

Robust localization of auditory and visual targets in a robotic barn owl

M. Rucci*, J. Wray, G.M. Edelman

The Neurosciences Institute, 10640 John Jay Hopkins Drive, San Diego, CA 92121, USA

Abstract

In the last two decades, the barn owl, a nocturnal predator with accurate visual and auditory capabilities, has become a common experimental system for neuroscientists investigating the biological substrate of spatial localization and orienting behavior. As a result, much data are now available regarding the anatomy and physiology of many neural structures involved in such processes. On the basis of this growing body of knowledge, we have recently built a computer model that incorporates detailed replicas of several important neural structures participating in the production of orienting behavior. In order to expose this model to sensorimotor and environmental conditions similar to those experienced by a barn owl, the computer simulations of the neural structures were coupled to a robot emulating the head of a barn owl, which was presented with auditory and visual stimulation. By using this system we have performed a number of studies on the mechanisms underlying the barn owl's calibration of orienting behavior and accurate localization of auditory targets in noisy environments. In this paper we review the main results that have emerged from this line of research. This work provides a concrete example of how, by coupling computer simulations of brain structures with robotic systems, it is possible to gain a better understanding of the basic principles of biological systems while producing robust and flexible control of robots operating in the real world. ©2000 Elsevier Science B.V. All rights reserved.

Keywords: Autonomous calibration; Spatial localization; Adaptive control; Neural networks; Optic tectum

1. Introduction

Accurate spatial localization is crucial in many animal species. Success in capturing prey or escaping from predators very often relies on the ability to precisely determine the position of targets through the available sensory modalities. Spatial localization is also important in artificial systems. By redirecting the cameras toward a target, it is possible to visually analyze stimuli detected by other sensory modalities, and allocate the available computational resources to a selected region of the image, an important step toward

the goal of achieving real-time performance [1,7,8]. In addition, in the presence of visual sensors with anisotropic resolution similar to the eyes of vertebrates [25], orienting behavior allows the examination of the target with the fovea, the region at the center of the visual field with highest spatial acuity.

Since targets may be perceived in any of the available sensory modalities, redirecting the direction of gaze is not a trivial operation. Different sensory inputs are represented in different manners, and spatial cues are not always explicitly available at the input stages, as in the case of vision. In all cases, in order to convert sensory stimuli into corresponding motor outputs, coordinate transformations are required that depend on the structure and characteristics of the considered system. These transformations need to be

* Corresponding author. Tel.: +619-626-2069; fax: +619-626-2099.

E-mail address: rucci@nsi.edu (M. Rucci).

tuned continuously to maintain good performance and compensate for changes in the system components. Although several methods have been developed for evaluating coordinate transformations (see, for example, [10]), most of these procedures often require the use of calibrating tools and the interruption of the normal functioning of the system, operations which are not always feasible. More recently, methods of self-calibration have been developed that are more suited for autonomous systems operating in unstructured environments [5,9,11,17–19]. However, the extension of these techniques to input modalities other than vision is not immediately obvious. For example, it is not clear how they could be applied to cases in which input signals are not passively propagated through the environment, as is the case with touch [21], or when signals are not explicitly spatially organized, as in the auditory domain.

In the last two decades, many neuroscientists investigating the neural mechanisms of spatial localization and the calibration of orienting behavior have focused their studies on the barn owl, a nocturnal predator that relies on accurate visual and auditory localization for hunting. Unlike humans, who combine extremely precise visual localization capabilities with poor auditory localization, the barn owl is able to identify the position of potential prey within a few degrees even in complete darkness, purely on the basis of audition. There is little doubt that the barn owl learns to orient toward targets on the basis of sensorimotor experience. During growth, when the body changes significantly in size and shape over a relatively short period of time, the brain of the barn owl must constantly tune the sensorimotor transformations in order to ensure accurate spatial localization. For example, in the case of audition, cues that determine the position of a sound source depend heavily on the morphology of the head and ears, necessitating the adjustment of behavior to compensate for corresponding changes in the values of the auditory cues. Experiments involving manipulations of the sensory conditions have clearly shown the great adaptability of the barn owl [14]. Young barn owls are able to recover accurate orienting behavior even in the presence of drastic alterations of their sensory inputs.

Recently we have developed a computer model of some of the neural structures dedicated to the localization of the azimuthal position of a target in the brain of

the barn owl [23]. This model has been interfaced to a robotic system emulating the head of the barn owl, so as to provide realistic auditory and visual inputs and produce motor behavior [22]. Two main goals stimulated this work: first, we were interested in achieving a better understanding of the neural principles underlying spatial localization in biological systems. Second, we wanted to provide an alternative to calibration techniques by developing a system that, similar to the barn owl, was able to acquire and maintain accurate localization of targets in different sensory modalities. The aim of this paper is to review some results produced by this line of research that are relevant to the development of autonomous robotic systems. We focus on two topics: the calibration of orienting behavior and the processes underlying a robust auditory localization. In Section 2, we present a brief description of the architecture of both the neural model and the robotic system. A general overview is given in this paper. More detailed descriptions can be found in previous papers that have dealt with specific aspects of this work [22–24]. In Section 3, we focus on the first topic of this review paper: the calibration of orienting behavior. We describe a learning paradigm based on a large amount of biological evidence that accounts for a wide range of experimental observations in the barn owl. Experiments with the model, in which neural changes occur according to the proposed learning paradigm, show that accurate orienting behavior toward visual and auditory targets is achieved and maintained even in the presence of severe manipulations of the sensory and/or motor conditions. In Section 4, we analyze the neural mechanisms underlying the barn owl's robust localization of auditory targets. By examining the activity of different areas of the model in the presence of auditory stimulation we draw a comparison between the selection of the target position in the model and the generalized cross-correlation algorithm for auditory localization. A brief discussion in Section 5 concludes the paper.

2. Modeling the barn owl's neural pathways to the optic tectum

In addition to the combination of sophisticated auditory and visual capabilities, several factors have contributed to the choice of the barn owl as one of the pre-

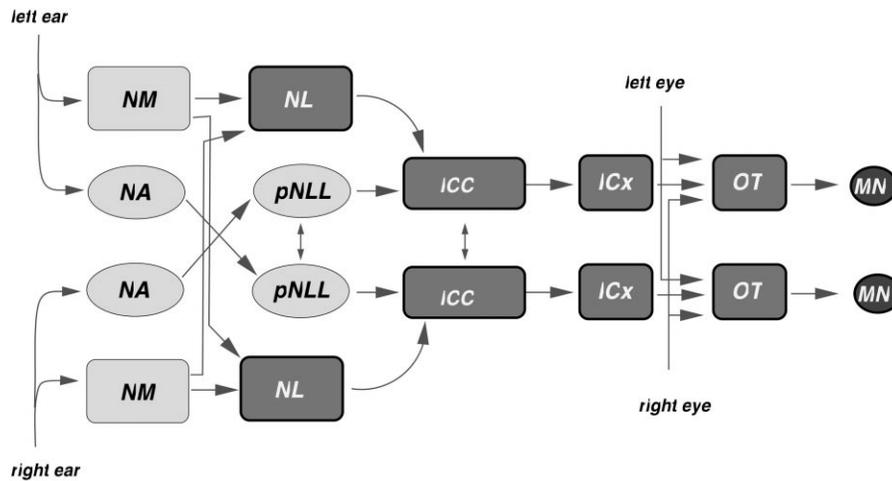


Fig. 1. The main subcortical neural structures involved in spatial localization in the barn owl. Similar to the brain of many other species, projections from visual and auditory structures converge in the optic tectum (OT), which, in turn, projects to motor areas (MN) that control the movement of the head. In the auditory pathway, separate structures process interaural time and amplitude differences (ITDs and ILDs). The nuclei magnocellularis (NM) and laminaris (NL) are dedicated to ITDs, the nucleus angularis (NA) and the pNLL process ILDs. The two pathways converge at the level of the central nucleus of the inferior colliculus (ICC), which in turn projects to the external nucleus (ICx) where a neural map of space in the auditory modality is found. This map is projected to the optic tectum where it is registered with a visual map of space deriving from topographic projections from the retina. Square (elliptical) boxes represent structures dedicated to the analysis of time (amplitude) binaural differences. Only some of the areas participating in the localization of the azimuthal position of a target have been included in the model (filled boxes). See text for details.

ferred experimental species by neuroscientists studying spatial localization. Among these factors is the inability of the barn owl to move its eyes with respect to its head which simplifies the problem of coordinate transformation by eliminating some degrees of freedom with respect to other species. Another important factor is the prototypical reflex response that consists of a rotation of the head toward a target, a behavior that can be easily and reliably measured. Most studies on the barn owl have focused on the neural pathways converging on, and departing from, the optic tectum, a subcortical multimodal center that is well known to participate in orienting behavior in many other species, including primates. These investigations have provided detailed anatomical and physiological characterization of many neural structures present in these pathways. It is now clear that 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 estimating

the elevation. Physiological investigations have shown the existence of two separate parallel neural pathways specialized for the independent analysis of ITDs and ILDs (see [16] for a review).

A schematic overview of some of the barn owl's sensory pathways to the optic tectum is illustrated in Fig. 1. In the figure, the neural structures that we have considered in our computer model are shown by dark boxes. In the barn owl, each area is duplicated on each side of the brain, where it processes the contralateral part of space. In the model the two parts were collapsed together so that each modeled area was dedicated to the analysis of signals from the whole surrounding space. The basic element of the model was a leaky integrator unit described by

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

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

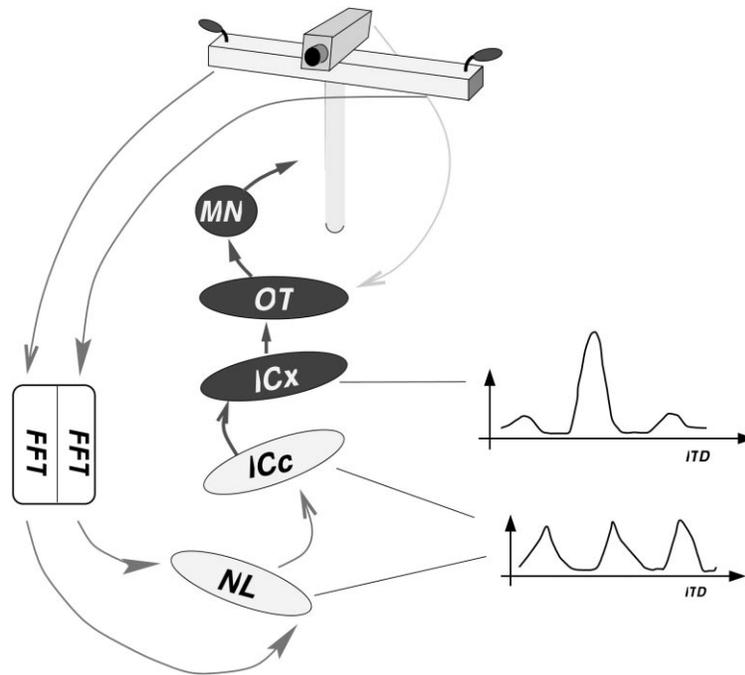


Fig. 2. Organization of the neural model. Lightly shaded boxes indicate areas in which units are arranged in two-dimensional maps, with unit sensitivity for frequency and ITD changing along the two axes. Dark boxes represent areas with units organized in one-dimensional arrays. On the right-hand side of the figure the representative responses of specific units in different areas to different values of ITD are shown. The periodicity in the response at the level of the NL and the ICc is lost in the ICx, where each unit responds strongly to a single value of ITD, i.e. to a single spatial location.

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. One main concern in building this model was to incorporate as much physiological and anatomical data as possible, in order to accurately replicate the characteristics of cells in different areas in the brain of the barn owl. Each structure in the model was composed of a collection of units. With the exception of the first two structures in the auditory pathway, the nucleus laminaris (NL) and the central nucleus of the inferior colliculus (ICc), in which units were arranged in two-dimensional arrays, all modeled areas consisted of a unidimensional array of units. All areas, apart from the input stages in the model (NL and the retina), included a pattern of intrinsic connectivity based on a local range of excitatory connections and a longer range of inhibitory ones.

That is, nearby units tended to activate each other, whereas units farther apart were mutually inhibited.

A schematic overview of the model is given in Fig. 2. The NL was the first stage included in the auditory pathway based on the fact that it was possible to accurately predict the response of NL units on the basis of the auditory inputs. In the model, the activation of NL units was empirically designed so as to fit as closely as possible the physiological data available in the literature. For binaural stimulation, when the two signals $s_L(t)$ and $s_R(t)$ with Fourier transforms $A_L(f)e^{j\phi_L(f)}$ and $A_R(f)e^{j\phi_R(f)}$ (evaluated over a time-window of observation W) were applied as inputs, the activation of NL units was dependent on the unit characteristic parameters \bar{T} , \bar{f} as follows:

$$U_{\bar{T}, \bar{f}}^{\text{NL}} = \mathcal{F}(A_L(\bar{f})A_R(\bar{f})) \cdot [\cos(\phi_{LR}(\bar{f}) - 2\pi \bar{f}\bar{T}) + 1]G_{\bar{f}\sigma}(f), \quad (2)$$

where \mathcal{F} is a logarithmic function that controlled the change in firing rate for different amplitudes of the input signals, $\phi_{LR}(\bar{f}) = \phi_L(\bar{f}) - \phi_R(\bar{f})$ is the difference of phase between the left and right input signals at frequency \bar{f} , and $G_{\bar{f}\sigma}(f)$ is a Gaussian function with mean \bar{f} and variance σ . Following anatomical studies on the barn owl, units in the NL were arranged in a two-dimensional array in which the unit preferred ITD \bar{T} and frequency \bar{f} varied systematically along the two axes. The ICc was also organized as a two-dimensional array of ITD versus frequency. Units in the ICc, received topographical connections from the NL (that is, each unit in the ICc received input from a unit in the same position in the NL). However, the pattern of activation in the ICc was different from the NL as a result of the intrinsic connectivity among ICc units.

As in the barn owl, the frequency dependency was lost at the level of the ICx. Each unit in the ICx received connections from ICc units in all frequency laminae, but only from a specific range of ITDs. Following physiological observations from experiments involving manipulations of the sensory inputs [6], the strengths of these connections were modified during sensorimotor experience. With training, each unit acquired a sensitivity for a specific value of ITD, so that it was activated only by sounds emitted from a single spatial position. The unit preferred ITD varied systematically in the array so as to create an auditory neural map of the surrounding space.

The optic tectum (OT) was modeled with two arrays of units: a sensory map, receiving topographical connections both from the retina and from the ICx, and a motor map receiving plastic connections from the sensory map. These two arrays replicated the superficial and deep layers of the optic tectum. As in the barn owl, units in different positions were sensitive to stimuli in specific positions of space independent of the sensory modality in which they appeared. In the presence of visual and/or auditory stimulation a corresponding pattern of activity was generated in the OT sensory map due to the specific pattern of connections from the retina and the ICx. This pattern of activity was transmitted to the motor map where motor units converted it into a specific direction of gaze. In this way, whenever an auditory or visual target was detected a reflexive orienting behavior was generated. The connections between the sensory and motor maps

in the OT were also modified on the basis of sensorimotor experience. Specific parameters describing the network used can be found in [22].

As illustrated in Fig. 3, the model was coupled to a robotic system composed of a camera and two microphones. The system was controlled via a VME bus with dedicated video and audio boards, interfaced to the workstation where the neural model was implemented. Similarly to the barn owl, the robot could aim in different directions, while maintaining the position of the microphones and the camera fixed with respect to each other. An array of 15 lights and loud-speakers provided auditory and visual stimulation in different spatial positions. Real-time preprocessing of the data included thresholding and compression along the azimuthal axis (an average of the pixel activation along each column was taken) for the visual signal, and simultaneous FFTs for the auditory signals. Preprocessing was carried out on dedicated boards on the VME bus.

3. Learning and adaptation of orienting behavior

3.1. Proposed learning paradigm

Young barn owls are able to recover accurate spatial localization even in the presence of drastic manipulations of their sensory inputs. A striking example is the recovery of precise auditory localization in barn owls exposed to a chronic translation of the visual field through the insertion of prismatic goggles on their eyes [15]. In this case, the barn owl has to learn that whenever a sound source is present, in order to look at the source with the fovea, it has to aim to the side of it by an amount equal to the translation introduced by the goggles. It has been shown that the adaptation of orienting behavior has a neural correlate in the change of preferred ITD of neurons at the level of the ICx. As a result of changes in their connections from the ICc, these neurons change their sensitivity so that their corresponding preferred spatial position translates by an amount equal to the visual shift [6]. Different experiments have also revealed that the alignment between visual and motor maps in the OT may also be modified, although the site of plasticity has not been unequivocally determined.

Although experimental evidence has elucidated some of the neural changes that occur during the adap-

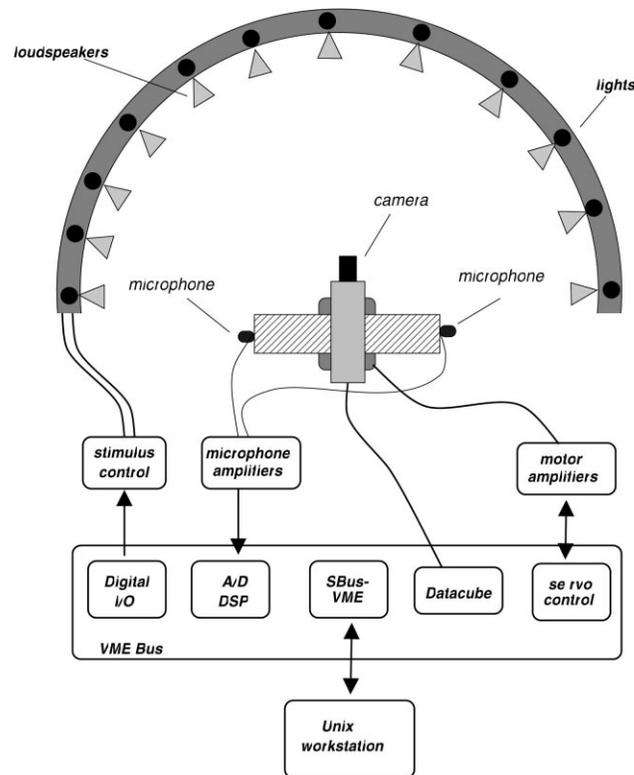


Fig. 3. Architecture of the robotic system controlled by the model of the barn owl neural pathways. A camera and two microphones were mounted on the pan axis of a TRC robotic head. Visual signals were acquired and processed by a Datacube MV200 board. Auditory signals were digitized by synchronized A/D channels and FFTs were evaluated in real-time on a DSP board. The system was positioned at the center of a circular array which included 15 lights and loudspeakers. Selected targets were activated via a digital I/O board on the VME bus (reprinted with permission from [22] ©1998 IEEE).

tation of orienting behavior, the mechanisms regulating such changes are not known. By using our computer model, we have proposed that the development and maintenance of accurate spatial localization can be explained in the context of a neurobiologically-based paradigm of reinforcement learning, which we call *value-dependent learning* [23]. Based on a large body of biological evidence, this paradigm proposes that signals related to the saliency of sensorimotor events are broadcast by the activation of diffuse-projecting neuromodulatory systems to different parts of the brain, where they modulate synaptic changes mediating neural plasticity. In the model, synaptic modifications were mediated by the activation of a modulatory system, triggered by the occurrence of salient events, that functionally replicated the proposed role of neuromodulatory subcortical centers. For simplic-

ity, the modulatory system consisted of a single unit, which received afferents from both the fovea and the motoneurons. In this way, the modulatory unit was moderately active when a stimulus was in the center of the visual field or when a movement was executed, and it was particularly active when the two events occurred simultaneously or shortly one after the other. Following experimental observations, the strengths of connections were allowed to change on the basis of experience both in the projections from ICc to ICx and in the projections between sensory and motor maps in the OT. The variable connection weights of plastic synapses were initially set to random values and then changed according to a modified version of Hebbian learning. 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. At each time step of the simulations, the strength of a connection ω_{ik} between unit i and unit k was modified according to

$$\frac{d\omega_{ik}}{dt} = \Phi_L(\epsilon_1 U_i U_k + \epsilon_2 V), \quad (3)$$

where U_i is the activation of unit i , and V is the activation of the modulatory system that acts as a reward signal. $\Phi_L(x)$ is a piecewise linear function characterized by two thresholds θ_{LTD} and θ_{LTP} . When the argument is lower than the first threshold θ_{LTD} the function has a value of zero; in between the two thresholds it has a negative value, thus inducing depression of synaptic strengths; and above θ_{LTP} it assumes positive values, thereby potentiating the connections. Typically, the values of the thresholds were selected so that potentiation occurred only for units whose activations were correlated when the reward signal was high. This happened after a successful saccade, which brought the stimulus onto the fovea. In this way, the probability of repeating the correct action in the future was increased. By contrast, when the reward signal was low, such as when the system moved but did not localize the target, highly correlated units were mostly depressed. This decreased the probability of repeating the wrong movement.

3.2. Experiments

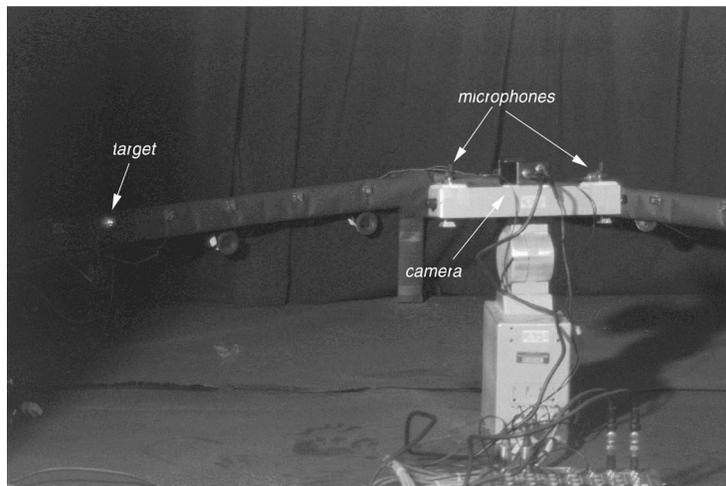
To analyze the capability of the model of producing and maintaining accurate orienting behavior, many experiments were performed in which the system was trained in the presence of different sensorimotor conditions [22]. The goal of the experiments was to discover the sensorimotor transformations leading 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 (see Fig. 4). After the execution of the movement, the stimuli were removed and the accuracy of localization performance was measured. These experiments 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.

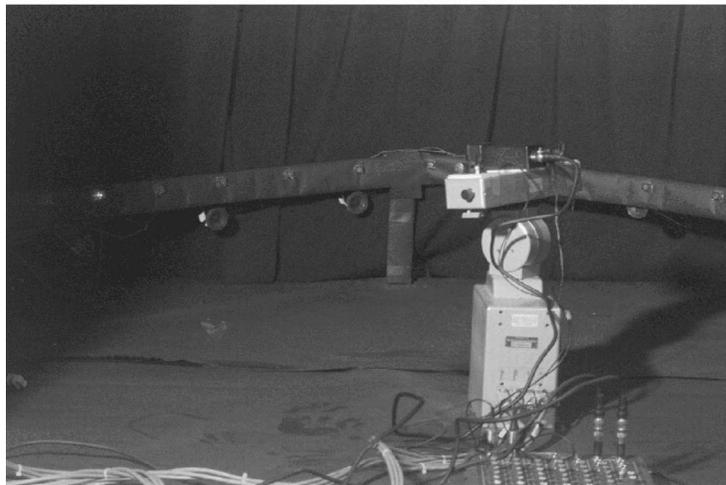
The results of three experiments of this type are documented in Table 1. The data refer to the accuracy of localization of auditory targets after the presentation of 15,000 targets, the equivalent of few hours of training (similar results were also obtained in the localization of visual targets, see [22]). In this case the dimension of the fovea was equal to a spatial extent of $\pm 3.5^\circ$. The first column of Table 1 shows the performance of the system after direct training, that is starting from the initial random patterns of connectivity. The data in the second column are relative to systems that were first trained with normal sensorimotor experience, and then exposed to an alteration of their sensorimotor characteristics. In each case, 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. The case of “normal” development in which no manipulation of the sensory or motor conditions was present is shown on the top row of Table 1. In this case, the sensorimotor transformations that need to be discovered are determined uniquely by the functional characteristics and relative arrangements of the sensors and the robot. Before training, the performance in the localization of auditory targets was $39.3^\circ \pm 32.7^\circ$. A more complex situation is the one shown in of the middle row of Table 1. This experiment replicated the insertion of prismatic goggles on the eyes of the barn owl by systematically

Table 1
Performance of auditory localization

Experiment	Direct training	Recovery
Normal development	$1.5^\circ \pm 1.0^\circ$	–
Visual alteration	$2.3^\circ \pm 1.1^\circ$	$2.2^\circ \pm 1.0^\circ$
Motor alteration	$2.2^\circ \pm 1.4^\circ$	$1.8^\circ \pm 1.3^\circ$



(a)



(b)

Fig. 4. An example of correct orienting behavior. An audiovisual target present at the left-hand side of the robot (top image) triggers a motor action that brings the target at the center of the visual field (bottom image).

translating the visual field by 20° to the right. As in the barn owl, the system learned that in order to foveate on an auditory target it was necessary to aim at the side by an amount equal to the visual displacement. Even in the case in which the system was first trained in the normal conditions and then suddenly exposed to a translation of the visual field accurate spatial localization was produced. A case of chronic manipulation of the motor outputs is illustrated in the bottom row of Table 1. In this case, the position of the system

was altered by simulating a nonlinear change of gain in the motors. The azimuthal position Φ determined from the motor units was systematically changed to $k\Phi^2$, where k is a constant that takes into account the dynamic range so as to maintain a fixed range of possible movements. In this case the system was also able to recover accurate orienting behavior even if it was previously trained with normal sensorimotor experience, and then suddenly exposed to the motor manipulation.

4. Some mechanisms underlying a robust auditory localization

As shown by the data in Table 1, after having been exposed to sufficient sensorimotor experience, the system was able to localize auditory targets with high accuracy. The origin of such precision in the determination of the position of a sound source, or equivalently the development of narrow tuning by ICx units to specific values of ITD, was the subject of a second study [24]. 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 [2,12]. This is a commonly used algorithm for estimating the binaural delay Δ between the two input auditory signals, $s_L(t)$ and $s_R(t)$, that is based on searching for the maximum of their cross-correlation function $c(t) = s_L(t) * s_R(t)$. However, it has also been observed that in the barn owl the ITD specificity of ICx neurons is less affected by noise than it would be predicted by binaural cross-correlation [12]. The exact mechanisms by which such robustness is achieved are not known.

4.1. Analysis of the feed-forward model

By using our model, we first examined the emergence of ITD sensitivity in ICx units in the case in which the intrinsic connectivity of the model was removed, i.e. when only the effect of feed-forward connections was considered. In this case the model was simple enough to allow an analytical examination. We found that the average activation of a ICx unit with preferred ITD T is given by

$$\begin{aligned} \langle U_T^{\text{ICx}} \rangle & \propto \int_{T-\Delta t}^{T+\Delta t} \int_{-\infty}^{\infty} \frac{\mathcal{F}(A_L(f)A_R(f))}{A_L(f)A_R(f)} \hat{G}_{\text{LR}}(f) e^{j2\pi f t} df \\ & = \int_{T-\Delta t}^{T+\Delta t} \int_{-\infty}^{\infty} \Psi_{\mathcal{F}}(f) \hat{G}_{\text{LR}}(f) e^{j2\pi f t} df, \end{aligned} \quad (4)$$

where $A_L(f)$ and $A_R(f)$ are the amplitudes of the Fourier transforms of the input signals, \hat{G}_{LR} is the estimate of the cross power spectrum in the window of

observation, and the function $\mathcal{F}(x)$ is the term present in the activation of NL units (see Eq. (2)). Details on how Eq. (4) was derived can be found in [22]. The time integral results from the spread of projections from the ICc. In this case, the considered ICx unit received strong connections from all ICc units with characteristic ITD in the range $(T - \Delta t, T + \Delta t)$. In a well trained systems Δt is small, and the time integral can be neglected in Eq. (4) by replacing t with T in the frequency integral. It is obvious from Eq. (4) that if $\mathcal{F}(x)$ is a linear function, the activation of an ICx unit would be indeed an approximation of the cross-correlation of the input signals, evaluated at the unit preferred ITD. However, since experimental observations indicate that $\mathcal{F}(x)$ is a sub-linear squashing function, in this simplified case of no lateral connectivity, the activation of ICx units deviate from a simple cross-correlation by the introduction of the term $\Psi_{\mathcal{F}}$. In order to understand how this term alters the sensitivity of ICx units with respect to the cross-correlation, it should be observed that Eq. (4) bears a strong resemblance to the so-called generalized cross-correlation method [13]. The generalized cross-correlation algorithm is a more robust method than direct binaural cross-correlation that is based on prefiltering the input signals so as to take into account the fact that a finite time window of observation is used in reality, and that multiple sound sources or echos may be present. The goal of this method is to make the detection of the binaural delay more reliable, by ensuring large and narrow peaks in the cross-correlation of the two signals. Instead of the direct cross-correlation $c(t)$ of the two inputs, the generalized cross-correlation $c_g(t)$ of the two filtered signals is used:

$$\begin{aligned} c_g(t) & = \int_{-\infty}^{\infty} H_R(f) H_L^*(f) G_{\text{LR}}(f) e^{j2\pi f t} df \\ & = \int_{-\infty}^{\infty} \Psi_g(f) G_{\text{LR}}(f) e^{j2\pi f t} df, \end{aligned} \quad (5)$$

where $H_L(f)$ and $H_R(f)$ are the transfer functions of the filters, and H^* is the complex conjugate of H . Several processing functions $\Psi_g(f)$ have been developed [13]. A strategy common to several of these processors is to perform a “prewhitening” of the input signals by equalizing the amplitude of their cross power spectrum G_{LR} . This operation sharpens the peaks of

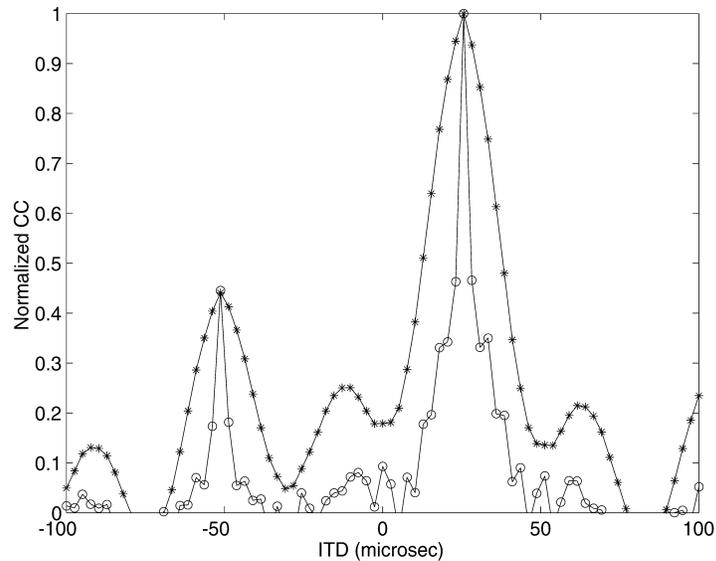


Fig. 5. Effect of a logarithmic compression on the spectral amplitude of the cross-correlation. The auditory inputs presented a main binaural delay at $25 \mu\text{s}$ and an echo at $-51 \mu\text{s}$. The stars indicate direct binaural cross-correlation. The circles refer to the cross-correlation after a logarithmic filtering as implemented by NL units.

the cross-correlation and makes the detection of ITD more reliable.

In the model, the processor $\Psi_{\mathcal{F}}$ produces an enhancement of the peaks of the cross-correlation similar to the generalized cross-correlation method. (see Fig. 5). This is a result of the particular shape of $\mathcal{F}(x)$ that has the effect of “prewhitening” the signals by reducing the range of variability of their spectral amplitudes. In this way, the amplitude dependency of NL units improves the signal to noise ratio for the detection of auditory sources by increasing the activation of ICx units with preferred ITDs corresponding to the input binaural delays, and decreasing the responses of the other units.

4.2. Effect of lateral connectivity

What happens in the more general case in which intrinsic connections at the level of the ICc and ICx are taken into account? In this case, the recursive interactions among units make computer simulations a more convenient tool to understand how the response of ICx units change with respect to the case of only feed-forward connectivity. The first outcome of these

simulations was the observation that the main effect of intrinsic connections at the level of the ICx was to suppress the mean activation of ICx units and enhance the peaks of activity. However, a more striking contribution of lateral connectivity was revealed in the presence of noisy signals. In this case, which is the norm in a real environment, the reciprocal inhibition of areas tuned to different interaural delays in the ICc produced ensembles of particularly active units only in adjacent frequency laminae possessing a similar pattern of activation. Since the input to all ICc neurons tuned for the same frequency depends on the estimate of the mean interaural phase ϕ_{LR} (see Eq. (2)), groups of active cells emerged only in adjacent frequency laminae with a similar estimate of the interaural phase. In the frequency bands in which the signal to noise ratio was lower, the variability of the estimate of the interaural phase was larger, and the activity of neurons in adjacent frequency laminae were more dissimilar than in frequency bands with higher signal to noise ratio. The result was a frequency weighting in which the contribution of different frequency bands to the activation of ICx units was inversely proportional to the variability of the phase estimate. This operation is extremely

similar to the maximum likelihood method, the most reliable version of generalized cross-correlation, in which the processing function takes into account the estimated level of noise in different frequency bands, so as to limit the effect of noise on the final result. It can be shown that in the maximum likelihood method the amplitude of the cross-power spectrum at a particular frequency is normalized by a factor proportional to the variability of the estimate of the phase at that frequency:

$$c_{\text{ML}}(t) \propto k \int_{-\infty}^{\infty} \frac{e^{j\hat{\theta}}(f)}{\sigma_{\hat{\theta}}^2(f)} e^{j2\pi f t} df \quad (6)$$

which is similar to the operation of intrinsic connections at the level of the ICc.

By comparing the processes of auditory localization in the barn owl to the generalized cross-correlation, one of the most robust algorithms for the detection of interaural delays, an elucidation of some of the mechanisms underlying precise localization in noisy environments has been possible. The model, by including in a parallel fashion signal prewhitening and weighting of the power spectrum based on the signal-to-noise ratio at different frequencies, was able to evaluate the position of sound source quickly and reliably even in the presence of echoes and noise.

5. Conclusions

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 will lead to a major revolution in the design of artificial systems. Nevertheless, neuroscience and robotics stand as two independent disciplines with only sporadic interactions (see, for example, [3,20,26,27]). A recent review of the emerging field of biorobotics is given in [4].

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.

More than 50 years ago, in the introduction of his book *Cybernetics: Or Control and Communication in the Animal and the Machine*, Norbert Wiener wrote:

“... [I had] the conviction that the most fruitful areas for the growth of the sciences were those which had been neglected as a no-man’s land between the various established fields. ... It is these boundary regions of science which offer the richest opportunity to the qualified investigator. They are at the same time the most refractory to the accepted techniques of mass attack and the division of labour. If the difficulty of a physiological problem is mathematical in essence, ten physiologists ignorant of mathematics will get precisely as far as one physiologist ignorant of mathematics, and no further” (Wiener, 1947).

Although many years have passed in which science and technology have exhibited dramatic advancements, these words are still valid. Never more than today has science become the tasks of specialists in fields that are growing progressively narrower, and similar problems are studied independently in different disciplines. 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. We believe that, by establishing a direct link between the natural and engineering sciences, interdisciplinary efforts like the one described in this paper can make significant contributions in both neuroscience and robotics.

Acknowledgements

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] A.L. Abbott, A survey for selective fixation control for machine vision, *IEEE Journal of Control Systems* 12 (4) (1992) 25–31.
- [2] Y. Albeck, M. Konishi, Responses of neurons in the auditory pathway of the barn owl to partially correlated binaural signals, *Journal of Neurophysiology* 74 (1995) 1689–1700.
- [3] J.S. Albus, *Brains, Behavior, and Robotics*, BYTE Books, Peterborough, NH, 1981.
- [4] R.D. Beer, H.J. Chiel, R.D. Quinn, R.E. Ritzmann, Biorobotic approaches to the study of the motor system, *Current Opinion in Neurobiology* 8 (1998) 777–782.
- [5] D.J. Bennet, D.Geiger, J.M. Hollerbach, Autonomous robot calibration for hand–eye coordination, *International Journal of Robotics Research* 10 (5) (1991) 550–559.
- [6] M.S. Brainard, E.I. Knudsen, Experience-dependent plasticity in the inferior colliculus: a site for visual calibration of the neural representation of auditory space in the barn owl, *Journal of Neuroscience* 13 (11) (1993) 4589–4608.
- [7] P.J. Burt, Smart sensing within a pyramid vision machine, *Proceedings of the IEEE* 76 (8) (1988) 1006–1015.
- [8] C. Colombo, M. Rucci, P. Dario, Attentive behavior in an anthropomorphic robot vision system, *Journal of Robotics and Autonomous Systems* 12 (1994) 121–131.
- [9] F. Du, M. Brady, Self-calibration of the intrinsic parameters of camera for active vision systems, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1993.
- [10] K.S. Fu, R.C. Gonzalez, C.S.G. Lee, *Control, Sensing, Vision and Intelligence*, McGraw-Hill, New York, 1987.
- [11] R.I. Hartley, Self-calibration from multiple views with a rotating camera, in: *Proceedings of the Third European Conference on Computer Vision*, 1994, pp. 471–478.
- [12] C.H. Keller, T.H. Takahashi, Binaural cross-correlation predicts the responses of neurons in the owl's auditory space map under conditions simulating summing localization, *Journal of Neuroscience* 16 (13) (1996) 4300–4309.
- [13] C.H. Knapp, G.C. Carter, The generalized correlation method for estimation of time delay, *IEEE Transactions on Acoustics, Speech and Signal Processing* 24 (1976) 320–327.
- [14] E.I. Knudsen, Experience shapes sound localization and auditory unit properties during development in the barn owl, in: G.M. Edelman, W.E. Gall, W.M. Cowan (Eds.), *Auditory Function: Neurobiological Bases of Hearing*, Wiley, New York, 1988, pp. 137–148.
- [15] E.I. Knudsen, M.S. Brainard, Visual instruction of the neural map of auditory space in the developing optic tectum, *Science* 253 (1991) 85–87.
- [16] M. Konishi, Listening with two ears, *Scientific American* 268 (4) (1993) 34–41.
- [17] S.D. Ma, A self-calibration technique for active vision systems, *IEEE Transactions on Robotics and Automation* 12 (1996) 114–120.
- [18] S.J. Maybank, O. Faugereas, A theory of self-calibration of a moving camera, *International Journal of Computer Vision* 8 (1992) 123–151.
- [19] P.F. McLauchlan, D.W. Murray, Active camera calibration for a head-eye platform using the variable state-dimension filter, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18 (1) (1996) 15–22.
- [20] C. Mead, *Analog VLSI and Neural Systems*, Addison-Wesley, Reading, MA, 1989.
- [21] M. Rucci, R. Bajcsy, Learning visuo-tactile coordination in robotic systems, in: *Proceedings of the IEEE Conference on Robotics and Automation*, 1995, pp. 2678–2683.
- [22] M. Rucci, G.M. Edelman, J. Wray, Adaptation of orienting behavior: from the barn owl to a robotic system, *IEEE Transactions on Robotics and Automation* 15 (1) (1999) 96–110.
- [23] M. Rucci, G. Tononi, G.M. Edelman, Registration of neural maps through value-dependent learning: Modeling the alignment of auditory and visual maps in the barn owl's optic tectum, *Journal of Neuroscience* 17 (1) (1997) 334–352.
- [24] M. Rucci, J. Wray, Binaural cross-correlation and auditory localization in the barn owl: A theoretical study, *Neural Networks* 12 (1) (1999) 31–42.
- [25] G. Sandini, G. Gandolfo, F. Grosso, M. Tistarelli, Vision during action, in: Y. Aloimonos (Ed.), *Active Perception*, Lawrence Erlbaum, Hillsdale, NJ, 1993, pp. 151–190.
- [26] M.V. Srinivasan, S. Venkatesh, *From Living Eyes to Seeing Machines*, Oxford University Press, Oxford, 1997.
- [27] B. Webb, A cricket robot, *Scientific American* 275 (6) (1996) 94–99.



Michele Rucci received a *laurea* degree *summa cum laude* in Bioengineering from the University of Florence and a Ph.D. from the Scuola Superiore S. Anna in Pisa. In 1994, before joining The Neurosciences Institute, he was a Visiting Researcher with the University of Pennsylvania, Philadelphia. His research interests are in the field of perception and learning both in biological and artificial systems. He has undertaken an interdisciplinary approach which involves the use of mathematical and computational techniques to investigate the basic principles of neural systems and the application of such principles to the development of new algorithms in

machine perception. He is particularly interested in the analysis of complete sensorimotor loops and in the use of robotic systems to expose neural models to the nonidealized characteristics of real environments. He is the author of publications in both biological and engineering journals.

Dr. Rucci is a Fellow in Computational Neuroscience at The Neurosciences Institute, San Diego, CA.



Jonny Wray was born in Yorkshire, England in 1966. He received the B. Eng. in Electrical and Electronic Engineering and Ph.D. in Physiology from the University of Newcastle upon Tyne, UK. His thesis work involved the mathematical analysis of neural networks, the formal comparison of such systems with classical techniques and the use of networks in nonlinear analysis of single cell biophysics. He is currently an Institute Fellow in Theoretical Neurobiology at The Neurosciences Institute, Sand Diego, CA, where he has developed his interests in systems neuroscience. His current research interests include the neural basis of perception and sensorimotor co-ordination which he studies using large-scale computer simulations of nervous systems.

Dr. Wray is a member of the IEE, the Society of Neuroscience and the Association for Research in Vision and Ophthalmology.



Gerald M. Edelman, M.D., Ph.D., Nobel Laureate in Physiology or Medicine in 1972, is President of Neurosciences Research Foundation and Director of The Neurosciences Institute, as well as the Chairman of the Department of Neurobiology at The Scripps Research Institute. He earned a B.S. degree from Ursinus College, and M.D. from the University of Pennsylvania, and a Ph.D. from the Rockefeller Institute. He also has numerous earned and honorary D.Sc. degrees from institutions as diverse as Williams College, the University of Paris, and the University of Cagliari in Sardinia. Dr. Edelman's early studies of the structure and diversity of antibodies led to his Nobel award; he then began research into the mechanisms involved in the regulation of primary cellular processes, particularly the control of cell growth and the development of multicellular organisms. In work focusing on cell–cell interactions, Dr. Edelman identified cell adhesion molecules (CAMs), which help guide the fundamental processes by which an animal achieves its shape and form and by which nervous systems are built. At the juncture of these two discoveries lies the basis of Dr. Edelman's theory of the development and organization of higher brain functions in terms of a process known as neuronal group selection, described in his 1987 book, *Neural Darwinism*.

Dr. Edelman is a member of the National Academy of Sciences, the American Philosophical Society, the Institute de France. He is the author of several books and over 400 research publications.