanding of the basic principles of how brains operate and adapt to the environment could lead to a major revolution in the design of artificial systems. Nevertheless, so far only sporadic interaction has occurred between neuroscience and robotics (e.g. [1, 2, 3]).

The work described in this paper provides one of the earliest examples of how an architecture specifically designed to replicate a biological system can efficiently control a robot operating in the real world. The system that we describe is based on a model of some neural structures in the brain of the barn owl which are dedicated to the production of orienting behavior. It is able to accurately orient toward auditory and visual targets located in different positions, and recover accurate localization behavior after sensorimotor manipulations.

2. MODELING THE BARN OWL NERVOUS SYSTEM AND PHENOTYPE

In the last two decades, neuroscientists have carefully investigated the process of spatial localization and orienting behavior in the barn owl, a nocturnal predator with accurate visual and auditory capabilities [4]. Behavioral experiments have shown that, unlike humans, barn owls rely on two separate auditory cues for the localization of the azimuth and elevation of a sound source: differences in the time of arrival of the sounds at the two ears (interaural time differences (ITDs)) are used for localizing the azimuth, and differences in amplitudes (interaural level differences (ILDs)) are used for the elevation (see [5] for a review). Fig. 1 (*a*) illustrates some of the structures in the brain of the barn owl involved in spatial localization. Two separate neural pathways process

auditory and visual inputs and converge in the optic tectum (OT), a structure well known to be involved in the production of orienting behavior. Within the auditory pathway, physiological investigations have shown the existence of two separate parallel neural pathways specialized for the independent analysis of ITDs and ILDs [6]. The system that we consider is schematically illustrated in Fig. 1 (*b*). A robotic head equipped with a camera and two microphones is controlled by a computer model of the neural pathways of the barn owl involved in the localization of the azimuth. Each structure in the model was composed of a collection of units, each implemented as a leaky integrator:

$$\frac{dy_i}{dt} = \sum_k \omega_{ik} y_k(t) - \delta y_i(t) + n(t) \qquad (1)$$

where $y_i(t)$ is the output of unit i, δ is a constant that determines the decay of activation, $n(t)$ is a noise term, and ω_{ik} is the weight of the connection between units k and i, which represents the strength of a synapse between the two units. The output, $y(t)$, of a unit can be viewed as representing the average firing rate of a collection of cells in the barn owl and its response properties can be considered as representative of a typical cell within such a group. The basic anatomical organization of the model is briefly summarized in Fig. 1. A more detailed description can be found in [7].

3. LEARNING SPATIAL LOCALIZATION

As with many other species, there is little doubt that the barn owl learns to orient toward targets on the basis of sensorimotor experience. During growth, its body changes significantly in size and shape, and its brain must constantly tune sensorimotor coordinations in order to ensure accurate localization.

Adaptation in the model was achieved by allowing synaptic modifications to occur in those structures that have been observed to be plastic in the barn owl. In accord with the results of physiological studies on the barn owl [8, 9], experience-dependent changes in the strength of connections occurred both in the projections from the *ICc* to the *ICx* and in the projections between sensory and motor maps in the *OT*. In the proposed learning scheme, synaptic modifications were mediated by the activation of a modulatory system (not shown in Fig. 1), which signaled the outcome of orienting actions. The modulatory system was modeled as a single unit, which received afferents from both the central region of the retina (the fovea) and the motoneurons. The variable connection weights of plastic synapses were initially set to random values and then changed according to a modified version of Hebbian learning [10]: following each movement, connections between highly correlated units (*i.e.*, those connections which gave a significant contribution to the selection of the motor action), were strengthened or weakened depending on whether a foveation event occurred or not, so as to increase or decrease the likelihood of repeating the same action in similar future situations.

To analyze the capability of the model of producing and maintaining accurate orienting behavior, we have performed many experiments in which the system was trained in in the presence of different sensorimotor conditions [11]. In each case the system had to discover the sensorimotor transformations that led to successful localization of visual and audio-visual targets in variable positions along the azimuthal axis, given the relative structural arrangements of the system components and their functional characteristics. The experimental paradigm consisted of three sequential steps: first, the system was positioned so as to aim in a randomly selected direction. A visual or an audio-visual target was then presented at one of the 15 available locations, and as a result the head moved according to the activity determined within the neural model. After the execution of the movement, the stimuli were removed and the accuracy of localization performance was measured. Periodically, in order to evaluate quantitatively the accuracy of the system, visual and auditory targets at several fixed distances from the current direction of gaze were presented and foveation errors) were measured.

Fig. 2 illustrates the foveation error at different stages of sensorimotor experience in two of such experiments. In the top row of Fig. 2, the initial learning of orienting behavior is shown. Before the exposure to sensorimotor experience, the orientation behavior of the system was poor, due to the initial random strengths of connections in the model. By means of learning, the performance of the system improved with sensorimotor experience, and the foveation error decreased. An example of recovery of accuracy after a manipulation of the motor outputs is shown in the bottom row of Fig. 2. In this case, a nonlinear change of gain in the motors was introduced after the system had already achieved accurate performance. Experiments of this type have shown that the system was able to discover the proper sensorimotor transformations in a wide range of initial conditions, and to quickly recover good foveation performance after sudden changes in both the sensory and motor characteristics.

4. THE "BARN OWL ALGORITHM" FOR AUDITORY LOCALIZATION

A common method used by engineers for estimating the ITD between two signals is by using their cross-correlation function. If $s_L(t)$ and $s_R(t)$ are the auditory signals at the left and right ears, their time-delay Δ can be estimated by searching for the maximum of their cross-correlation function $c(t)$

$$\Delta = \max_T c(t) = \max_T \int_{-\infty}^{\infty} s_L(\tau) s_R(\tau + t) d\tau \quad (2)$$

It has long been hypothesized that ITDs could be derived in the brain of the barn owl by cross-correlating the signals at the two ears. Indeed, several experimental observations suggest that the ITD tuning of ICx cells emerges as a result of a cross-correlation-like treatment of the input signals [12] [13]. However, it has also been observed that space-specific ICx neurons are less affected by noise than would be predicted by direct binaural cross-correlation [13]. The exact mechanisms by which such a noise rejection is achieved are not known.

The so-called generalized cross-correlation method is a more robust algorithm than direct binaural cross-correlation. It is based on a prefiltering of the input signals, so as to compensate for the errors introduced by a finite time window of observation, and by the presence of multiple sound sources or echos. The goal of this method is to ensure large and narrow peaks in the cross-correlation of the two signals, while at the same time ensuring stability with respect to the problem of a finite time window of observation. Instead of the direct cross-correlation $c(t)$ of the two inputs, the cross-correlation $c_g(t)$ of the two filtered signals is used:

$$c_g(t) = \int_{-\infty}^{\infty} H_R(f) H_L^*(f) G_{LR}(f) e^{j2\pi ft} df = \quad (3)$$
$$\int_{-\infty}^{\infty} \Psi_g(f) G_{LR}(f) e^{j2\pi ft} df$$

where $H_L(f)$ and $H_R(f)$ are the transfer functions of the filters used, H^* is the complex conjugate of H, and $G_{LR}(f)$ is the cross-power spectrum of the input signals. Several processing functions $\Psi_g(f)$ have been developed [14]. A strategy common to several of these processors is to operate on the input signals so as to "prewhiten" them by equalizing the amplitude of the cross power spectrum. In this way the peaks in the cross-correlation become sharper and more easily detectable. In the most reliable versions of this method the cross-power spectrum is weighted by taking into account the level of noise at different frequencies. This limits the effect of noise on the final result.

Analysis of our model shows that the processes by which the barn owl determines the azimuth of a sound source bear a strong resemblance to the generalized cross-correlation method [15]. In the case in which the intrinsic connectivity of the model is neglected, so as to consider only the effect of the feedforward connections, the activation of ICx units can be evaluated analytically. It can be shown that the average net input of ICx units is proportional to

$$< U_T^{ICx} > \propto \int_{T-\Delta t}^{T+\Delta t} \int_{-\infty}^{\infty} \Psi_{\mathcal{F}}(f) \hat{G}_{LR}(f) e^{j2\pi ft} df \quad (4)$$

where $A_L(f)$ and $A_R(f)$ are the amplitudes of the Fourier transforms of the input signals, $G_{LR}^g = 2A_L(f)A_R(f)e^{j\phi_{LR}}/W$ is the estimate of the cross power spectrum in the window of observation and

$$\Psi_F(f) = \frac{\mathcal{F}(A_L(f)A_R(f))}{A_L(f)A_R(f)}$$

It follows that if $\mathcal{F}(x)$ is a squashing function as experimental observations seem to indicate, the processor $\Psi_{\mathcal{F}}(f)$ operates in a similar way to the phase transform processor. It introduces a "prewhitening" of the signals by reducing the range of variability of the spectral amplitudes of the input signals. As a result, the peaks of the cross-correlation become sharper. In addition to the action of feed-forward connections, analysis of the activation of units in different frequency bands has shown that lateral connections in the inferior colliculus operate a frequency weighting of the input power spectral densities, that further improve the reliability of ITD estimates in noisy environments in a similar way to the most efficient processors of the generalized cross-correlation algorithm.

5. CONCLUSIONS

In the past, researchers in robotics and AI have often looked at biology as a source of inspiration for solving their problems. From the opposite perspective, neuroscientists have recently turned their attention toward the use of robotic systems as a way to quantitatively test and analyze theories that would otherwise remain at a speculative stage. While computer models have recently gained popularity in the neuroscience community as a way to analyze the operations of complex neuronal ensembles, these models are usually activated with simplified artificial patterns that bear little resemblance to natural stimuli. The use of robotic systems has the advantage of introducing phenotypic and environmental constraints similar to

those that brains of animals have to face during development. Consideration of these constraints is particularly important in light of modern brain theories, that emphasize the importance of the environment and sensorimotor experience during neural development.

It is commonly accepted that the comprehension of how the brain works will dramatically affect the design of artificial systems operating in the real-world. On the other hand, the use of real-world devices seems to be an increasingly promising avenue for understanding the fundamental functional principles of neural systems. We believe that an interdisciplinary approach like the one described in this paper can make significant contributions in both neuroscience and robotics and, at the same time, by establishing a direct link between the natural and engineering sciences, offer new ideas to workers in robotics and artificial intelligence.

ACKNOWLEDGMENTS

This work was carried out as part of the theoretical neurobiology program at The Neurosciences Institute, which is supported by Neurosciences Research Foundation. Support for this program is received in part from the van Ameringen Foundation and Novartis Pharmaceutical Corporation.

References

[1] Albus, J. S., *Brains, Behavior, and Robotics.* Peterborough, N.H: BYTE Books, 1981.

[2] Mead, C., *Analog VLSI and neural systems.* Reading, MA: Addison-Wesley, 1989.

[3] Srinivasan, M. V. and Venkatesh, S., *From living eyes to seeing machines.* Oxford: Oxford University Press, 1997.

[4] Konishi, M., "How the owl tracks its prey." *American Scientist*, vol. 61, pp. 414–424, 1973.

[5] Konishi, M., "Listening with two ears." *Scientific American*, vol. 268, 4, pp. 34–41, 1993.

[6] Konishi, M., Takahashi, T. T., Wagner, H., Sullivan, W. E., and Carr, C. E., "Neurophysiological and Anatomical Substrates of Sound Localization in the Owl." In G. M. Edelman, W. E. Gall, and W. M. Cowan, eds., *Auditory function: neurobiological bases of hearing*, New York: J. Wiley & Sons, pp. 721–744, 1988.

[7] Rucci, M., Tononi, G., and Edelman, G. M., "Registration of neural maps through value-dependent learning: modeling the alignment of auditory and visual maps in the barn owl's optic tectum." *The Journal of Neuroscience*, vol. 17, 1, pp. 334–352, 1997.

[8] du Lac, S. and Knudsen, E. I., "Early visual deprivation results in a degraded motor map in the optic tectum of barn owls." *Proc. Natl. Acad. Sci. USA*, vol. 88, pp. 3426–3430, 1991.

[9] Brainard, M. S. and Knudsen, E. I., "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*, vol. 13, 11, pp. 4589–4608, 1993.

[10] Hebb, D. O., *The organization of behavior.* New York, NY: Wiley, 1949.

[11] Rucci, M., Edelman, G. M., and Wray, J., "Adaptation of orienting behavior: from the barn owl to a robotic system." *IEEE Trans. on Robotics & Automation*, (in press) 1998.

[12] Albeck, Y. and Konishi, M., "Responses of neurons in the auditory pathway of the barn owl to partially correlated binaural signals." *The Journal of Neurophysiology*, vol. 74, pp. 1689–1700, 1995.

[13] Keller, C. H. and Takahashi, T. H., "Binaural cross-correlation predicts the responses of neurons in the owl's auditory space map under conditions simulating summing localization." *The Journal of Neuroscience*, vol. 16, 13, pp. 4300–4309, 1996.

[14] Knapp, C. H. and Carter, G. C., "The generalized correlation method for estimation of time delay." *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 24, pp. 320–327, 1976.

[15] Rucci, M. and Wray, J., "Binaural cross-correlation and auditory localization in the barn owl: a theoretical study." *submitted to Neural Networks*, 1998.

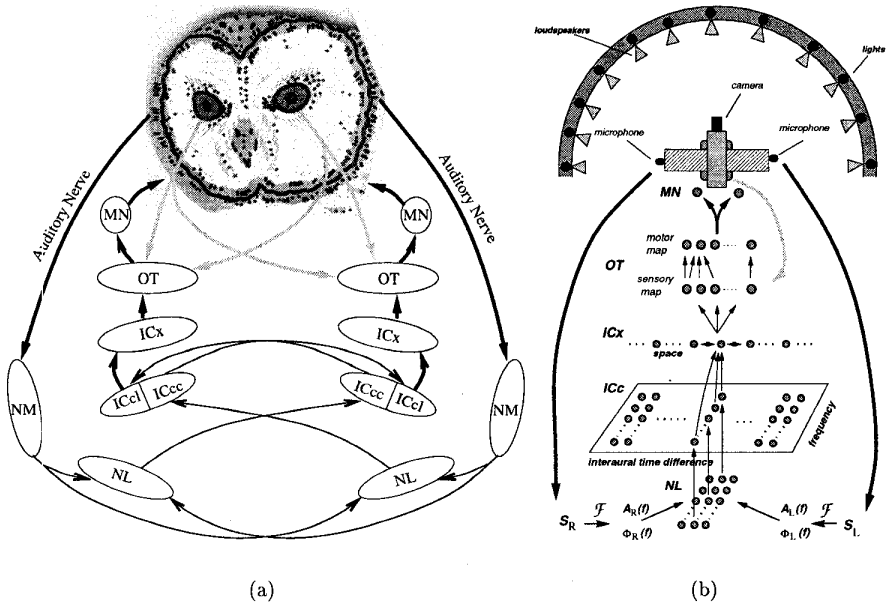

Figure 1: The main neural structures involved in spatial localization in the barn owl and the modeled system. (*a*) In the brain of the barn owl, similarly to the brain of many other species, projections from visual (grey arrows) and auditory (black arrows) structures converge in the optic tectum (OT), which, in turn, projects to motor nuclei (MN) that control the movement of the head. (*b*) The neural model included the main structures dedicated to the localization of the azimuth. Unit activation in the nucleus laminaris (NL) was designed as a function of the spectral components of the two input signals so as to replicate the activation of neurons in the barn owl. NL units were arranged as a bidimensional array in which unit sensitivity varied systematically with respect to frequency along one axis and interaural time delays along the other. The NL projected topographically to the central nucleus of the inferior colliculus (ICc), which was also modeled as a bidimensional map. Following the data from the barn owl, units in the ICc projected to the external nucleus of the inferior colliculus (ICx), which was implemented as a one-dimensional array of units. Each ICx unit received connections from ICc units in all the frequency laminae, but only from a limited range of preferred ITDs. The OT included two unidimensional arrays: a sensory map composed of bimodal units that were activated by both visual and auditory stimuli, and a motor map whose units projected to the motor system. In addition to the connections from the ICx, OT sensory neurons received visual input directly from the retinal map by means of a set of topographically organized connections. A local pattern of short-range excitatory connections and longer-range inhibitory ones was present in the ICc, ICx and OT. The neural model was interfaced to a robotic system composed of a TRC robotic head equipped with a camera and two microphones. The camera was mounted at the center of the pan axis of the head, and the two microphones were located at the opposite sides of the camera. Visual and auditory targets were generated by activating one of 15 lights and loudspeakers located on a circular frame directly in front of the robot. Auditory stimuli consisted of a burst of pseudo-white noise in the frequency range $1-9$ Khz. The two auditory channels were simultaneously sampled and their FFTs estimated continuously on a dedicated multi-channel DSP.

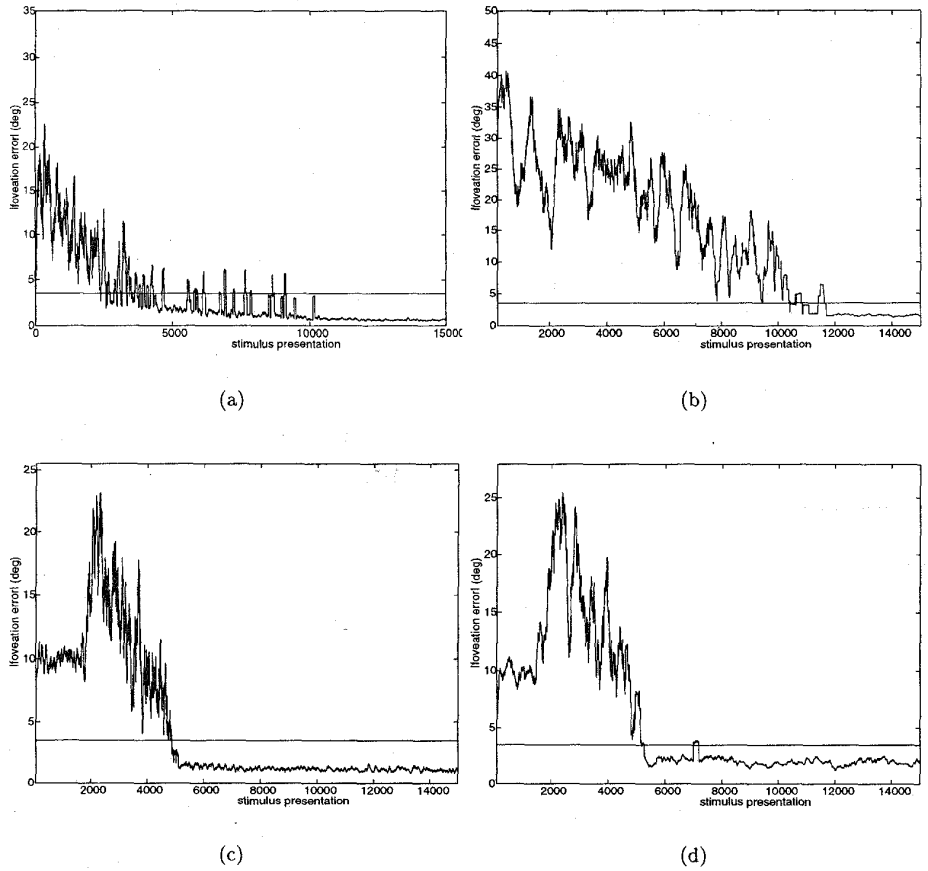

Figure 2: Improvement in the accuracy of orienting behavior toward visual (a, c) and auditory (b, d) targets with learning. The plot is a running average over 50 movements. The horizontal lines in the graphs indicate the dimension of the fovea, which consisted of 3 receptors corresponding to a spatial extension of $\pm 3.5^o$. These results refer to a model composed of 50×300 units in both NL and ICc, and 50 units in all the other maps. (a, b) Initial learning of orienting behavior: the system discovers the correct sensorimotor transformations to localize visual and auditory targets starting from a random connectivity. (c, d) Recovery of accurate performance after manipulations of the motor output. The motor commands sent by the model to the robot were suddenly changed after accurate performance had been already reached.