ELSEVIER

Robotics and Autonomous Systems 39 (2002) 87–106

Robotics and Autonomous Systems

www.elsevier.com/locate/robot

Development of auditory-evoked reflexes: Visuo-acoustic cues integration in a binocular head

Lorenzo Natale*, Giorgio Metta, Giulio Sandini

Laboratory for Integrated Advanced Robotics (LIRA), Department of Communication, Computer and Systems Science, University of Genoa, Via Opera Pia 13, 16145 Genoa, Italy

Received 7 May 2001; received in revised form 22 January 2002
Communicated by F.C.A. Groen

Abstract

The goal of this paper is to propose a biologically plausible, functional model of the acquisition of visual, acoustic and multi-modal motor responses. Within this context visual and acoustic spatial cues are considered, fused in a coherent percept and eventually employed to control the orienting behavior of a humanoid robot. The rationale of the approach lies in the possibility to test and empirically prove the correctness of the model through the embodiment and the real interaction of the system with the environment.

The model takes into account the fact that (i) acoustic and visual cues are represented with respect to different coordinate frames (head-centric versus retino-centric) and consequently they need to be "aligned" in order to be properly fused, (ii) a teaching signal has to be generated in order to inform the system that the motor performance is not adequate to perform the task (i.e. orient toward the stimulus) and thus adaptation is required, and (iii) vision plays a major role in driving the acquisition of the appropriate map of space but other sources of feedback might be employed as well. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Sound localization; Visuo-acoustic integration; Biomimetic robotics

1. Introduction

Many species including humans rely on acoustic cues to behave appropriately in their environment. If only sound interpretation were the concern, perhaps a single ear would be sufficient. But for a very peculiar task two ears are indeed necessary: that is sound localization.

Although in some cases—such as vision and touch—the brain might recover spatial information from a simple mapping of the sensory epithelium (stereoscopic vision is an exception), sound localization is based on a precise computational process. One fascinating characteristic of this process is the ability to detect time differences as small as few microseconds in spite of the duration of neural spikes, which is in the order of milliseconds.

The binaural information relevant to sound localization is usually considered to be the interaural phase difference (IPD) and the interaural level difference (ILD), due respectively to the different position of the ears and to the shape of the head [1]. The IPD is mainly related to the horizontal position of a sound source (azimuth) while the ILD can be useful to perform estimation of both elevation and azimuth (see Fig. 1). In

* Corresponding author. Tel.: +39-10-3532946; fax: +39-10-3532948.
E-mail addresses: nat@dist.unige.it (L. Natale), pasa@dist.unige.it (G. Metta), giulio@dist.unige.it (G. Sandini).

0921-8890/02/$ – see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0921-8890(02)00174-4

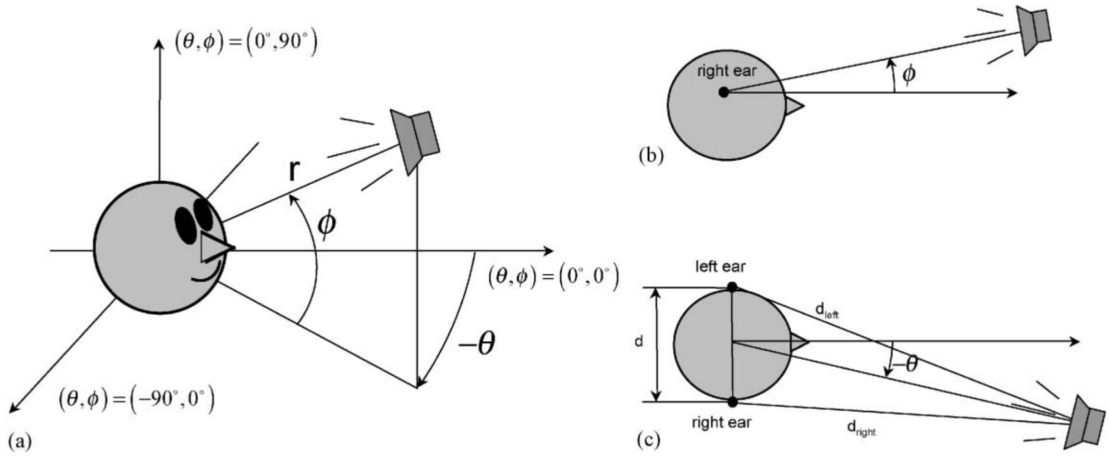

$(\theta,\phi)=(0^\circ,90^\circ)$
r
ϕ
$(\theta,\phi)=(0^\circ,0^\circ)$
$-\theta$
$(\theta,\phi)=(-90^\circ,0^\circ)$
(a)
right ear
ϕ
(b)
left ear
d_{left}
d
$-\theta$
right ear
d_{right}
(c)

Fig. 1. Head-centric reference frame (a), lateral (b) and top view (c). The position of a sound source at a distance *r* from the listener can be represented by means of the angles of azimuth and elevation (being, respectively, θ and ϕ).

particular it has been shown that, given a constant azimuthal angle, the ILD varies with continuity with the elevation of the sound source [27]. However, the IPD and the ILD by themselves are not sufficient to explain the ability of some species (including humans) to perform accurate sound localization. Directional filtering of the part of the external ear known as the pinna, in fact, produces spectral changes that can be used to estimate the position of a sound source even in absence of interaural differences [3]. Given the monaural nature of these cues, the exact mechanism behind this ability is not yet known. However, several computation theories have been proposed [10,29,36].

In the animal world, the barn owl is the one species with the finest sound localization apparatus [19]. This is because in nocturnal hunting the owl relies mostly on sound to locate the prey. Further, because hunting involves flying, in order to plan a successful attack trajectory the barn owl has to locate its prey in both the horizontal and vertical direction with respect to itself. The localization system of the barn owl is unique in many aspects. Due to the peculiar disparity of the arrangement of the feathers covering the left and the right ear, the ILD at high frequencies is highly dependent on the position of the sound source on the vertical plane. The IPD and the ILD at low frequency are used to estimate the horizontal position [22].

Each of the above localization cues is inherently ambiguous if considered on a single or narrow frequency range (this can be easily understood considering that periodic signals produce periodic values of the IPD; the same is still true for the ILD due to less obvious geometric properties of head and ears). In any case, natural sounds are usually broadband and, because the IPD and the ILD change as a function of frequency, it has been suggested that an integration procedure is possible [5]. As long as pure tones are considered, it is still possible to speak in terms of the IPD; conversely in the case of non-periodic broadband signals, the interaural time difference (ITD[1]) is more often considered as a global measure directly related to the angular position of the sound source in the horizontal plane.

Some studies such as those of Knudsen and co-workers [4,20] also investigated where in the brain and what signals might be at the basis of the plasticity of the spatial representation that is found in the *optic tectum* (or *superior collicolus* (SC) in mammals) at the end of the sound localization pathway. They reared some barn owls with distorting prisms (shifting the visual world horizontally) and observed which sort of modifications appeared in the response of some collicolar neurons. They noted that although the hearing system by itself had been left untouched, the localization of sound sources in the dark suffered

[1] ITD is the delay between the instants an auditory signal reaches the left and the right ear.

roughly the same error of the visual system. What has happened? The acoustic representation was shifted according to the distortion produced by the prisms. On the basis of these results they concluded that one sensory modality (i.e. vision) dictated how the other (i.e. sound) had to develop in order to generate a coherent percept. [2]

This multi-sensory map has to be subsequently linked to a motor map, which drives the orienting behavior of the animal. The SC is a possible site where this link takes place [25]. SC neurons were studied extensively and are thought to mediate through efferent pathways the orienting behavior by moving the eyes, head, and body. Plastic changes in the visual, acoustic, and motor maps are thought to depend on visual feedback, but other explanations are possible involving either sound or motor cues alone. In some experiments in fact Knudsen and Mogdans have been able to show that plastic changes might be observed even in the absence of vision [21].

Engineers have proposed several solutions to the problem of sound localization and, sometimes, artificial systems were realized [6,7,9]. All these approaches used a head mockup made of plastic together with rubber pinnae in order to extract the estimation of the position of the sound source mainly from the knowledge of the Head Related Transfer Function (HRTF). In robotics active sensors such as ultrasounds provided usually some sort of acoustic sensation. They were used particularly for navigating through an unknown environment and building sometimes occupancy maps for later use (see, for example, [24,35]). Though there have been examples of use of acoustic cues for navigation [13], the recent appearance of humanoid robots, and their consequent interaction in a human populated environment, increased the relevance of sound localization [1]. In some cases there have already been attempts to implement the localization procedure by means of connectionist models [15,31,32]. In the case of Rucci et al. [32] only a single camera was employed and localization was limited to the horizontal direction. In spite of this, they modeled accurately the processing pathways by means of artificial neural networks and showed as this mapping led to an appropriate visuo-acoustic integration. The work of Rosen et al. [31] also proposes an accurate model of the neurons in the ascending pathway to the optic tectum. Irie [15] took a different approach designing a neural network to learn the relationship between a middle-level representation of the binaural signals (i.e. a filtered version of different portions of the signals) and the position of the sound source in space. In this case vision was used as the feedback signal for a self-supervised backpropagation training of the neural net.

The goal of this paper is to propose a biologically plausible, functional model for the acquisition (i.e. learning) of appropriate visual, acoustic and motor responses for a humanoid robot interacting in a real environment. We investigated how acoustic and visual feedback might be used to autonomously develop a representation for directing the gaze toward visuo-acoustic sources. In particular, we investigated how a neural-like representation can be assembled autonomously; we can talk of self-supervised learning. This approach is similar to the one proposed by Kuperstein [23] although implemented in a real rather than simulated robot. With respect to the implementation proposed by Irie [15], our goal was to learn the sensori-motor coordination rules, and not the localization per se. Another example where sound and vision were integrated can be found in [12,28]. In this latter case though, the emphasis was more on the localization coupled to source separation rather than the learning aspects we focused on. Finally, in all these examples, only the azimuthal component was considered and in some cases more than two microphones were employed. The core aspect of the paper is on showing that (i) two microphones are enough and (ii) visual and auditory cues can be combined in such a way that autonomous on-line learning is possible. In the next section (i.e. Section 2) a more detailed description of the robot head and the background information is provided. Section 3 provides the rationale behind the experiments and the assumptions made. Section 4 develops the complete control and learning model in details with emphasis on the acoustic aspects, which are the major contribution of this paper. Experiments showing the performance of the system are presented in Section 5. Finally, Section 6 draws the conclusions and discusses the results.

[2] Plasticity in the acoustic representation has been observed also in humans [11].

2. Experimental setup

The experimental setup consists of a 5 degree-of-freedom robot head. It can independently pan the eyes, which are mounted on a common tilt. Further, the head can pan and tilt at the level of the neck. As for the visual sensing the robot acquires and processes images in a space-variant format also known as log-polar [34]. The robot's eyes observe the world through a high-resolution *fovea* and a lower resolution periphery. Fig. 2 shows an example of the actual images used for the experiments (b), together with a picture of the setup (a). Note the "tennis-like" balls covering the cameras and the asymmetric ear lobes on top of the head. In this respect, the microphones and the ear lobes were mounted, for practical reasons, on the common tilt of the eyes, i.e. the only feasible mechanical arrangement was on the link connecting the cameras. The robot mount possesses also a *vestibular* sense provided by a three-axis gyro mounted on the neck. The role of inertial sensors within the robot controller was studied elsewhere; a detailed description can be found, for example, in Panerai et al. [30]. Each joint has also a "proprioceptive" sense provided by high-resolution optical encoders.

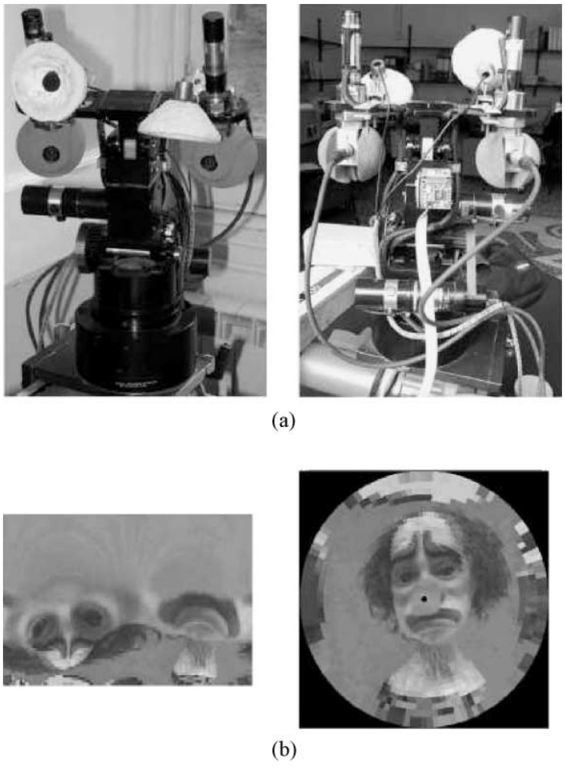


Fig. 2. (a) The experimental setup—front view on the left and back view on the right. The microphones and the ear lobes are mounted on the topmost part of the head. The cameras are within the two "blue" tennis-like balls. (b) An example of log-polar image as acquired by the robot visual system (left) and its reconstructed Cartesian counterpart (right). The resolution is 64 eccentricities by 128 angles. Note the different resolution in the fovea with respect to that in the periphery.

As far as the signal processing is concerned, the ITD and the ILD were computed for sound localization, while color segmentation was used to visually locate interesting objects in the environment. Color segmentation in this case proved to be reliable and robust as a source of positional information. The particular algorithm implemented here is also flexible enough, i.e. it is not devoted to segment but only to a particular color. In other words the robot utilizes a cueing mechanism (e.g. motion detection) to initially spot the possibly interesting objects in the environment and only in a second stage identifies the "principal" color of the object for tracking, saccading, etc. [26].

The robot is controlled by a set of Pentium-based machines running Windows NT and connected through a 100 Mbps Ethernet. The whole software architecture is based on DCOM, an object-oriented standard that among other functionalities allows remote-calling methods through a network. Visual processing is carried out by one machine, sound processing by another, and motor control by a third one. A fourth computer is dedicated simply to data logging, display, and control of the status of the system.

In summary, the platform employed for the experiments consisted of the following:

- a 5 degree-of-freedom robot head mount as described above;
- two cameras and a software simulation of the space-variant (log-polar) geometry;
- a couple of electret condenser microphones Sony ECM-T140 and plastic ear lobes;
- four Pentium-based machines (ranging from 400 to 750 MHz), two frame grabbers, a motion control board (Motion Engineering), and a Sound-Blaster PCI 128 for 44.1 kHz and 16 bits for sound digitalization.

3. Model and assumptions

As discussed above in Section 1 the principal cues for sound localization are the ITD and the ILD. They are not the only cues employed by humans but they are the most easily extracted by processing the sound waves impinging on the ears or microphones, and we assumed they possess an almost one-to-one relationship with azimuth and elevation. In the artificial implementation they are estimated by computing the normalized cross-correlation and the ratio of the average power, respectively. In the barn owl the same processing is thought to be carried out through two different pathways starting from the cochlea and converging in the central nucleus of the inferior colliculus (ICc). Neurons of the ICc are directly connected to the external nucleus of the inferior colliculus (ICx) where, finally, an auditory map of space can be found. At least in the barn owl, ICx is thought to be the first stage where plasticity could be noted [5]. Moreover, ICx neurons showed to be sensitive to the delay between the sound signals: the final response of the neurons in this region seems to reflect a cross-correlation-like processing, albeit with a high immunity to noise and echoes [17]. It has been argued that the generalized cross-correlation [18] by compensating for noise and echoes can provide a response similar to that of the neural pathway of the barn owl [33].

Given the ITD and the ILD there are still two separate problems to be solved: (i) how to integrate visual and acoustic spatial information; (ii) how to convert the spatial percept in an appropriate motor command.

The former is known to be solved in the barn owl by aligning the representations of the visual and acoustic space. This kind of plasticity is driven by visual information but changes were also observed in the absence of vision. Although this is relatively simple for the barn owl, which has a limited ocular mobility, the same is more complicated for primates including humans, which possess an efficient oculo-motor apparatus. In this case, the two maps have to be kept aligned by taking into account the position of the eyes with respect to the head. The retino-centric representation of the visual space might shift with respect to the head-centric representation of the acoustic space. This mechanism has to include proprioceptive or efference-copied information to dynamically realign the maps.

Our robot implements such an aligning mechanism, although a simplified version of it, i.e. for a given eye–head orientation and only for the azimuthal component. A bimodal map integrates vision with sound and generates the motor commands required to gaze toward a sound, a visual or multi-modal stimulus. In this case, the movement of the head is subsequently accomplished by driving the system in order to roughly face the target by rotating the neck. Adaptation in this case is driven by vision. Even if vision might drive learning appropriately, we decided also to explore whether sound alone could generate an appropriate spatial map (of course of sound alone in this case). A second experiment by using the ITD and the ILD as teaching signals is thus reported.

The second problem, i.e. generating the proper motor commands, is thought to be solved by the SC in mammals, which is interconnected with other brain structures, including premotor and motor nuclei in the brainstem and spinal cord. In our model, maps convert their inputs into motor commands, and the motor response is tuned on the basis of a measure of the current performance of the system. This measure might be either visual or acoustic.

It is worth noting that when we use the word “localization” referring to our artificial system, we never mean the actual computation of either the position of the target in Cartesian coordinates or the identification of the direction of the source with respect to the robot. What we mean is always the estimation of some quantities that can be used to move the robot in a meaningful way with respect to the environment, i.e. gazing toward the sound. It is the “pragmatic” of the perception what we are interested in, in other words, which sort of signals might represent perception with respect to the task to be solved.

To be more precise the experiments will show the following:

- The robot can autonomously learn the mapping between the retinal position of the target and its acoustic counterpart (e.g. the ITD).
- The robot can convert the “fused” percept into a motor command, which can be used to generate

saccades toward an acoustic, a visual, or a visuo-acoustic target.

- The robot can employ only acoustic cues to learn a map of the sound space, which allows moving the neck toward a noisy target without any visual contribution.

To recapitulate the assumptions made: the system knows that the ITD and ILD are in an almost one-to-one relationship with azimuth and elevation, respectively. This is not a limiting assumption because it has been shown [8] that it is indeed possible to easily learn this relationship (for example, in the form of a gain matrix). The same is true if we consider the retinal error instead of the ITD and ILD pair. Given the gain matrices the proposed model can learn all the inverse maps needed to align visual and acoustic cues, and to generate a proper orienting eye–head movement as detailed in the next sections; no further a priori knowledge is required.

4. Computation of ITD and ILD

4.1. Interaural time difference

In the previous sections, we mentioned that the ITD is the main cue for estimating the horizontal component of the position of a sound source. The sound waves impinging on the microphones are delayed one with respect to the other by an amount, which is a function of both the azimuthal angle of the source and the relative distance between the microphones or baseline d (see Fig. 1(c)).

By applying this definition of the ITD, we can work out the relationship between the geometry of the head, the microphones, and the sound source:

$$\mathrm{ITD} \triangleq \frac{\Delta R}{c}, \tag{1}$$

where ΔR is the difference of space traveled by the sound wave to get from the source to the left and right microphone, and c is the speed of sound. From simple geometry (see also Fig. 1(c)), considering a target positioned at θ degrees on the horizontal plane and at a distance r from the center of the baseline d, we can expand Eq. (1) into

$$\begin{aligned}\mathrm{ITD} &= \frac{d_{\mathrm{right}} - d_{\mathrm{left}}}{c} \\ &= \frac{1}{c}\left(\sqrt{r^2 + \frac{d^2}{4} + dr\sin\theta}\right. \\ &\qquad \left. - \sqrt{r^2 + \frac{d^2}{4} - dr\sin\theta}\right),\end{aligned} \tag{2}$$

which, under the assumption $r \gg d$, becomes

$$\mathrm{ITD} \approx \frac{d}{c}\sin\theta. \tag{3}$$

This function is monotonic and continuous within the interval $\pm\pi/2$. The maximum and minimum values ($\pm d/c$) are obtained in the limit, when θ is $\pm\pi/2$. If the computation is carried out in the discrete, then the ITD is measured as number of samples k. The relationship between the two previous quantities ITD and k is

$$\mathrm{ITD} = k\frac{1}{f_{\mathrm{s}}}, \quad k = f_{\mathrm{s}}\frac{d\sin\theta_k}{c} = N\sin\theta_k, \tag{4}$$

where f_{s} is the sampling frequency. The ITD samples, in this case, are not uniformly distributed. They are denser near $\theta = 0$ and sparser as we approach the limits. Given Eqs. (3) and (4) the distance between two consecutive values of ITD is

$$\Delta\theta_k = \frac{1}{\sqrt{N^2 - k^2}}, \quad k \in [-N+1, N-1], \tag{5}$$

which is obtained by deriving the inverse of Eq. (4), and where

$$N = \left\lfloor \frac{d}{c} \cdot f_{\mathrm{s}} \right\rfloor. \tag{6}$$

For example, in our implementation the distance between the microphones is about 16 cm, that leads, along with a sampling rate of 44,100 Hz, to $N = 21$ and a maximum accuracy of about 2.5°.

The actual computation of the ITD is carried out by using the generalized cross-correlation method as described by Knapp and Carter [18]; as shown in Fig. 3, the signals are initially filtered and subsequently the ITD is estimated as the maximum of their cross-correlation function. The goal of the initial filtering is that of preshaping the resulting cross-correlation in order to improve the peak detection procedure.

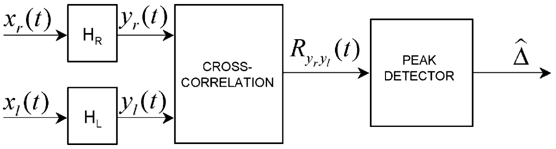


Fig. 3. Estimation of the ITD, general schema. The maximum of the cross-correlation function $R(t)$ is taken as a measure of the delay between the signals; the waveforms are prefiltered in order to enhance both accuracy and reliability (see text for discussion).

Though the optimal filtering would require some a priori knowledge about source and noise spectra, we decided to avoid any noise estimation procedure and made no assumptions about the spectral characteristics of the sound sources to be located.

If $G_{y_r y_l}$ is the cross-spectrum of the filtered signals (y_r and y_l), then their cross-correlation can be obtained in the frequency-domain by inversely Fourier transforming:

$$R_{rl}(\tau) = \int_{-\infty}^{+\infty} G_{y_r y_l}(f)\, e^{j2\pi\tau f}\, df. \quad (7)$$

The effect of the filters is to weight the cross-spectrum of the original signals, so that

$$G_{y_r y_l} = H_r(f) H_l^*(f) G_{x_r x_l}(f). \quad (8)$$

Knapp and Carter described several methods to choose H_r and H_l; we implemented and compared the Phase Transform (PHAT) and the Smoothed Coherence Transform (SCOT), which, at least under the above mentioned assumptions, provided comparable results. According to the SCOT H_r and H_l are chosen as follows:

$$H_r(f) H_l^*(f) = \frac{1}{\sqrt{G_{x_r x_r}(f) G_{x_l x_l}(f)}}. \quad (9)$$

Substituting Eqs. (8) and (9) in Eq. (7),

$$R_{rl}(\tau) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{G_{x_r x_r}(f) G_{x_l x_l}(f)}} \cdot G_{x_r x_l}(f)\, e^{j2\pi\tau f}\, df. \quad (10)$$

Eq. (10) can be interpreted as a prewhitening of the signals before computing the cross-correlation. Since the side effect of the filter is to emphasize those parts of the spectrum characterized by a low signal-to-noise ratio, their contribution was suppressed by employing a heuristically estimated thresholding procedure (i.e. where spectra were below a certain threshold, no computation was performed).

The SCOT provided a good response to both broadband signals (e.g. white noise) and relatively narrowband signals (e.g. voice). Figs. 4 and 5 show examples

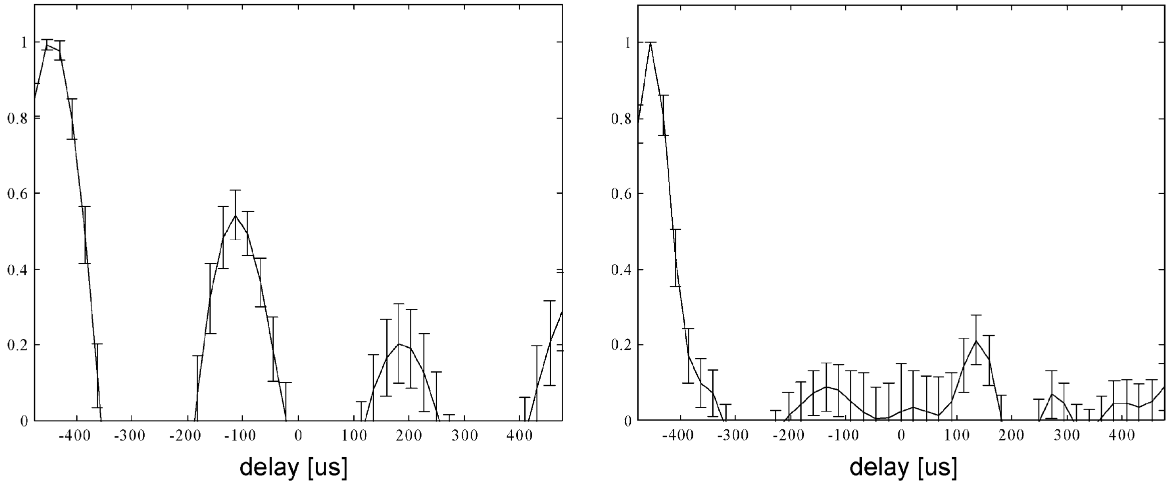


Fig. 4. Comparison between standard cross-correlation (left) and Smoothed Coherence Transform (right) for a broadband sound source. The computation was carried out over 2048-sample long windows out of about six seconds of data. Results were averaged and standard deviation computed.

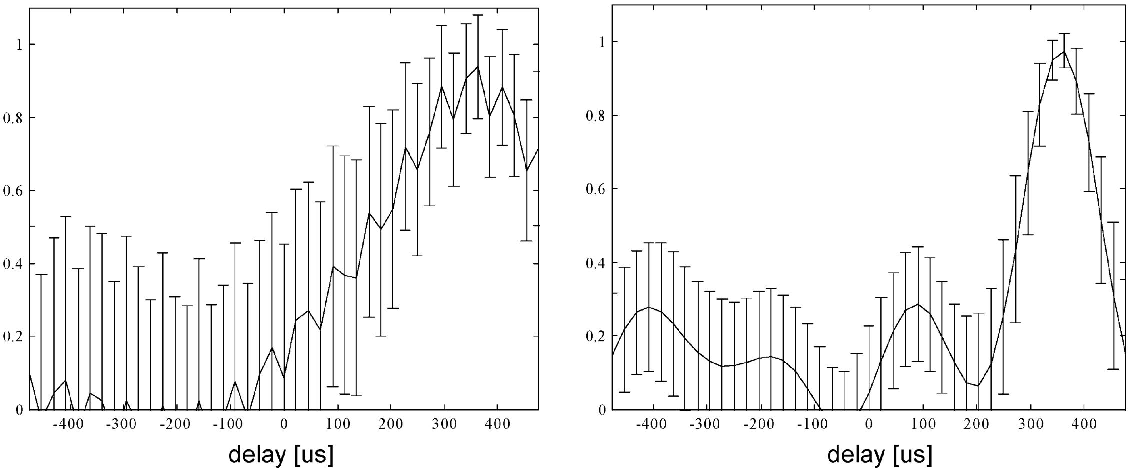


Fig. 5. Comparison between standard cross-correlation (left) and Smoothed Coherence Transform (right) for a narrowband source (voice). The computation was carried out over 2048-sample long windows out of about six seconds of data. Results were averaged and standard deviation computed.

of cross-correlation functions obtained by using real voice and noise data. It is worth noting that the SCOT (right column) provides a much sharper peak, which of course enhances the peak detection, and thus the ITD estimation in both cases.

4.2. Interaural level difference

With only two microphones—unlike the human ears—there are no cues that can be exploited to estimate the elevation of the sound source (Fig. 1(b)). As mentioned in the introduction we had to break the symmetrical configuration and to further employ artificial ear lobes meant to convoy the sound coming from a particular direction. This, along with the different orientation of the microphones (respectively upward and downward for the right and left ear), made the ILD to directly depend—at least at the frequencies above a certain value—on the elevation of the target source. The ILD is hence computed as a function of the average power of the signals:

$$\mathrm{ILD} = 10 \log \frac{\int G_{x_l x_l}(f)\,\mathrm{d}f}{\int G_{x_r x_r}(f)\,\mathrm{d}f}. \tag{11}$$

Experiments with either broadband or narrowband signals led to the decision to compute the ILD by taking into account only the frequency in the range 3–10 kHz. This is because in that frequency range, the effect of the different orientation of the microphones and of the ear lobes is stronger and provides a better dynamics. This effect is of course specific to our experimental setup: a different arrangement might require a different filtering.

4.3. System overview

The block diagram of the complete system is shown in Fig. 6. As mentioned before, we used standard equipment, made up by a couple of electret condenser microphones and a preamplifier. A SoundBlaster PCI 128 at 16 bits and 44.1 kHz sound card samples the signals, while the processing (the estimation of the ITD and ILD) is carried out every 46 ms by an x86 processor using Intel Signal Processing Library [14].

The overall performance was measured using a white noise source, placed in about 200 different locations, randomly distributed within the interval of ±60° for the angle of elevation and ±80° for the azimuth angle. In each location the values of the ITD and the ILD were recorded several times (more than 10) and averaged yielding the maps shown in Fig. 7. The two-dimensional maps represent the measured signals as a function of the position of the sound source. Fig. 7 shows also that the system performs as expected, though it is clear that the estimation of the ILD is reliable only in a relatively small region at

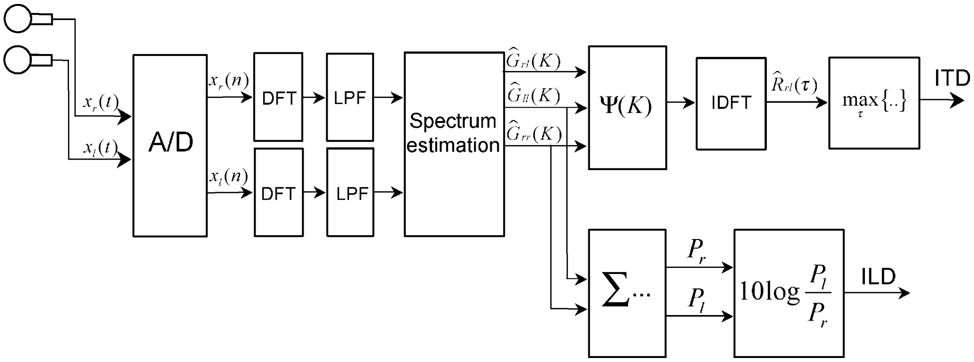


Fig. 6. System overview. Sound is acquired through the SoundBlaster PCI 128 (indicated here by A/D) and Fourier transformed. The initial low pass filters remove the frequencies above 10 kHz. Spectra and cross-spectrum are subsequently estimated and used to compute the ITD and ILD. The cross-spectrum is weighted and inversely transformed to estimate the cross-correlation function, and, eventually, the ITD (as described by the SCOT algorithm). The two spectra are also used to determine the ratio between the average power of the signals and, through its logarithm, the ILD (see text).

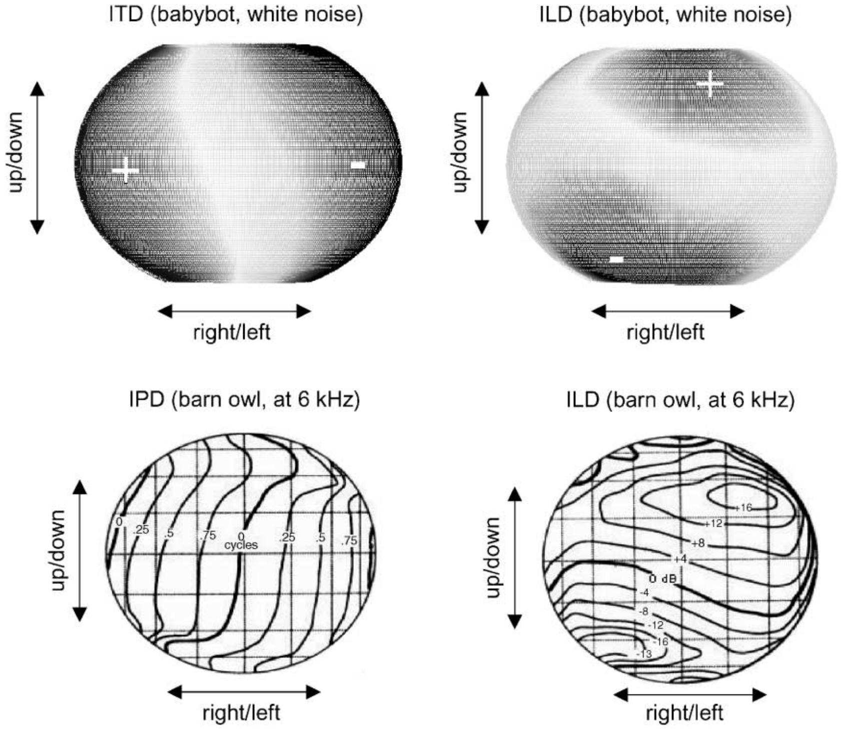


Fig. 7. The ITD and ILD as a function of θ and ϕ. A sound source was positioned in proximity of the robot's head at 200 different positions randomly distributed on the surface of an imaginary sphere. The ITD and ILD vary, respectively, within the interval of ±440 μs and ±12 dB (regions relative to positive and negative values marked, respectively, with '+' and '−'). The lower pictures show the same measurements taken in the barn owl although within different frequency ranges (pictures adapted from [8] with permission of the authors). In spite of some differences—mainly in the absolute values of the signals—a strong resemblance can be noticed. Both graphs are plotted from the listener's point of view.

the center of the axes; outside this region, and even at high frequencies, the ILD no longer depends on the elevation alone and correlates also to the position in the horizontal plane. The latter effect is probably due to a reduced directionality of the ear lobes as the azimuth increases. Besides, it is important to note that, because of the different inclination of the microphones, the median plane defined as the locus of the points equidistant from the ears—where the ITD is zero—is tilted. For all these reasons the ITD and the ILD are neither mutually orthogonal nor independent as shown also by the contour lines in Fig. 7.

Finally, the lower panel of Fig. 7 shows the response of the barn owl external ear measured by inserting two small microphones within the ear canals [5]. Basically the plots show the response—in terms of the ITD and ILD—of the particular configuration of feather and trough to an incoming sound in a given frequency range. We reported this response here in order to compare it to the response of the robot's ear lobes. They are qualitatively similar, a part from the plot of the ITD which is horizontally flipped because the microphones, in our implementation, besides being tilted are also not horizontally aligned.

5. Experiments

As outlined in Section 3, several experiments were planned, whose main goal was to show the following:

- The robot can autonomously learn the mapping between the retinal position of the target and its acoustic counterpart and convert the "fused" percept in a motor command, which can be used to generate saccades both toward an acoustic and a visual target.
- The robot can employ only acoustic cues to learn a map of the sound space, which allows moving the neck toward a noisy target without any visual contribution.

Two main classes of controllers were employed for these tasks:

- Closed loop, whenever continuous information is available, such as during the smooth tracking of a visually identified target. Closed loop gains were manually tuned in order to obtain stability, i.e. no learning involved. The controller is typically a proportional-integral-derivative (PID).
- Open loop, when only intermittent information is available. Of course a further, possibly non-linear, computation is necessary here to convert the sensory signal into a motor command. The robot acquires the relevant transformations during learning.

5.1. *The control of the eyes*

Consider the problem of moving the eyes toward a spotted acoustic target: there is no sound feedback to be exploited, since, as the eyes move, no variation is elicited in terms of the perceived sound. The solution we suggest here is to take advantage of the integration between acoustic and visual information. The main concern becomes that of matching two signals expressed with respect to different, decoupled, reference frames—eye centered versus neck centered. Formally, whenever a target appears in a certain spatial position, the robot senses two different spatial "quantities" with respect to either an eye centered or neck centered reference frame. The first one, s_{v}, is the *retinal error*, the second, s_{a}, is simply the ITD. Given a particular joint configuration q, a function f links the two quantities:

$$s_{\mathrm{v}} = f(s_{\mathrm{a}}, q). \tag{12}$$

This hypothetical function could be used to match visual and acoustic information or to express an acoustical spatial perception in the eye-centric reference frame. At least in theory, at this point there is no difference between the two signals and a common control system could be used.

However, an approximation of Eq. (12) cannot be easily obtained unless a few other simplifications are made:

- The function f in Eq. (12) is estimated only for a limited range of values of q (i.e. the eyes in an almost-centered configuration with respect to the neck $|q| <$ threshold). As a matter of fact this is not a strong constraint since saccades are mostly generated after a previous movement has been completed (no double saccades are allowed in this experiment) thus the joint configuration is often within the "limited range".
- The task is limited also to one dimension (i.e. horizontal movements). This again is not too restrictive since in our present setup the same schema cannot be applied for tilt movements because the

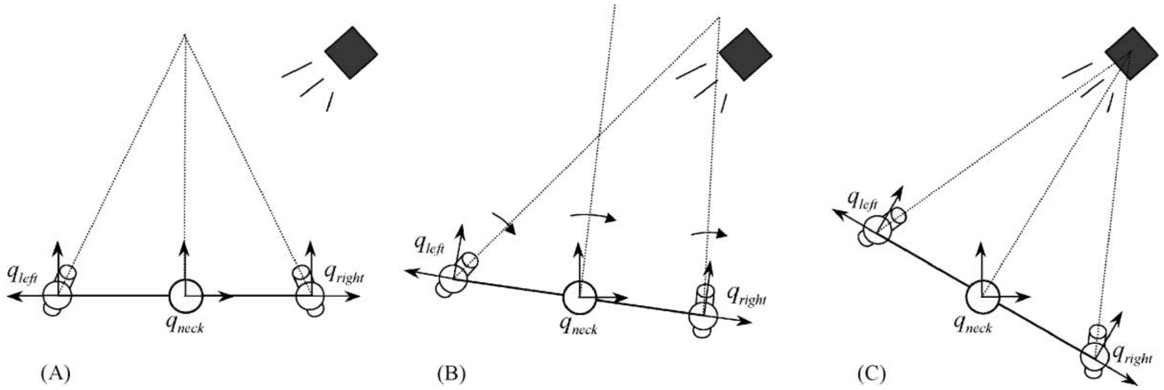


Fig. 8. Eye–head coordination. (A) The situation preceding a saccade (i.e. eyes centered with respect to the neck) with a target appearing within the robot's field of view. The position of the target is estimated using both visual and auditory information and a saccade is thus started. The neck is controlled in order to keep the robot head roughly facing the target. (B) The neck moves in the same direction of the eyes. (C) At movement completion the robot is again in a symmetrical vergence configuration and ready to perform another saccade.

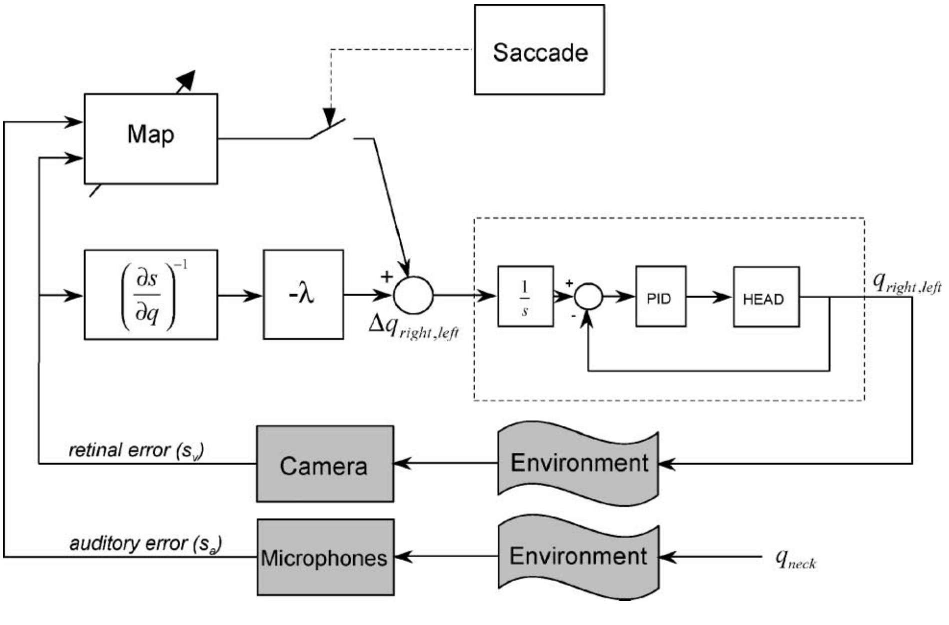


Fig. 9. Block diagram of the eye controller. It consists of two loops: a closed loop and a feed-forward loop. The former uses the inverse Jacobian as in the classical visual servoing approach. The latter consists of an inverse model (indicated by "Map"). Whenever the error is greater than a certain threshold the block identified by 'saccade' issues the start signal; the map puts together auditory and visual error and it is employed to compute a saccade. The schema stresses the fact that the auditory feedback exists only when the neck moves. For this reason, only the visual error drives the learning of the map. The auditory error is used for querying the map at the beginning of the saccade when certain assumptions are satisfied (see text). A low-level PID controller (within the control board) takes care of computing the torques to drive the motors.

microphones are mechanically mounted on the common tilt of the cameras. As a consequence the eyes cannot tilt independently of the ears. It is not difficult though to imagine a similar mechanism applied to vertical movements.

The following visuo-acoustic to motor map is hence estimated:

$$\Delta q = f(s_{\mathrm{v}}, s_{\mathrm{a}}), \tag{13}$$

which blends together visual and acoustic error signals and gives the required motor command in order to foveate the target.

Whenever a target appears, the visual information drives a feedback-error learning mechanism [16], whereas the acoustic information is used only at the beginning of the motion when the coordinate frames are essentially aligned (and q is within the mapped range). In defining a feedback-error mechanism here what is necessary is the simultaneous perception of a

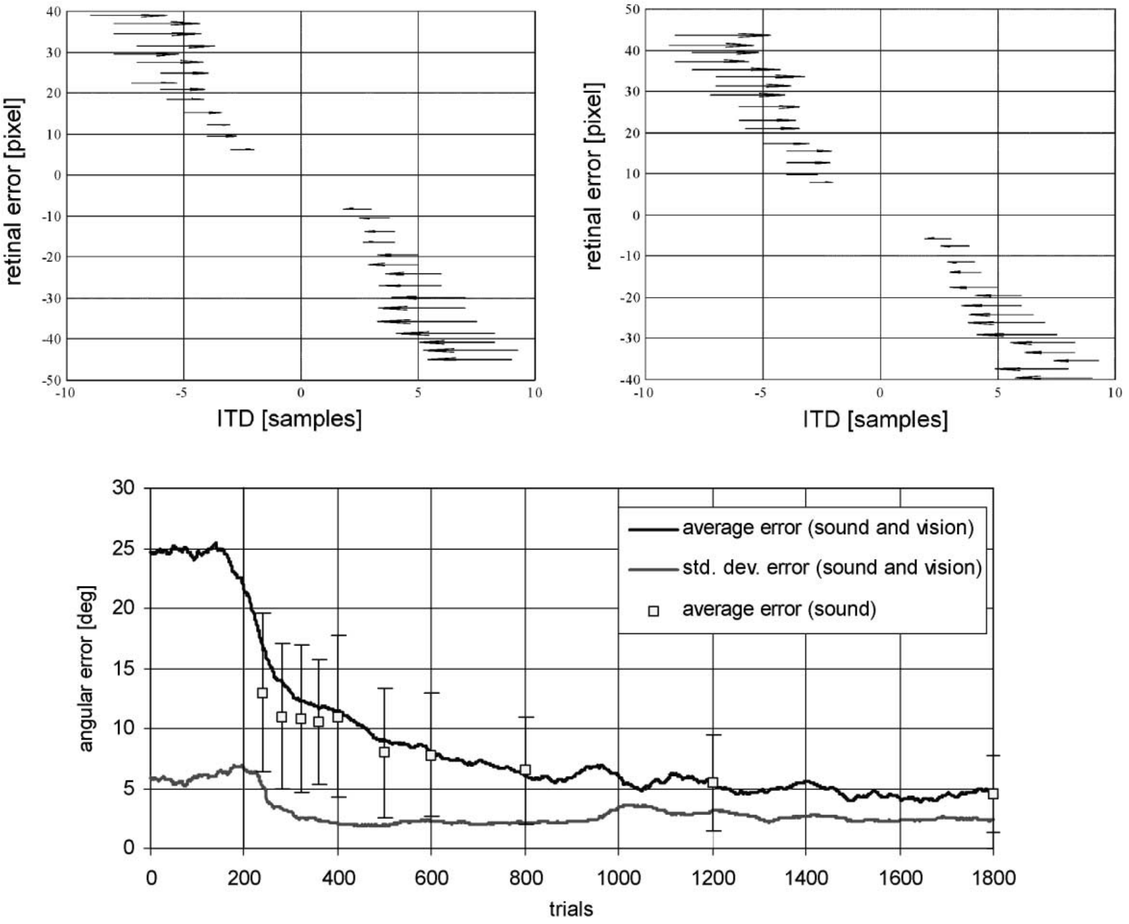


Fig. 10. Learning the eye maps. (Top) Eye maps for the right and left eye respectively. The input is in both cases the retinal error measured in pixels and the ITD measured in samples (respectively visual and acoustic information). The output (vectors) is the motor command required to foveate the target. The motor command here is the velocity required to fulfill a saccade in a given time interval. The linear relationship between the retinal error and the ITD is easily spotted in both maps. (Bottom) Measured residual errors during the learning process. The plots show moving average (solid line—black) and standard deviation (solid line—light gray) computed over 100-sample long windows. The process was interrupted from time to time (determined in order to sample the learning process non-uniformly) and 100 acoustic random stimuli were presented to the robot to test also the estimation of the visual information (see text)—the average error and its standard deviation is superimposed to the same graph.

visuo-acoustic target, an error signal (i.e. the visual feedback) and how to subsequently query the learnt map. By approximating the head with a simple integrator (i.e. assuming a negligible head dynamics and a high stiffness low-level PID controller), the feedback loop is constructed by using visual information as follows:

$$\Delta q_{\mathrm{left}} = k \cdot s_{\mathrm{vleft}}, \qquad \Delta q_{\mathrm{right}} = k \cdot s_{\mathrm{vright}}, \tag{14}$$

where k is the proportional gain and Δq_{left} ($\Delta q_{\mathrm{right}}$) is the velocity motor command for the left (right) eye. k was chosen beforehand in order to obtain a stable closed loop system. The position on the image plane s_{vleft} (s_{vright}) represents the error for the foveation task (zero meaning perfect foveation). This sort of controller is proven to be stable although the performance is not always uniform because dynamics is not properly accounted for [2].

When a saccade is attempted, its precision/performance is evaluated and eventually criticized by the closed loop controller (overshoot or undershoot of the target). This resulting error signal can be used to modify the future behavior of the system as follows:

$$\Delta f(s_{\mathrm{v}}^{t-n}, s_{\mathrm{v}}^{t-n}) = k_{\mathrm{learn}} \cdot s_{\mathrm{v}}^{t}. \tag{15}$$

Eq. (15) simply says that the current estimate of the function f computed in $(s_{\mathrm{v}}^{t-n}, s_{\mathrm{a}}^{t-n})$ is changed by an amount proportional to the retinal error s_{v}^{t} at the end of the saccade. Time is made explicit and t represents the end of the saccade, while $t-n$ is the instant the saccade is started (of length n). Moreover, when a sensory pair $(s_{\mathrm{v}}^{t-n}, s_{\mathrm{a}}^{t-n})$ is observed, an approximation of Eq. (12) is learnt too. In our constrained experiment Eq. (12) can be well approximated by

$$s_{\mathrm{v}} = a \cdot s_{\mathrm{a}} + b, \tag{16}$$

where s_{v} (retinal error) and s_{a} (ITD) are scalars, and a and b are two parameters estimated using an on-line least-square algorithm. This relationship appears to be linear in our case although this is not generally true for every possible experimental arrangement. In that case, a more general neural network or suitably chosen polynomial can replace the linear approximator.

The map (Eq. (13)) and Eq. (16) can be used even if one of the two sensory components is not available. Roughly speaking, when visual information is not available, Eq. (16) is employed to determine the "missing" visual percept and address the map correctly to determine the saccadic motor command.

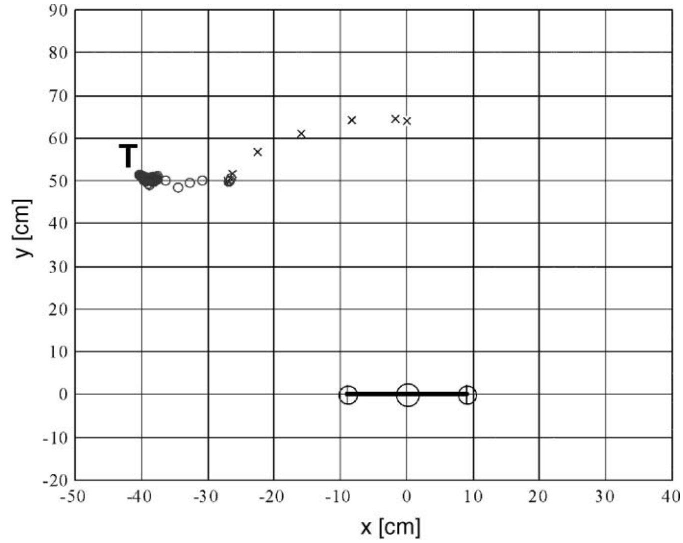


Fig. 11. An exemplar trajectory of the fixation point: top view. The simple sketch represents the robot—small circles are the eyes and big circle is the neck. A target originally outside the robot's visual field generates a short sound; the gaze shifts toward it along an iso-vergence line (cross marks). The saccade is accurate enough to bring the target near the fovea in both retinas. A few "visual" closed loop control steps then complete the movement (circle marks). It is worth noting that part of this second movement is needed to adjust vergence (depth). The neck was not moved in this particular experiment.

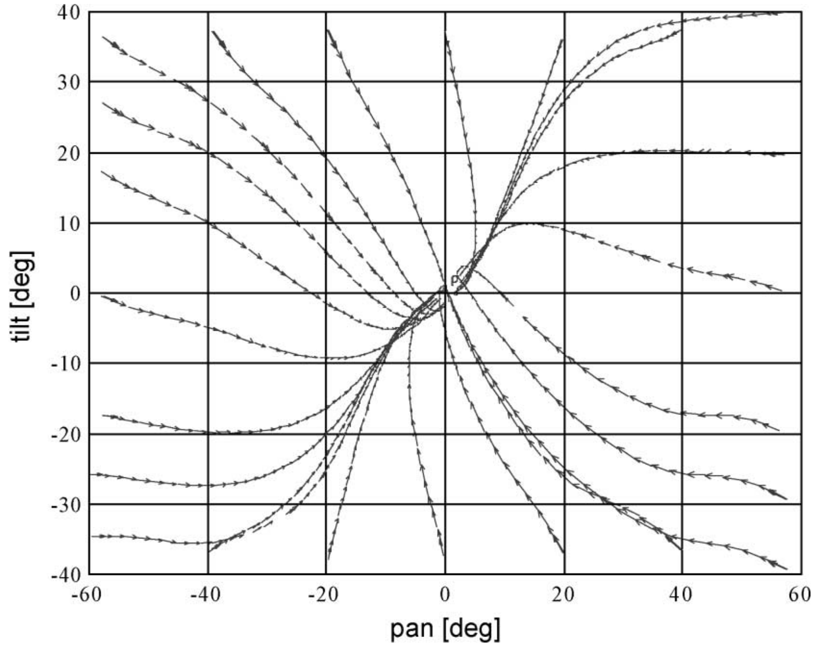


Fig. 12. Head trajectories, closed loop. Some trajectories of the head in the space (pan, tilt), in presence of a central target. The convergence is guaranteed only inside a relatively small region. Outside it, the convergence of the ITD drives the system toward the center of the axis where, at last, the error is zeroed.

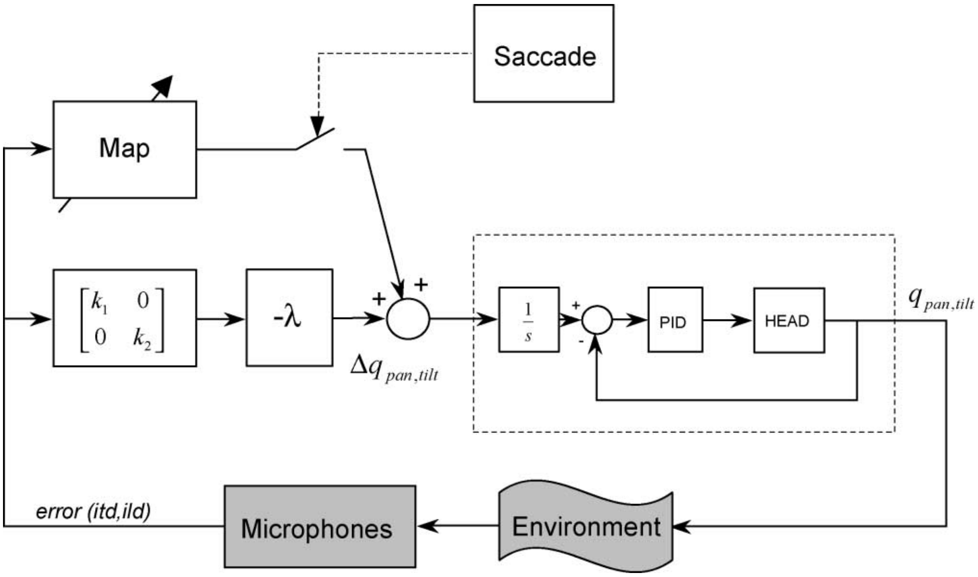


Fig. 13. The neck control schema. It is made up of two loops: a closed and an open loop controller. The gain k_1, k_2, and λ are tuned beforehand in order to obtain stability. Whenever the error is greater than a certain threshold the block identified by 'saccade' issues the start signal; the map is employed to compute a saccade. A low-level PID controller (within the control board) takes care of computing the torques to drive the motors.

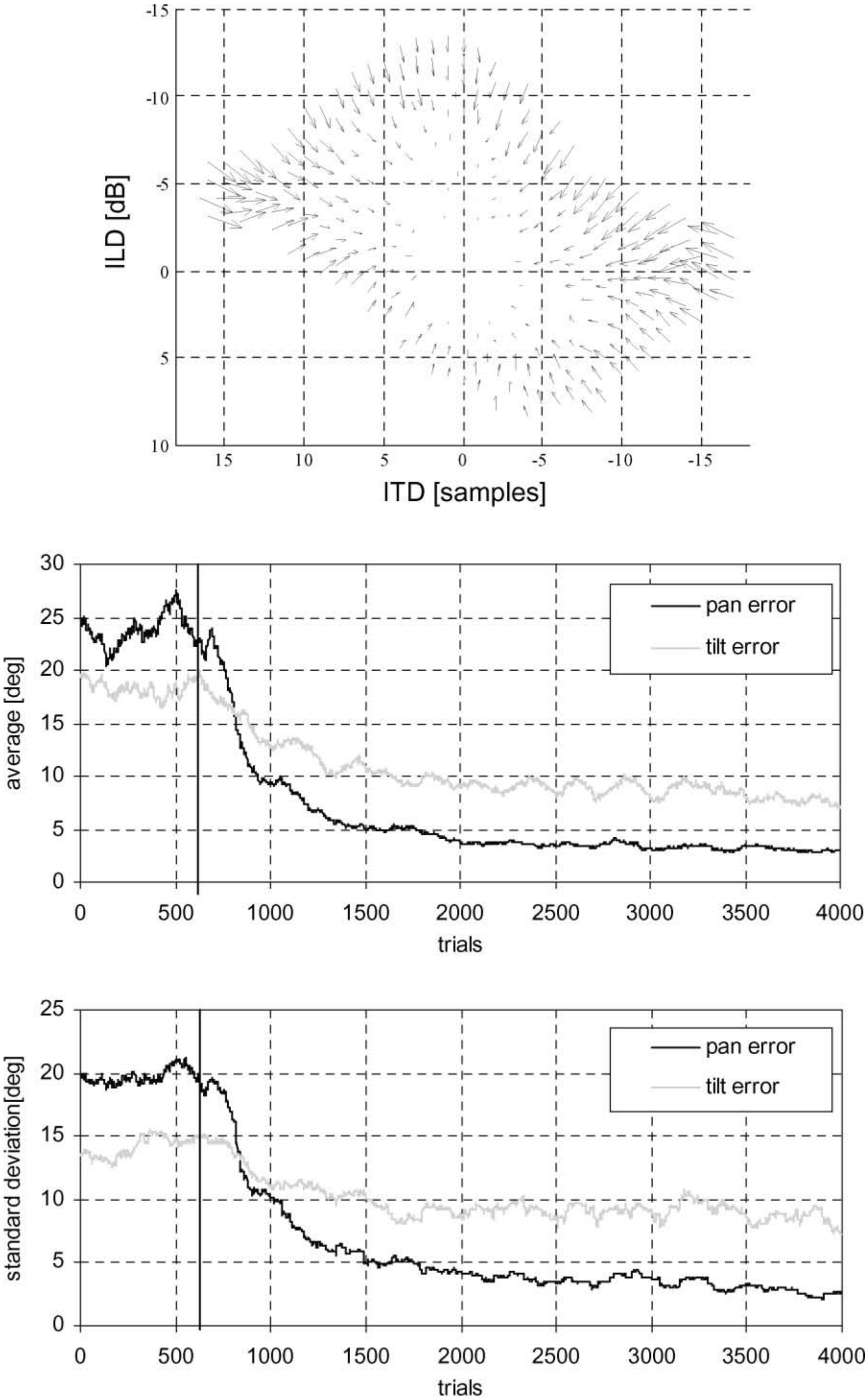


Fig. 14. Learning the neck map (1). Moving window average and standard deviation (both over windows of 150 trials) of the residual error at the end of the saccade. After the activation of the learning (vertical solid line) a sharp increase of motor performance can be noted. The topmost panel shows the map as obtained at the end of the learning phase: the output is the required saccadic command, the input the initial error (ITD and ILD).

Within our conceptual schema thus Eq. (16) plays the role of keeping visual and acoustic information aligned (coordinate conversion), and Eq. (13) acts as the SC by generating the appropriate motor command from the sensory information.

A final note for what concerns the control of the neck. In this experiment, it has been controlled in order to keep the robot head roughly facing the target, i.e. once the eyes start moving, also the neck moves in the same direction so that eventually, at movement completion, the robot is again facing the target in a symmetrical vergence configuration. Formally,

$$\Delta q_{\text{neck}} = \text{PID}(q_{\text{right}} - q_{\text{left}}). \tag{17}$$

This strategy is advantageous since it keeps the system ready for a new movement and maximizes the chances of being able to correctly complete subsequent movements by keeping the head far from limit configurations (see Fig. 8). Fig. 9 shows the complete block diagram of the implemented controller. The low-level PID controller is indicated, together with visual and acoustic loops. Maps are activated by a common logic whenever the conditions to generate a saccade are met (e.g. a target is not already foveated).

During the experiments reported here, learning was stopped from time to time and a sequence of 100 saccade performed using only acoustic stimuli. The visual position was recovered from Eq. (16); average and standard deviation of the error computed to be compared with the average and standard deviation of the error measured during learning. Fig. 10 shows this result and the effectiveness of the learning procedure: saccades become more precise, and the number of closed loop control steps required to foveate the target is reduced. At the same time, the ability of the robot to move the eyes toward an acoustic stimulus is increased. Maps shown (topmost plots) were simple nearest-neighbor lookup tables. More sophisticated neural network could in principle be used but, for these cases of moderate difficulty, lookup tables were simpler to implement and easier to analyze and tune (e.g. learning rate, number of samples to converge, analysis at a single location, etc.). An exemplar trajectory of the fixation point of the robot in this situation is shown in Fig. 11.

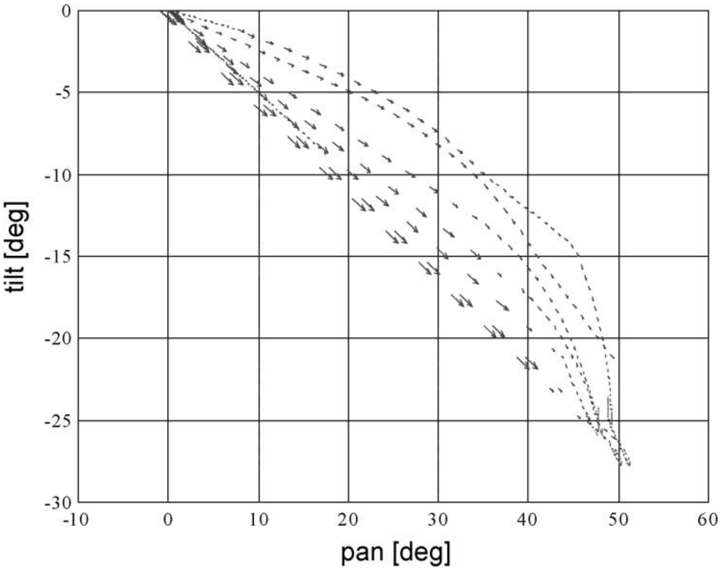
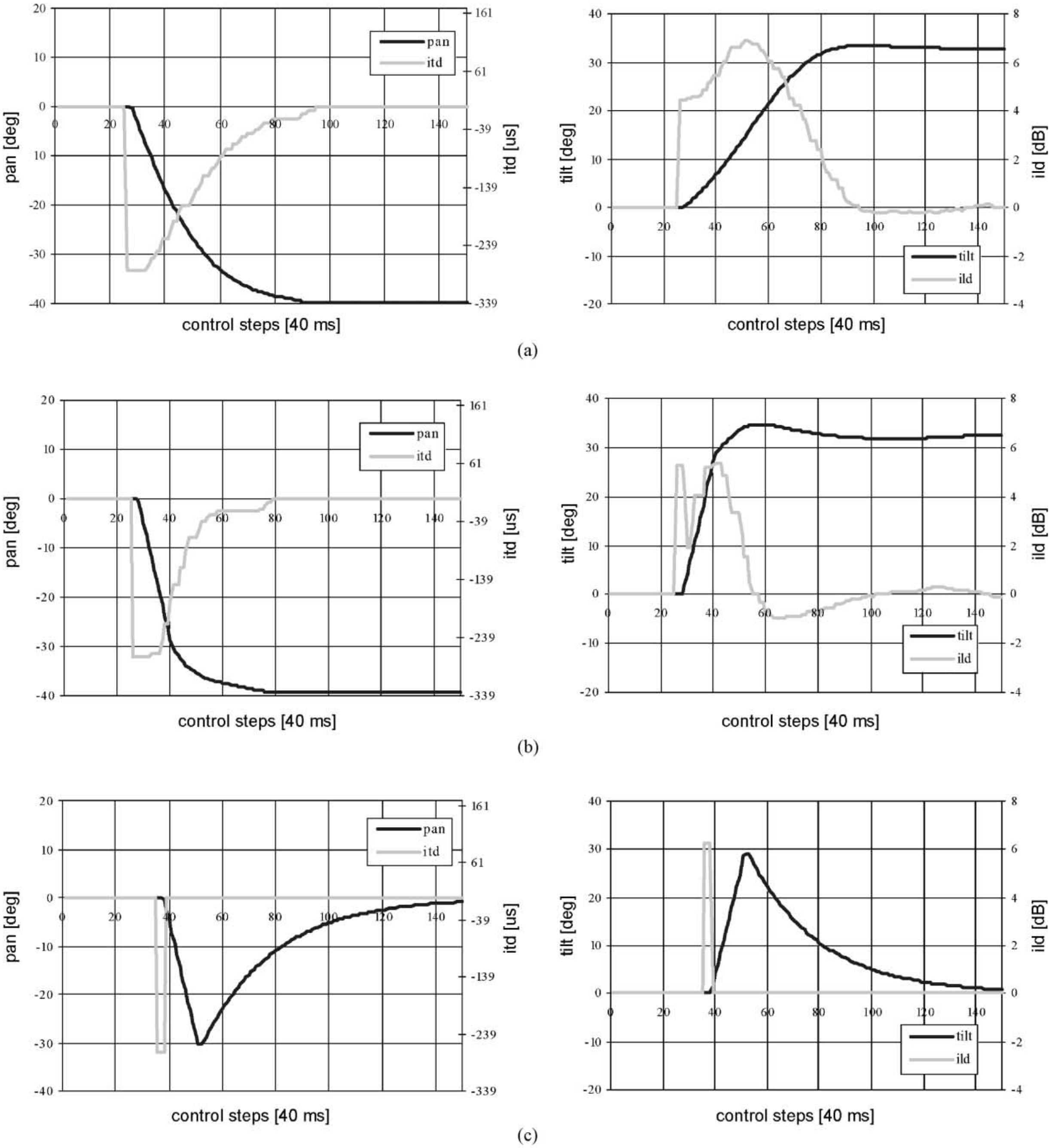


Fig. 15. Learning the neck map (2). The head velocity at each control step is plotted in the (pan, tilt) plane. The same trajectory, relative to a target placed at $\theta \approx 50^\circ$ and $\phi \approx -30^\circ$ is repeated during the learning phase; the increment of performance, is shown by the straightening of the trajectory.



(a)

(b)

(c)

Fig. 16. Examples of trajectories. The error signals (ITD and ILD) and the movement of each joint are shown, entailed by a target at $\theta \approx -40^\circ$ and $\phi \approx 30^\circ$. Two different conditions were tested: closed loop and feed-forward (saccadic) control, after a period of learning. The availability of the sound signal can be noticed from the error plot: in (a) and (b) the sound was "on" for the whole motion; in (c) it lasted only for a few control steps (in this case, at the end of the saccade, the head automatically returns to the zero position). It is worth stressing how an accurate response is obtained even in presence of a very short signal (c).

5.2. Using sound only

This last experiment addresses the issue of the role of acoustic information alone during learning. As in the previous experiment, we propose to use a feedback-error learning schema. The difference here is that the only available feedback signal is sound. Of course, in this case, the eyes do not move; only the neck moves to track the noisy target. It is worth noting that the cameras will tilt as the microphones and the cameras tilt altogether. The two components of the controller need to be defined: (i) the closed loop; (ii) the map to be learnt.

The closed loop controller can be imagined as zeroing the ITD and ILD. According to standard control theory, under certain hypotheses the ITD and ILD can be used directly in order to drive a simple PID controller. Strictly, the ITD and ILD do not jointly respect these hypotheses; it can be proven though that exists a small region around $(0, 0)$ where the stability of the system is guaranteed. Outside this central region (that is whenever the ITD is above a certain threshold), the ITD by itself can drive the system toward $(0, 0)$, so that eventually, the error is zeroed. From the robot behavior point of view, this means that the movement is not straight to $(0, 0)$ but rather the trajectory describes a curve in joint space (see Fig. 12), and sometimes a local divergent behavior of the tilt angle can be noticed.

The map has to approximate the inverse transform between the error and the required motor command Δq, i.e.

$$\Delta q = f(s). \tag{18}$$

Function f is learnt by means of the feedback-error learning mechanism as shown in Fig. 13 where a sketch of the complete control schema is presented. Learning has been carried out in a real environment (no acoustic-isolated environment nor anechoic chamber) by using real and natural stimuli (voice, metal objects, etc.); however, in order to make the whole process repeatable we used also a continuous broadband sound produced by a speaker connected to a computer and placed at the distance of 1.5 m from the robot. The head of the robot was then passively displaced in different positions, for example, $(q_{\text{neck}}, q_{\text{tilt}})$ thus simulating the presence of a target in $(\theta = -q_{\text{neck}}, \phi = -q_{\text{tilt}})$ (θ and ϕ, respectively, azimuth and elevation as in Fig. 1). Afterward, the robot was free to redirect the gaze toward the target generating one or more new learning samples.

Moving window average and standard deviation of the positioning error of the robot were computed. The plots show a quick drop of both quantities after the activation of learning, corresponding to a sharp increase of motor performance (Fig. 14). However, it is also clear that the system performs less reliably when a movement of the tilt axis is involved due to the poorer quality of the ILD. A further understanding of the learning process can be drawn from Fig. 15 where a repetitive saccade is performed toward the same target; it is easy to spot that the trajectory becomes progressively straighter as learning proceeds. A final comparison between the behavior of the robot before and after learning (closed loop versus feed-forward control) is shown in Fig. 16.

6. Discussion

The theoretical and experimental data presented in this paper aimed at demonstrating the following issues: (i) binaural cues such as the ITD and the ILD are valid measures of the position in space of a sound source and can be used to drive appropriate motor responses aimed at foveating the target; (ii) visual information can drive the acquisition of appropriate acoustic-visual to motor maps which can be used to direct gaze toward interesting events in the environment; (iii) even sound alone can suffice in building a spatial map of the "acoustic" environment in cases where the visual information is lacking. Furthermore, integration of visual and acoustic cues might lead to a more complete range of behaviors and, though not analyzed here, to the enhancement of the response in cases where both cues are present.

For humanoid robots these abilities might prove to be fundamental in operating in a human environment where the ability to direct the attention toward different speakers, or unexpected events (e.g. an object falls) could foster more sensible actions to be undertaken. We have shown in different experiments that the robot by simply interacting in such an environment and provided appropriate learning rules, could independently and autonomously learn the coordinate transformations and motor commands required for shifting its attention appropriately.

For biological systems we already know, at least partially, that these maps and controllers are present. What we do not know is exactly how the different signals interrelate, and how learning could be carried out appropriately. So we would like to stress the fact that for the first time we were able to condense, albeit in a simplified form, our knowledge onto a working physical setup, interacting in a real environment. This "learning" by doing approach might prove useful in understanding how and why some variables are necessary, and how the brain might be using such information.

Future work will include a more accurate exploitation of the frequency information, for example, it is known that the ILD at low frequencies provides information on the azimuthal angle, and thus can be integrated with the ITD for better performance. Further, temporal delays at stimulus onset or termination are valid cues for improving ITD estimation. The model will be also extended to take into account the eyes position for a large range of values (and not only for $|q| <$ threshold) thus removing some of the assumptions made in the proposed implementation. In this paper we focused the attention on the aspects concerning the learning of the motor response; future experiments will investigate the ability of the robot to keep an adequate motor performance even in presence of alterations in the sensory and motor subsystems.

Acknowledgements

The research described in this paper has been funded by the Italian Space Agency (ASI), and by the European Union grants MIRROR (IST-2000-28159) and CVS (IST-2000-29375).

References

[1] B. Adams, C. Breazeal, R.A. Brooks, B. Scassellati, Humanoid robots: A new kind of tool, IEEE Intelligent Systems 15 (4) (2000) 25–31.

[2] A. Bernardino, J. Santos-Victor, Binocular visual tracking: Integration of perception and control, IEEE Transactions on Robotics and Automation 15 (1999) 1080–1094.

[3] J. Blauert, Spatial Hearing: The Psychophysics of Human Sound Localization, MIT Press, Cambridge, MA, 1983.

[4] 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, The Journal of Neuroscience 13 (1993) 4589–4608.

[5] M.S. Brainard, E.I. Knudsen, S.D. Esterly, Neural derivation of sound source location: resolution of spatial ambiguities in binaural cues, The Journal of the Acoustical Society of America 91 (1992) 1015–1027.

[6] R.O. Duda, Estimating azimuth and elevation from the interaural intensity difference, Technical Report No. 4, Department of Electrical Engineering, San Jose State University, San Jose, CA, 1993.

[7] R.O. Duda, W. Chau, Combined monaural and binaural localization of sound sources, in: Proceedings of the 29th Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, 1995.

[8] B. Espiau, F. Chaumette, P. Rives, A new approach to visual servoing in robotics, IEEE Transactions on Robotics and Automation 8 (1992) 313–326.

[9] N. Henderson, Estimating azimuth from speech in a natural auditory environment, Technical Report No. 14, Department of Electrical Engineering, San Jose State University, San Jose, CA, 1996.

[10] P.M. Hofman, A.J. Van Opstal, Spectro-temporal factors in two-dimensional human sound localization, The Journal of the Acoustical Society of America 103 (1998) 2634–2648.

[11] P.M. Hofman, J.G.A. Van Riswick, A.J. Van Opstal, Relearning sound localization with new ears, Nature Neuroscience 1 (1998) 417–421.

[12] J. Huang, N. Ohnishi, N. Sugie, A biomimetic system for localization and separation of multiple sound sources, IEEE Transactions on Instrumentation and Measurement 44 (1995) 733–738.

[13] J. Huang, T. Supaongprapa, I. Terakura, F. Wang, N. Ohnishi, N. Sugie, A model-based sound localization system and its application to robot navigation, Robotics and Autonomous Systems 27 (1999) 199–209.

[14] Intel, Intel® Signal Processing Library 4.5, 2000.

[15] R.E. Irie, Robust sound localization: an application of an auditory perception system for a humanoid robot, Master's Thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, 1995.

[16] M. Kawato, K. Furukawa, R. Suzuki, A hierarchical neural network model for control and learning of voluntary movement, Biological Cybernetics 57 (1987) 169–185.

[17] C.H. Keller, T.H. Takahashi, Binaural cross-correlation predicts the response of neurons in the owl's auditory space under conditions simulating summing localization, The Journal of Neuroscience 16 (1996) 4300–4309.

[18] 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.

[19] E.I. Knudsen, The hearing of the barn owl, Scientific American 245 (1981) 82–91.

[20] E.I. Knudsen, P.K. Knudsen, Vision guides the adjustment of auditory localization in young barn owls, Science 230 (1985) 545–548.

[21] E.I. Knudsen, J. Mogdans, Vision-independent adjustment of unit tuning to sound localization cues in response to monaural occlusion in developing owl optic tectum, The Journal of Neuroscience 12 (1992) 3485–3493.

[22] M. Konishi, Listening with two ears, Scientific American 268 (1993) 66–73.

[23] M. Kuperstein, Neural model of adaptive hand-eye coordination for single postures, Science 239 (1988) 1308–1311.

[24] O. Manlov et al., Indoor mobile robot control for environment information gleaning, in: Proceedings of the IEEE Intelligent Vehicles Symposium, Dearborn, MI, 2000, pp. 602–607.

[25] M.A. Meredith, B.E. Stein, Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration, Journal of Neurophysiology 56 (1986) 640–662.

[26] G. Metta, Babyrobot: A study on sensori-motor development, Ph.D. Thesis, Dipartimento di Informatica, Sistemistica e Telematica, University of Genoa, Genoa, Italy, 1999.

[27] J.C. Middlebrooks, J.C. Makous, D.M. Green, Directional sensitivity of sound-pressure levels in the human ear canal, The Journal of the Acoustical Society of America 86 (1989) 89–108.

[28] H. Nakashima, N. Ohnishi, Learning sound source localization by interaction between motion and sensing in a nonlinear motor system, in: Proceedings of the Fifth International Conference on Neural Information Processing, Kitakyushu, Fukuoka, Japan, Vol. 2, IOS Press, Amsterdam, 1998, pp. 861–864.

[29] C. Neti, E.D. Young, M.H. Schneider, Neural network models of sound localization based on directional filtering by the pinnae, The Journal of the Acoustical Society of America 92 (1992) 3141–3155.

[30] F. Panerai, G. Metta, G. Sandini, Visuo-inertial stabilization in space-variant binocular systems, Robotics and Autonomous Systems 30 (2000) 195–214.

[31] D. Rosen, D.E. Rumelhart, E.I. Knudsen, A connectionist model of the owl's sound localization system, in: Proceedings of the Advances in Neural Information Processing Systems, Vol. 6, Morgan Kaufmann, San Mateo, CA, 1994, pp. 606–613.

[32] 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 (1999) 96–110.

[33] M. Rucci, J. Wray, Binaural cross-correlation and auditory localization in the barn owl: A theoretical study, Neural Networks 12 (1999) 31–42.

[34] G. Sandini, V. Tagliasco, An anthropomorphic retina-like structure for scene analysis, Computer Vision, Graphics and Image Processing 14 (1980) 365–372.

[35] V. Santos, J.G.M. Goncalves, F. Vaz, Perception maps for the local navigation of a mobile robot: A neural network approach, in: Proceedings of the IEEE International Conference on Robotics and Automation, San Diego, CA, 1994, pp. 2193–2198.

[36] P. Zakarouskas, M.S. Cynader, A computational theory of spectral cue localization, The Journal of the Acoustical Society of America 94 (1993) 1323–1331.

Lorenzo Natale was born in Genoa, Italy, in 1975. He received the M.S. degree in Electronic Engineering from the University of Genoa, Italy, in 2000. His thesis concerned the study of sound localization with the consequent development of a system learning to direct the gaze toward auditory targets on the basis of both visual and auditory cues. In January 2001 he has started Ph.D. in Robotics at the Laboratory for Integrated Advanced Robotics (LIRA-Lab, Genoa). His current research involves the study of motor control and sensori-motor coordination.

Giorgio Metta holds an M.S. degree and Ph.D. in Electronic Engineering from the University of Genoa, Italy. He worked on a humanoid robot from a biologically motivated perspective, with the ultimate goal of learning how to model biological agents by building complex artificial systems. Since 2001, he is Postdoctoral Associate at the MIT, AI-Lab. His current interests are man-machine interaction, imitation learning and the development of gesture/language in robotics.

Giulio Sandini teaches the course of Natural and Artificial Intelligent Systems for students of the Electronics and Computer Science curriculum. He currently coordinates the activity of researchers at the Laboratory for Integrated Advanced Robotics (LIRA-Lab, Genoa) where research related to robotics and computational neuroscience is carried out. Among the ongoing projects the control of a binocular head using space-variant, anthropomorphic sensors and the guidance of mobile robots on the basis of visual information. Giulio Sandini has been a member of program committees of international conferences and chairman and co-chairman of international conferences and workshops.