ABSTRACT. This chapter is an essay on the relationship between audition, action, and space. In its main features, our auditory system resembles that of other vertebrates: mammals, birds, reptiles, amphibians or fish, and this testifies to its great phylogenetic age. Perceptual systems were originally inseparable from motor systems, and this applies in particular to the sense of space that guides navigation and adjusts spatial parameters of action. Hearing is complementary to other senses, but with the advantage that it works in dark and cluttered environments. A closer look at the auditory system reveals a largely centrifugal structure, in which the most central levels can act on the most peripheral, and many species adopt active strategies to probe space acoustically.

1 Introduction

Perception and action are intimately linked in the life of an organism, but psychology and physiology commonly treat them as separate. Boring [1942] dates this divorce to the discovery by Bell and Magendie, at the dawn of the 19th century, of distinct motor and sensory pathways stemming from the spinal cord. Thanks in part to this reductionist focus, much progress has been made in understanding both pathways, but sometimes it is useful to take a less focused perspective and look again at perception within the context of action. This chapter attempts to do this for the auditory perception of space. It is divided into four parts. The first part recalls the “active” roots of perception, still apparent in primitive organisms. The second reviews active aspects of hearing, both in the structure and physiology of the system and in active perceptual strategies. The third focuses on the perception of space. The fourth asks, a contrario, how perception and action are distinct (in the system and in our inquiry).

2 Perception and action

To illustrate the link between perception and action, Szentgothai and Arbib [1975] used the example of a hypothetical organism equipped with two fins and two eyes. The right eye was connected to the left fin by a neuron, and the left eye to the right fin. When a prey appears within the field of the right eye, a command is sent to the left fin to instruct it to move (Fig. 1). The organism then turns towards the prey, and this orientation is maintained by bilateral activation until the prey is reached. Perception for this primitive organism is not distinct from action, and Szentgothai and Arbib suggest
For this primitive organism, perceiving of a prey is equivalent to sending a command to the opposite fin to move. Unilateral activation orients the organism towards its prey, and bilateral activation then ensures a trajectory that reaches it that the same may be true for more evolved organisms, except that multiple levels of inhibition are built over this basic design to allow a more sophisticated behavior.

A similar mechanism (without the crossing) would allow avoidance of a predator. To take an example from the acoustical domain, the ears of certain moths (Notodontidae) feed a neuron that causes the moth to drop if a bat approaches [Fullard, 1998; Pollack, 1998]. A very simple mechanism may thus be sufficient to produce a behavior that promotes survival.

Things become more complex if the environment contains both predators and prey. Depending on whether the organism meets one or the other, it must choose whether to approach the target or else move away. Together with a richer repertoire of behaviors, this requires the competence to discriminate between different kinds of objects. Some crickets thus react to two types of sound stimuli, one that attracts them (positive phonotaxy) and another that makes them flee (negative phonotaxy) [Pollack 1998].

The difficulty increases if the environment contains multiple objects: prey, predators or obstacles. In the presence of multiple objects the simple algorithm of Fig. 1 would fail, and the same might apply to a more sophisticated discriminative mechanism. A predator that meets a group of prey might fail to recognize them, or else might choose a median trajectory and miss them all. To address this situation an organism must perform perceptual segregation. The importance of this task was recognized only recently, with the development of Auditory Scene Analysis (ASA) by Bregman [1990]. It is likely that the sophisticated structure of our auditory system is in part the result of a selective pressure to perform this analysis efficiently [de Cheveigné 2001].

Let us return to the organism of Figure 1. The information at its disposal is crude (two "bits" of information: presence vs absence, left vs right) but it is enough to perform a complex task that results in a precise determination of the spatial position of the prey. Turning towards the prey determines its bearing, and moving to it determines (a posteriori) its distance. Through its action, the organism translates rudimentary spatial cues (left-right) into accurate spatial coordinates. However this mode of acquisition has a cost: the organism must travel to the object in order to localize it, and cannot use that information to choose whether it should go there, nor can later remember having
been there. Its action reveals it to prey and predators. In a more complex organism the link is looser, which allows it to create motor plans, construct a model of the world, and choose between several possible actions. The resulting dissociation between perception and action explains why we can justifiably abstract one from the other, in the tradition of physiological psychology handed down to us from Bell and Magendie [Boring 1942]. The nature of their relations is nonetheless important, in particular as far as hearing is concerned.

3 Active hearing

Plato and Euclid thought that the eye projects rays or particles towards objects to perceive them (extramission). Pythagoras thought the same of the ear [von Békésy and Rosenblith 1948]. The idea of an active and exploring ear has since been abandoned, but several aspects of hearing recall this notion: the largely efferent (descending, centrifugal) structure of the auditory nervous system, the intriguing phenomena of otoacoustic emissions, and of course the echolocation strategies of many species. The next paragraph dwells on some active and efferent aspects of the auditory system.

3.1 Anatomy and physiology

The auditory system is often described as having a bottom-up structure (afferent, centripetal). Acoustic waves from a source travel to the eardrum (through air or water) after having been reflected and refracted by the torso, head, pinna and auditory meatus. Displacements of the eardrum are transmitted by the chain of ossicles of the middle ear until the cochlea, causing the basilar membrane to vibrate. Displacements of the basilar membrane produce a shearing effect between the organ of Corti and the tectorial membrane, and this excites a population of sensory cells: the inner hair cells (IHC). These are, as it were, the "microphones" of the ear. Depolarizing a hair cell causes it to release neurotransmitters that trigger spikes within dendrites of spiral ganglion neurons that project, via the auditory nerve, to the brain. The pathway to the cortex crosses a series of relays: cochlear nucleus, nuclei of the superior olivary complex and of the lateral lemniscus, inferior colliculus, and medial geniculate body (MGB) [Popper and Fay 1992; Webster, Popper and Fay 1992; Ehret and Romand 1997]. Figure 2 privileges, among these ascending branches, those that allow exchange of information between hemispheres (and thus the binaural interactions that play a role in the perception of space).

In addition to these ascending pathways which play an obvious role there is a massive presence of descending pathways (Fig. 3) the purpose of which is more obscure. Why send information from center to periphery within a system who's role is to listen to the environment? In some cases the efferent pathways seem to involve mere reverberation between relays (e.g. cortex and thalamus) or a feedback loop. In others, their mechanism and role are still to be determined.

Prominent among these efferent pathways, there is one that terminates on the outer hair cells (OHC). After many years of speculation about the sensory role of outer hair cells cells (and the reasons for two distinct populations of hair cells), it finally became
clear that the role of OHCs is motor: they inject enough mechanical power into the cochlea to compensate for losses and avoid excessive damping. In normal situations their “operating point” is adjusted, under the control of the efferent system, just below the limit of spontaneous oscillation [Warr 1992; Guinan 1996; Sahey et al. 1997]. However that limit is sometimes exceeded and the ear then produces a sound called otoacoustic emission [Kemp 1978; Probst et al. 1990]. Puel et al. [1989] found that the emission can vary during a visual task, thus showing that the effects of efferent control extend to the periphery and beyond. Aside from the curiosity and wonder that such phenomena evoke, they betray the active mechanisms within the auditory system. The sound level produced is not enough (and probably never was throughout evolution) to “illuminate” an object acoustically, but the existence of otoacoustic emissions nevertheless shows that Pythagoras was right, metaphorically.

Fig. 2. Schematic representation of the main auditory relays and ascending pathways. The cochlear nucleus is subdivided into parts labeled DCN, PVCN and AVCN (dorsal, posteroventral and anteroventral), the superior olivary complex into parts labeled MSO, LSO and MNTB (medial and lateral superior olives and medial nucleus of the trapezoid body), and the nuclei of the lateral lemniscus into DNLL, INLL et VNLL (dorsal, intermediate and ventral). The inferior colliculus is labeled IC, the medial geniculate body MGB, and the auditory cortex AC [from Cheveigné 2003, 2006].

It has been suggested that efferent control adjusts the sensitivity of the ear to interaural intensity differences that cue spatial location (see below) [Sahey et al. 1997]. Outer hair cells, and the efferent system that controls them, are phylogenetically ancient since they are found in all vertebrates as well as some invertebrates [Warr, 1992; Guinan, 1996; Carr and Code, 2000; Sahey et al. 1997]. The muscles of the middle ear, also under efferent control, probably serve to attenuate loud and predictable sounds such as produced by vocalization (Du Verney [1636] thought that they tuned the resonant frequency of the eardrum). Certain birds have movable “covers” at the entrance of
their auditory meatus [Klump 2000] the role of which might be to adjust sensitivity to spatial cues. Finally, many animals possess movable pinnae that adjust the directivity of the ears. Those of certain bats move (one forward the other backward) at each pulse produced [Popper and Fay 1995].

3.2 Active strategies

Echolocation is used by bats and marine mammals [Popper and Fay 1995; Au et al. 2000]. Specialized production processes are involved to produce a high level of acoustic power and focus it in space and time, and specialized perception processes to receive and decode the echos. Signals themselves are tailored for this purpose, with a wide range of frequencies (up to 150 kHz) and various forms of modulation to exploit the delay of the echo (that indicates distance from the target), the Doppler effect (that indicates its speed, wing beats, etc.) and spectro-temporal statistics (that reveal its texture). Echolocation allows a creature to localize and characterize its prey, avoid obstacles, and probe the structure of a space (for example a cave).

Echolocation is also used by cave-dwelling birds [Klump, 1998]. Truax [1984] mentions the case of sailors that used whistles to navigate at the entrance of a harbor in
fog or darkness. The blind use the tapping of their stick, or mouth-produced clicks to characterize spaces and find their way. Reznikoff [1987] even put forward the hypothesis that prehistoric humans used their voice to navigate within caves, using particular resonances as landmarks. In contrast to “clicks” produced by the blind, that are useful to judge the proximity of nearby surfaces, the voice has controllable power, directivity and spectral properties adequate to excite resonance modes of remote cavities and volumes. It would be worth exploring this hypothesis experimentally.

The range of echolocation is limited by the double pathway to and from the target (especially in an absorbing or noisy environment), and by the weak and unreliable nature of the reflection. The range is greatly extended if, instead of an echo, the creature uses the response of another creature. Such an acoustic exchange is useful to keep close to a conspecific (partner, parent or offspring), or instead to avoid it or maintain it at a distance (rival) [Hyman 2003], or to locate oneself with respect to a herd or flock. A remarkable example is location by penguins of their offspring among thousands of others [Aubin and Jouventin 1998]. Selective pressure also affects the targets: certain flowers evolved with a catadioptric shape that makes them easier to spot by bats. Conversely, certain moths have developed acoustical responses to bat calls to discourage them, a mechanism that later may have evolved to allow communication between partners [Conner 1999]. Might this be the distant origin of language?... To summarize, active mechanisms are at work within the auditory system, and active strategies are employed for perception, in particular to probe space.

4 Spatial hearing

Hearing plays an important role in the perception of space. True, its angular resolution is much more crude than that of vision: 1 to 3 degrees for azimuth and 10 to 20 degrees for elevation in man [Blauert 1997; Heffner and Heffner 1992] as opposed to 5 seconds for vision [Barlow and Mollon 1989]. However hearing allows a faster reaction time, works in the darkness, and informs about sources located outside the field of vision. Low frequency sounds propagate around obstacles, and in some case cover long distances. Some media are thus more transparent to sound than to light.

Whereas retina offer “multichannel” visual information that is spatially rich, the auditory system makes do with only two samples of the acoustic field, extracted by the ears. Cues to location of a source are mainly the interaural intensity difference (IID) and interaural time difference (ITD) [Thompson, 1882; Blauert, 1997]. Sound from a source on one side of the head reaches the opposite ear with a greater delay (due to a longer acoustic path) and attenuation (due to the shadow effect of the head). For most animals including man [Heffner and Heffner, 1992], interaural differences in time and intensity jointly determine the azimuth (bearing) of a source. The former are most useful at low frequencies, the latter at high frequencies. Perception of elevation, less reliable, depends mainly on variations, as a function of elevation angle, of filtering properties of pinnae, head and torso. Perception of distance, even less reliable, depends on cues such as loudness (for sources of known power), ratio between low and high frequencies (since the former are less attenuated than the latter by propagation through air), ratio between direct and reverberated sound in a closed space, etc. Each of these cues requires some degree of familiarity with the source to be effective, and thus operates
via inference on the basis of multiple clues, only some of which are acoustical. We are justified to speak of a model of the scene that, once built, allows inferences that go beyond mere localization.

The barn owl (Tyto Alba) has asymmetric ears and uses interaural intensity differences to determine elevation rather than azimuth as in other species. It thus forms a precise bidimensional map within which the direction of a source can be located (this amounts, for a bird perched or hovering, to locating the position of a prey on the ground). Its exceptional auditory acuity is paired with excellent vision, and one can find in its superior colliculus a spatial map that is common to both modalities.

Processing of binaural cues by the auditory system is reminiscent of the organism of Fig. 1. Interaural intensity differences are measured in a brainstem relay, the lateral superior olive (LSO, Fig. 2). Measurement involves excitatory-inhibitory interaction between neural patterns coming from both ears. Interaural time differences are measured in another relay, the medial superior olive (MSO, Fig. 2), by excitatory-excitatory interaction. According to a classic model due to Jeffress [1948], patterns from one ear undergo an internal delay before converging with those from the other ear on a coincidence detector neuron. MSO contains a population of cells of this type, each with its own delay. The cell with an internal delay that compensates the external propagation delay is activated, and this indicates the bearing of the source. It is as if the organism had interiorized, within its MSO, a population of primitive organisms (Fig. 1), each attending to its own azimuth.

This discussion (and most of the psychoacoustics of localization) assumes that the source to be localized is isolated. However most acoustic scenes are populated by concurrent sources, and it is not easy to extend to them results obtained with isolated sources. Complex scenes raise the problem of segregating cues to each source superimposed within the acoustic signal sampled by the ears. Supposing that cues of concurrent sources can be teased apart, there remains the problem of the perceptual organization of the scene as a whole. Here again binaural cues are put to work. First for segregation: according to the equalization-cancellation (EC) model of Durlach [1963; Culling et al. 1998], patterns from one ear are adjusted in delay and amplitude to compensate for interaural disparities of a masking source, and then subtracted from the patterns from the other ear. This has the effect of attenuating correlates of the masker and favoring the extraction of the other sources. Next for perceptual organization: spatial separation between sources eases the perceptual aggregation of fragments belonging to the same source, and thus reduces informational masking of one source by another [Darwin and Hukin 1999, 2001a,b; de Cheveigné et al. 2002]. To summarize, hearing is an important modality for perceiving space, and spatial cues themselves are useful for the analysis of complex acoustic scenes.

5 Perception without action?

The creature of Fig. 1 locates its prey with accuracy and a remarkable economy of sophistication, but for that it must move. This costs it energy and time and reveals it to predators and prey. A mechanism that avoids displacement would give it a selective advantage. IID and ITD measurement circuits within the brainstem can be interpreted as such. Source direction is represented internally by activation of one neuron in MSO
The organism thus knows where the source is without having to go to it. Indeed, it can use this information to choose whether or not to go and, supposing it chooses to do so, to prepare its move. The requirements on stimulus duration are reduced. For example a sound of 75 ms is enough for the barn owl to direct its head towards a source with an accuracy of about 2° in both azimuth and elevation (within a field of 30°). This is much too short for a feedback loop.

The more sophisticated perceptual system thus buys economy, speed and efficacy. To take a counter-example, the zooplankton *Copelia quadrata* owns a unique eye that sweeps within the focal plane of the lens in the manner of a fax or television screen [Gregory et al. 1964]. A retina or compound eye would spare it the trouble. Decoupling perception from action allows information to accumulate to form abstract *models* or *motor plans* that guide future action. Indeed it has been suggested that such “potential motor plans” constitute the representation of visual space [Mattelli and Lupino 2001]. They are reminiscent of feedforward control mechanisms in the cerebellum, in which a predictive model guides a complex and rapid movement without the time and stability constraints of feedback control [Miall et al. 1993].

The evolution of perceptual systems can thus be seen as a trend towards economy of action. This does not eliminate the usefulness of action for a perceptual system. We turn our head to disambiguate azimuth or elevation, we move closer to the source or away from noise, we wander around the room to locate a ringing cell phone, we invent stethoscope or microphone to hear subtle sources, etc. It has been suggested that action is required to learn something as basic as the three dimensional structure of space [O’Regan and No 2001, Philipona et al. 2004, Aytekin et al. 2006]. Action is also necessary to *calibrate* perception: without action, how can we know if perception is veridical? And, last but not least, action for survival remains the ultimate goal of every organism.

## Conclusion

This chapter is an essay on the relations between audition, action, and space. In its main features, our auditory system resembles that of other vertebrates: mammals [Fay and Popper 1994; Au et al. 2000], birds or reptiles [Dooling et al. 2000], amphibians or fish [Fay et Popper 1999], which testifies to its phylogenetic age. It is likely that perceptual systems, hearing in particular, originally were inseparable from motor systems. This is probably most true of the sense of space, the main goal of which is to guide navigation and adjust spatial parameters of action. In this role, hearing is complementary of other senses, but with the advantage that it works in dark and cluttered environments. Hearing sometimes employs active strategies, and a closer look at the auditory system reveals a centrifugal structure in which central levels act on the most peripheral. The inclusion of active dimensions in the study of perception may promote our insight into perceptual phenomena.
Acknowledgments

This work was supported by the Cognitique programme of the French Ministry of Education, Research and Technology (MENRT). This paper is adapted from de Cheveigné [2006]. Thanks to Maria Chait for comments on the manuscript.

References


Keio University Press Inc. 2006