ARTICLE Communicated by J. Kevin O'Regan

A Sensorimotor Approach to Sound Localization

Murat Aytekin
aytekin@umd.edu
Neuroscience and Cognitive Science Program, University of Maryland, College Park, MD 20742, U.S.A.

Cynthia F. Moss
cmoss@psyc.umd.edu
Neuroscience and Cognitive Science Program, Department of Psychology and Institute of Systems Research, University of Maryland, College Park, MD 20742, U.S.A.

Jonathan Z. Simon
jzsimon@umd.edu
Neuroscience and Cognitive Science Program, Department of Electrical and Computer Engineering, Department of Biology, University of Maryland, College Park, MD 20742, U.S.A.

Sound localization is known to be a complex phenomenon, combining multisensory information processing, experience-dependent plasticity, and movement. Here we present a sensorimotor model that addresses the question of how an organism could learn to localize sound sources without any a priori neural representation of its head-related transfer function or prior experience with auditory spatial information. We demonstrate quantitatively that the experience of the sensory consequences of its voluntary motor actions allows an organism to learn the spatial location of any sound source. Using examples from humans and echolocating bats, our model shows that a naive organism can learn the auditory space based solely on acoustic inputs and their relation to motor states.

1 Introduction

It is largely accepted that the relative position of a sound source is determined by binaural acoustical cues, such as interaural level and time differences (ILD and ITD) and monaural spectral features, embedded in the acoustic signals received at the ears. Recent advances in our understanding of sound localization, however, suggest that sound localization is not purely an acoustic phenomenon, an inherent assumption of any solely cue-based model. Studies report that aside from the acoustic information received at the ears, input to other sensory modalities can also affect a

subject's estimate of sound position in space. Vision, for instance, can influence and guide calibration of sound localization (Knudsen & Knudsen, 1985, 1989). Proprioceptive senses, such as position and the motion of the head, as well as the perceived direction of gravitational forces, and gaze direction (Lewald & Ehrenstein, 1998; Goossens & van Opstal, 1999; Lewald, Dörrscheidt, & Ehrenstein, 2000; Lewald & Karnath, 2000; DiZio, Held, Lackner, Shinn-Cunningham, & Durlach, 2001; Getzmann, 2002; Prieur et al., 2005; Sparks, 2005), also play essential roles in sound localization.

There is strong evidence that normal development and maintenance of the ability to localize sound requires auditory experience (Knudsen, 1982; Wilmington, Gray, & Jahrsdoerfer, 1994; King, Parsons, & Moore, 2000), a finding also proposed for the visiomotor system (Held & Hein, 1963; Hein, Held, & Gower, 1970). The auditory system has the capability to adapt to changes that occur in the directional characteristics of the external ears during development (Moore & Irvine, 1979; Clifton, Gwiazda, Bauer, Clarkson, & Held, 1988) and in adulthood (Hofman, van Riswick, & van Opstal, 1998; van Wanrooij & van Opstal, 2005; Kacelnik, Nodal, Parsons, & King, 2006). Subjects can adapt to artificial changes to specific sound localization cues (Held, 1955; Loomis, Hebert, & Cicinelli, 1990; Javer & Schwarz, 1995) and changes introduced to the information content of acoustic input; for example, blind infants can learn to use a sonar device to reach objects in their environments (Bower, 1989). The notion that the auditory system has the capability to learn to localize sound suggests that it should not rely on an innate, preexisting representation of how to interpret spatial acoustic cues. This line of evidence further suggests that the auditory system is plastic enough to acquire the spatial information and perform the computations that are needed to determine it, through experience.

Here we demonstrate an approach that aims to provide a comprehensive computational scheme that incorporates the adaptive and multisensory nature of the computation of spatial acoustic information by the nervous system. The approach addresses the question of how a naive nervous system might develop the ability to compute the spatial locations of sound sources. By naiveness, it is implied that the nervous system is not yet specialized to process spatial information provided by the sensory inputs. This approach to the sound localization problem is complementary to, but fundamentally different from, those that rely solely on acoustic cues for sound localization. Most acoustic cue-based approaches are limited to specific computational strategies that use exclusively the head-related transfer functions (HRTFs), that is, the direction-specific acoustic filtering by the pinnae and the head (Colburn & Kulkarni, 2005). Standard modeling approaches rely on acoustic information alone, and so cannot explain the effects of motor actions and other sensory modalities on the computation. Moreover, unlike the proposed model, by taking the outside observer's point of view, they are not concerned with how the auditory system acquires the knowledge of the spatial coordinates to utilize these acoustic cues. We propose a

sensorimotor approach (Poincaré, 1929; O'Regan & Noë, 2001; Philipona, O'Regan, & Nadal, 2003; Philipona, O'Regan, Nadal, & Coenen, 2004) to the general problem of sound localization, with an emphasis on questions of development and learning that allow spatial information to be acquired and refined by a mobile agent. Thus, the sensorimotor approach does not require a priori representation of space by the system.

2 The Sensorimotor Approach

How can a naive animal acquire the sense of space, that is, how does the brain acquire enough information about its surroundings to localize a sound source? If experience is necessary for the development of spatial hearing, that is, sound localization, one might infer that this also involves learning to interpret spatial information embedded in the acoustic signals it receives. This would require learned associations between certain features of the acoustic input received at the ears and the spatial information carried by sounds. Although behavioral evidence suggests that a reflexive orientation to sound sources by neonatal animals and human infants is hardwired (Kelly & Potash, 1986; Muir, Clifton, & Clarkson, 1989), this reflex disappears and is later replaced by spatial cognition that develops through experience (Clifton, 1992; Campos et al., 2000). The purpose of this early orientation reflex, controlled by the lower brain stem, may then provide an initial state for this learning process (Metta, 2000; Muir & Hains, 2004). Without the knowledge about associations between acoustic cues and spatial information, a naive nervous system would be unable to interpret spatial characteristics of the acoustic inputs (Poincaré, 1929; Philipona et al., 2003).

What gives rise to auditory spatial perception? Poincaré (1929) and recently Fermuller and Aloimonos (1994) and O'Regan and Noe (2001) argue that if an organism has no means to generate movements and sense them (proprioception), the perception of space cannot develop. With voluntary movements, the nervous system learns sensorimotor contingencies (O'Regan & Noë, 2001), which in turn reveal the spatial properties of the acoustic inputs. To clarify this point, we adapt an example given by Poincaré for visual perception of space. Consider two sound sources located at a given position with reference to a listener's head at different times. Assume that both sounds are displaced 30 degrees to the right with reference to the listener. These relative displacements will induce different acoustic inputs at the subject's ears. Despite the differences in the acoustic sensations at these two instances in time, the nervous system will interpret them both as sound sources that underwent the same spatial displacement. There is something in common between the two sets of acoustic inputs that allows the brain to calculate the change in spatial position of the two sound sources. The acoustic changes associated with the 30 degree displacement of each sound source are different, but the argument is that these changes are interpreted

by the auditory system as comparable because they can be compensated by the same set of motor actions. By compensation, it is implied that sensory inputs are brought back to their initial conditions, for example, 30 degree head movement to the right to recover the initial positions of the sound sources. As a realization of Poincaré's insight, it follows that in order to perceive the space acoustically, self-generated movements are required. These movements sample acoustic space by changing the positions of the ears and, crucially, by conveying the corresponding proprioceptive sensation.

Acoustic inputs received at the ears vary with the dynamics of a sound source (external changes) and according to the motor state of the body (internal changes). External changes may be due to nonstationary acoustic properties or a spatial displacement of a sound source. External and internal changes can be distinguished by the nervous system, since only internal changes are also sensed proprioceptively. Note that spatial external changes can be mimicked by certain motor actions that move the ears and head rigidly (rotation and translation). However, this cannot be used to mimic external changes due to dynamic acoustic properties of the sound source. Stated another way, spatial-external changes can be compensated by a certain set of changes imposed on internal states by motor actions. It is this notion of compensation that leads to the commonality between the internal and external changes, which then give rise to the entity of space encoded by the nervous system (Poincaré, 1929; Vuillemin, 1972; Philipona et al., 2003). Thus, the compensation is a direct consequence of the organism-environment interaction. The brain can distinguish the body and its exterior and, using this dichotomy, can learn and explore space through observation of the body's actions and the resulting sensory consequences. Sensorimotor laws (contingencies) should be invariant under changes in the way the acoustic information is encoded by the nervous system and changes in the sensor structure, provided that these changes do not hinder the notion of compensation (Philipona et al., 2003, 2004). In other words, spatial displacements of sound sources can be compensated by the same physical displacements of the sensors, regardless of the shape of the ear, head, or body or the way the information is represented in the central nervous system. Changes in the sound-source properties and of body state can be thought of as transformations acting on the acoustic environment-body system. Explorations of those transformations that lead to compensable actions yield the mathematical foundation that is necessary to formalize a mechanistic way to explore the acoustic space.

3 Sound Localization and Sensorimotor Approach

We envision the problem of learning auditory space as the organism's acquisition of the spatial relations between the points that can be occupied by sound sources in space. Our goal here is to show that using movements of the listener, that is, compensable motor actions, it is possible to identify

points in space and represent them with coordinates, and hence their geometric relations in the auditory space. The learning of the auditory space could then allow the system to determine acoustic or audio-motor features associated with a particular point in space, that is, sound localization. In order to achieve this task, we make three assumptions. First, we assume that an organism can distinguish the differences between the exteroceptive and proprioceptive inputs. Based on the classification and separability of sensory inputs in these two groups, an organism can identify the dimension of space through its interaction with the environment (Poincaré, 2001; Philipona et al., 2003). Thus, we also assume that the dimension of auditory space is known. Since only two parameters are necessary to identify a sound-source location in auditory space in the far field, auditory space is assumed to be two-dimensional. Finally, we assume that the organism can distinguish the motor actions that can induce spatial displacement (spatial motor actions). Philipona et al. (2003) demonstrate how a naive system can identify these special movements.

Our initial method will be limited to the learning of points and their spatial neighborhood relations in auditory space by the organism, without concern for the detailed metric properties. This approach is substantially simpler in its goals (and weaker in accomplishments) than that applied by Philipona et al. (2004) in the visual system. For our purposes, we need a way to identify points in the space from the organism's perspective.

3.1 Manifold of Spatial Acoustic Changes. One can view the set of all possible states of a sound source, its time, frequency and spatial properties as a manifold, E, a topological space that is locally diffeomorphic to a linear space. Similarly, all the motor states (including, e.g., head position) can be thought of as elements of a manifold, M, and all the acoustic sensory inputs received at the ears also constitute a manifold, S. Since the acoustic sensory input is determined by both the sound source and the current motor state, there is a functional relationship between the manifolds E, M, and S:

$$S = f(E, M).$$

If an organism makes all the possible head movements when the state of the sound sources is e_o, the resulting set of sensory inputs, $f(e_o, U)$, will be an embedding, where U represents a subset of motor states in M. For sound sources in the far field, these embeddings will be two-dimensional, since the relative position change of a sound source caused by head movements can be represented with two parameters. These embeddings can be referred as the orbits of a sound-source state. Sensory inputs corresponding to the different sound sources with the same relative spatial positions will lie on different orbits: two sound sources with different spectra will result in different sets of acoustic inputs during similar relative head motions. The orbits are smooth surfaces, $f(e_i, U)$ and $f(e_j, U)$, where e_i and e_j

represent two different sound sources, and consist of sets of sensory inputs that correspond to the environment-organism states. There can be as many orbits as there are different sound sources. As discussed earlier, the entity of space emerges as a result of observations that identical displacements of any two sound sources can be compensated by the same motor outputs. Exteroceptive and proprioceptive sensory inputs associated with the same displacement, however, could be different for different motor states in general.

Exteroceptive sensory changes that are compensated for by a particular motor action, such as a head rotation, may not be unique to particular spatial locations and so alone are insufficient to give rise to the concept of a point in space. Yet listeners, when they localize the spatial position of a sound source, do perceive it as a unique spatial point. Thus, there must be something in common (an invariant property of a spatial point) among these different sensory inputs so that the organism can identify their spatial parameters as characterizing a single point. From the organism's point of view, the locations in space of two potential sound sources' positions would be identical if both could be brought to a reference sensorimotor state with the same set of motor actions, for example, head movements. The set of potentially identifiable points (by movements of a head with immobile ears) results in a two-dimensional representation of space, since there are only two independent parameters, azimuth and elevation, available to identify the direction of a sound source in space with reference to the body.

If there were only a single orbit in the sensory input space (only one kind of sound), the solution to the auditory space learning problem would have been described as fitting a global coordinate system to this surface with two parameters. In this case, each sound-source location could have been represented by the global parameters of its corresponding point on the orbit, and any arbitrary point on this surface could have been picked as the reference sensory input or the origin of the global coordinate system.

The practical case of multiple sound sources (i.e., multiple orbits), however, poses a more challenging problem in terms of finding the parameters that represent the spatial positions of the sound sources. We can determine each orbit and fit a global coordinate system to each one of them, but as a result, we would obtain as many global coordinate systems as the number of orbits in the sensory input space. A critical task, then, is for each spatial source position to find the points on the orbits that correspond to that same spatial position and assign a single parametric representation. In general, there might not be additional information available to register these separate global coordinate systems with reference to each other. An extra sensory input, such as vision, favoring a particular set of organism-environment states (W, m_o) such that W represent the subset of sound-source states with the same position (0 degree azimuth and 0 degree elevation, for instance), could be sufficient as a reference. We argue, however, that under circumstances special to the auditory system, such an extra input is not necessary.

In fact, we show that the directional acoustic properties of the external ear make it possible to solve the localization problem without the need for another exteroceptive sense.

3.1.1 Directional Properties of the Acoustic Inputs at the Ears. Directional features of the acoustic signal received at the ears can be completely captured by linear-time-invariant (LTI) system models of the transformation of the sound caused by its interaction with the head and the pinnae. The transfer functions of these LTI models are the HRTFs (Blauert, 1997). Acoustic inputs received at the ears caused by a sound source in space can be mathematically depicted as follows:

$$S_{left}(f) = A(f) \cdot H_{left}(f, \theta, \phi)$$
$$S_{right}(f) = A(f) \cdot H_{right}(f, \theta, \phi).$$

The H_is represent the HRTF of left and the right ears for a given source at position, θ azimuth and ϕ elevation, respectively, for frequency f. $A(f)$ is the sound source's frequency representation (spectrum), and the $S_i(f)$ results are the spectra of the acoustic inputs as measured at the left and the right ears. Now we will show that the orbits that can be obtained from such a sensory system have common features (invariants) for the sensory inputs that are generated by the sound sources with the same relative positions. The existence of such features allows a definition for a point in space.

3.2 Computation of the Space Coordinate System. Head movements induce changes to the acoustic inputs produced by a sound source located at a fixed point in space. These changes can be thought of as transformations acting on the points on the orbit of this sound source. For simplicity, we assume that head movements are fast enough that during motion, there is no change in the state of the acoustic environment, and we assume that the head motion occurs only when there is a single sound source. Without loss of generalization (though see section 6), we may limit the organism's spatial motor actions to infinitesimal head movements around a fixed head orientation (0 degree azimuth and 0 degree elevation, for instance) when a measurement from the environment is taken. Each measurement is assumed independent of all others. Each head movement induces a vector tangent to a point, $f(e_i, m_o)$, on the sensory input manifold, S, representing the change in the sensory input. Each tangent vector is attached to the observed sensory input before the head movement.

In order for the nervous system to represent spatial points independent of their sound source, based on audio-motor contingencies, there must be a unique (i.e., sound source invariant), feature of spatial displacements that is present for all sets of exteroceptive sensory input changes associated with

particular motor state change. Note that each member of such a set arises from one point on a sound-source orbit. Only by utilizing source-invariant features is it possible to bring points on different orbits associated with the same relative spatial position into a single representation of that point in space.

How this feature manifests itself depends on the neural representation of the acoustic inputs. For instance, an organism that encodes changes in acoustic energy in logarithmic units (dB) could realize that some changes in the sensory inputs associated with its head movements are identical. In the case of acoustic energy differences encoded linearly in amplitude, a head movement would generate external changes that are linear functions of the acoustic inputs. These linear relationships would be constant for sound sources at the same relative position (e.g., $\Delta S = S(f)\frac{\partial H(\theta,\phi,f)}{\partial \theta}$), which would qualify as the necessary sound-source-invariant feature. Thus, different coding schemes require different computational solutions for the system to identify the invariance property, some of which might be easier to implement by the nervous system. For each coding scheme, an appropriate metric on the neural signals is needed to determine similarity of the neural representations of any two acoustic signals. In pure logarithmic coding, this metric takes its simplest form, in terms of the level of computation required, since similar inputs result in similar changes of the neural signals. But the specific neural coding of acoustic sensory input employed is not critical: it can be shown that under any sufficiently complete neural coding scheme, a diffeomorphism exists between the submanifolds of E and S.

In order to obtain a full coordinate representation of auditory space, no two sound-source-invariant features can be identical. Failure of this uniqueness would naturally result in (well-established) psychoacoustical spatial ambiguities.

3.2.1 From Sound-Source-Invariant Features to Coordinate Representation of Spatial Locations. Consider the sensory input changes caused by a set of infinitesimal head movements obtained for a given sound-source location. If we assume logarithmic coding for simplicity, the tangent vectors obtained by the same head movements for different sound sources will be identical, (i.e., a sound-source-invariant feature) and will change smoothly with sound-source location. If we further assume that the set of tangent vectors from a spatial location is unique to that location, then the set of tangent vector sets constitutes a two-dimensional manifold: the manifold of spatial points. Since this manifold is independent of the sound source's internal characteristics and dependent only on the relative position of the sound sources, one may then attempt to assign to it a global coordinate system, corresponding to the relative sound locations. The manifold can be represented as a two-dimensional embedding in an N-dimensional space, where N is the number of frequencies discriminable by the auditory system

and a manifold learning algorithm can be employed to obtain a coordinate system.

4 Demonstration 1: Obtaining Spatial Coordinates in Humans

We implement the proposed computational scheme for simplified human subjects. We show that by using voluntary head movements and observing their sensory consequences, it is possible to compute the (two-dimensional) global parameters that represent the direction of the far field sound sources in space. We use 45 human HRTFs from the publicly available CIPIC database (Algazi, Duda, Morrison, & Thompson, 2001), in magnitude only (i.e., no ITD information is used). The simulated humans take measurements from the environment by making small (1 degree) head movements from a resting state where the nose points at 0 degree azimuth and 0 degree elevation. For each measurement, the head makes the following three head movements; rightward, upward, and tilting downward to the right. Each head movement results in a vector that is on the tangent space at $f(e_i, m_o)$, the sensory input caused by a particular sound source. Sensory inputs are assumed to be represented in logarithmic (dB) coding. An extended-tangent vector is then produced for each measured sound source by concatenating the three measured tangent vectors. Since extended-tangent vectors are sound-source invariant, the set of them qualifies as a location-dependent feature, and the set of location-dependent features constitutes a manifold isomorphic to auditory space. The particular choice of head movements is not important as long as they generate a set of independent vector fields on the sensory input space. Any arbitrary vector field caused by an arbitrary head movement can be written as the linear combination of any independent vector field set. The number of independent vectors is limited by the dimension of the manifold. We have chosen three head movements to guarantee that at each spatial point, there are at least two independent tangent vectors that can be obtained. For example, at the north pole, right-to-left turns of the head cannot introduce any external changes, but the other two head movements can still generate two independent tangent vectors.

4.1 Simulation of the Sensory Inputs. Simulated sound sources are represented as vectors that comprise the Fourier transform magnitudes of sounds in the frequency interval of 1 kHz to 20 kHz, in steps of 200 Hz, resulting in 95-dimensional real vectors. These vectors are generated from a uniform random process and have units of dB. The HRTF phase information, normally available to humans up to 3 kHz (Stevens & Newman, 1936), is ignored only to simplify the analysis.

We have selected our source positions on the hemisphere that extends between two polar points located at -45 degrees elevation, 0 degree azimuth and 45 degrees elevation, 180 degrees azimuth (elevations below -45 degrees are not available for CIPIC data). Each sound spectrum is filtered

with their corresponding HRTF to compute the acoustic signals at the ear canal (represented as 190-dimensional vectors). The HRTFs are smoothed spectrally using a gaussian smoothing filter with a constant quality factor ($Q = 24$) and interpolated spatially using spherical harmonics (Evans, Angus, & Tew, 1998). The spatial positions of the sound sources are chosen using uniform spiral sampling (Saff & Kuijlaars, 1997) with the spiral axis extending from one ear to the other (total of 1748 points). This sampling approaches a uniform distribution on the sphere when the number of points approaches infinity. The tangent vectors are computed as the difference of the spectra of the simulated sound at the ears before and after the three head movements. Later these vectors are concatenated to build a 570-dimensional extended-tangent vector. Each extended-tangent vector is used to represent a spatial point in the space since they are independent of the sound-source spectrum under the logarithmic coding. If the extended vectors are unique for each sound-source location in (far-field) space, then they should lie on a two-dimensional embedding (manifold) in 570-dimensional space.

4.2 Determining the Spatial Parameters. Assuming the manifold of extended-tangent vectors exists, we can use a manifold learning algorithm, for example, adaptive-LTSA (local tangent space alignment), to assign global parameters to its points. There is a one-to-one association between these parameters and the sound-source locations. Adaptive LTSA is a nonlinear manifold learning algorithm with three steps. First, using an adaptive nearest-neighbors method (Wang, Zhang, & Zha, 2005), the algorithm finds the local neighborhood around each point. Then for every point, local coordinates of its neighboring points on the local tangent space are computed. In the third step, these local coordinates are aligned to construct a low-dimensional global representation of the points on the manifold.

The algorithm can capture local isometry only if the manifold is locally isometric to a Euclidean parameter space, but a hemisphere is not: one cannot flatten a hemisphere to a plane without distorting the pairwise distances between the points. Thus, if the algorithm can successfully capture the topological relations, the pairwise distances between any two points in the maps will be distorted. In addition, like other well-known spectral manifold learning methods—for example, locally linear embedding (Roweis & Saul, 2000) or Laplacian eigenmaps (Belkin & Niyogi, 2003)—LTSA (Saul, Weinberger, Ham, Sha, & Lee, 2006) cannot guarantee the recovery of the global parameters obtained from the alignment step (Zha & Zhang, 2005). To remedy the recovery problem, we implement the conformal component analysis (CCA) method proposed by Sha and Saul (2005) on the results obtained from LTSA. CCA computes the low-dimensional embedding by trying to preserve the angles in between the extended-tangent vector as a result of minimizing total local dissimilarity, and hence generates a conformal map of the points on the manifold. CCA can also be used to predict intrinsic dimension of the locally isometric manifolds. Local dissimilarity,

$D_i(s_i)$, is a measure based on the similarity (similarity of the corresponding angles) of the triangles in the neighborhood of an extended-tangent vector, $\{x_i\}$, and its image, $\{z_i\}$, in the low-dimensional embedding obtained via manifold learning methods:

$$D_i(s_i) = \sum_{jj'} \eta_{ij}\eta_{ij'}(\|z_j - z_{j'}\|^2 - s_i\|x_j - x_{j'}\|^2)^2.$$

Here x_i and z_i represent the extended-tangent vectors and their low-dimensional images, respectively. $\eta_{ij} = 1$ if x_j is a member of x_i's neighborhood and $\eta_{ij} = 0$ otherwise. s_i is the constant representing the scaling as a result of the conformal mapping at the neighborhood of x_i. The total local dissimilarity is obtained as $\sum_i D_i\,(s_i)$. This measure allows one to quantify and compare the two-dimensional embeddings obtained from this method using a local dissimilarity measure (Sha & Saul, 2005). Although this manifold learning method is not necessarily the best suited for obtaining the metric structure on the manifold, embeddings obtained from LTSA are sufficient to show that the neighborhood relationships of spatial points can be learned using auditory-motor contingencies.

In Figure 1 total local dissimilarity values are given for each subject. Forty-three out of 45 subjects gave similar dissimilarity values. The manifold learning method failed to give reliable low-dimensional embeddings for the remaining two subjects (subject 008 and subject 126) with the highest values of total local dissimilarity. Examples of global parameter maps are also provided in Figure 1. These maps correspond to subjects with 15th (subject 154), 30th (subject 050), and 43rd (subject 131) highest dissimilarity values. The stereographic projection of the spatial distribution of the sampled azimuth and elevation values can be seen in the insets on the left corner of the figure.

An ideal map of the global parameters of the sound-source locations should maintain the pairwise neighborhood relationship between the positions of the sound sources, hence showing a smooth change in relation to their spatial parameters. Notice that the relationship between the sound-source directions in space is preserved in the global parameter maps. For many subjects, the two-dimensional extended-tangent manifold requires three dimensions from the CCA results to globally describe the parameter space. This is expected since the manifolds studied here are not locally isometric. In these cases, the points in the lower-dimensional representation of the manifold manifest as a two-dimensional surface embedded in three-dimensional coordinate system. For the global parameter maps given in Figure 1, a two-dimensional projection was possible. We found that the density of the points in the global parameter maps increases at and around the north pole for all subjects. With increasing dissimilarity, value deformations are observed in the same region (subjects 050 and 131). The overall shapes

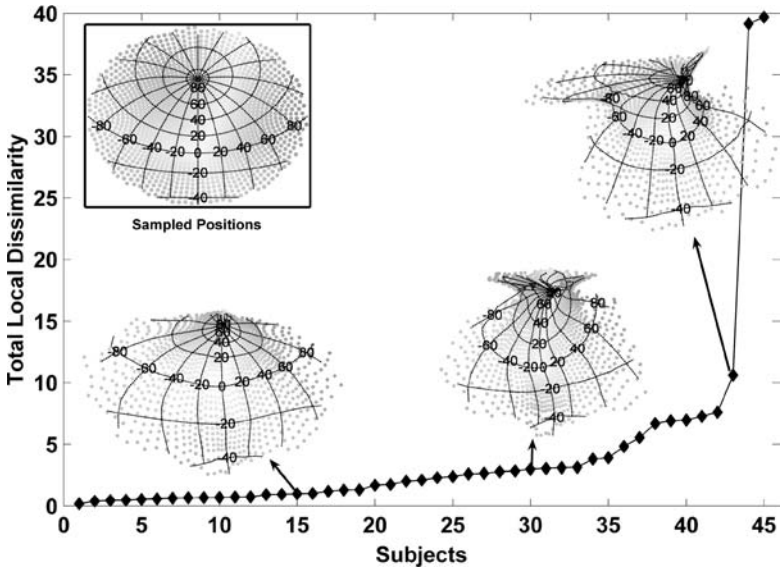


Figure 1: Total local dissimilarity for each subject. Spatial points are sampled from the hemisphere of interest in stereographic projection depicted in the inset. The coordinate grids indicating azimuth and elevation values of the sound-source locations allow comparisons with the global parameter maps obtained from the manifold learning method. Global maps of three subjects corresponding to 15th (subject 154), 30th (subject 050), and 43rd (subject 131) largest total local dissimilarity values are also shown (arrows). The manifold learning step failed to capture geometric organization of the spatial points for the subjects with two outlying local dissimilarity values: the 44th (subject 008) and 45th (subject 126).

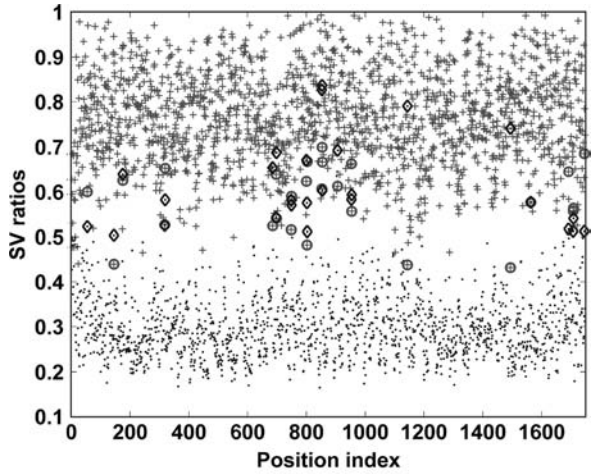
of the global parameter maps, however, are not particularly important as long as a unique pair of parameters is obtained for each sound-source direction (though it is interesting that the two dimensions correspond roughly to the directions of elevation and azimuth). Smoothness of the global parameters provides the ability to interpolate (i.e., to predict any inexperienced sound-source direction on this hemisphere).

Deformations are always observed at high elevations (more than 60 degrees). We have investigated potential reasons for these effects. Accurate determination of the local tangent space is dependent on the local curvature, local sampling density, noise level, and regularity of the Jacobian matrix (Zhang & Zha, 2004). The estimated pairwise distances between any two neighbor points are less accurate in the high curvature regions, and thus, uniform sampling of the manifold can introduce a bias in the alignment of the local coordinate system when a global minimization scheme

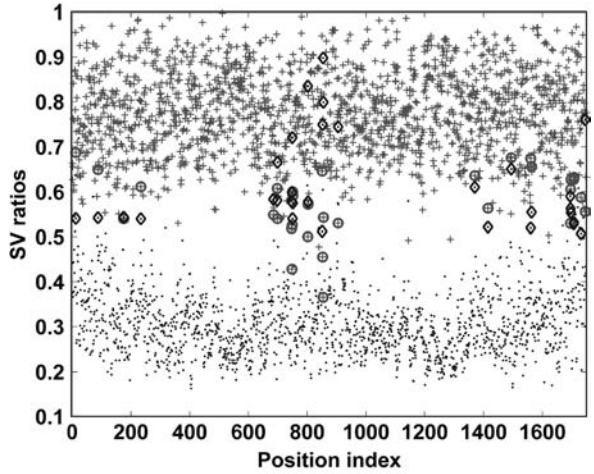
is employed. The process of global alignment of the local coordinates uses a weighting procedure inversely proportional to the local curvature estimates, which minimizes any potential bias that could be introduced by the high curvatures on the manifold (Wang et al., 2005). The local curvature values estimated by LTSA, as well as the condition of the Jacobian matrix, were examined in detail.

Spatial distribution of the local curvatures showed higher curvature values at and around the north pole, coinciding with the previously mentioned high-density regions. Mean curvature value above 60 degrees elevation across all subjects is found to be 1.07 ± 0.27 and 0.51 ± 0.04 below this elevation. Although these curvature estimations are only approximations, it is clear that at higher elevations, extended-tangent vector manifolds show higher local curvature values.

The Jacobian matrix represents the best linear approximation to the local tangent space at a given extended-tangent vector. Since the manifold of the extended-tangent vectors should be two-dimensional, the rank of the Jacobian matrix should be equal to two. Nonuniformity of the local tangent space dimensions on the manifold can result in nonoptimum solutions by the LTSA (Zhang & Zha, 2004). In order to investigate the condition of the Jacobian matrix, we compared the singular values of the matrices comprising the local coordinates of the points in each neighborhood obtained in the first step of the manifold learning algorithm. These matrices and the corresponding Jacobian matrix should span the same local linear subspace and thus have the same rank. Singular values of a (noisy) local coordinate matrix are expected to have two large values and $K - 2$ smaller values for a two-dimensional local linear space. Here K represents the number of points in the neighborhood. We investigated singular values obtained at each point's neighborhood for subjects with the smallest (subject 147) and largest (subject 131) dissimilarity values (see Figure 2). In Figures 2a and 2b we have shown the ratios of the singular values $\frac{\sigma_2}{\sigma_1}$ and $\frac{\sigma_3}{\sigma_2}$, depicted as plus signs and filled circles, for each neighborhood, where σ_i is the ith largest singular value ($\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_K$). For both subjects, the majority of the points show well-separated singular value ratios, confirming that first two singular values, σ_1 and σ_2, are comparable and larger than the rest of the singular values. However, this separation for subject 131 is not as robust as it is in subject 147, particularly for the points corresponding to the most peripheral left and right sound-source locations, as well as points near the north pole. Closer inspection reveals that σ_1 is moderately larger ($\frac{\sigma_2}{\sigma_1} < 0.7$, $\frac{\sigma_3}{\sigma_2} > 0.5$) than the rest of the singular values, suggesting that the dimension of the local linear spaces is less than 2. These points are highlighted in Figures 2a and 2b by marking corresponding values by open circles and open diamonds, respectively. Based on these observations, deformations observed in the global coordinate maps could be caused by the irregularity of the local tangent space dimensions in the extended-tangent vector set.



(a)



(b)

Figure 2: Singular value ratios $\frac{\sigma_2}{\sigma_1}$ and $\frac{\sigma_3}{\sigma_2}$, where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_K$. (a) Subject 147 (lowest-ranked total local dissimilarity) and (b) subject 131 (43rd ranked total local dissimilarity). Ratios obtained within the neighborhood of each point (total 1748 points uniformly distributed on the hemisphere) are depicted in '+' and '.', respectively. Moderate separation of the two sets of values implies full-rank Jacobian matrix for two-dimensional local tangent space. The ratios with potential rank deficiency problem are stressed with $\circ$ (for $\frac{\sigma_2}{\sigma_1}$) and $\diamond$ (for $\frac{\sigma_3}{\sigma_2}$). A local neighborhood is determined as problematic if the largest eigenvalue is moderately larger than the rest of them (if $\frac{\sigma_2}{\sigma_1} < 0.7$, $\frac{\sigma_3}{\sigma_2} > 0.5$).

In addition to the condition of the Jacobian matrix and local curvatures, local geodesic distances can also be influential on the global parameter maps. Because the global maps are only conformal representations, the local distances between the global parameters are not directly interpretable. Thus, we cannot directly conclude that points corresponding to the higher-density region are simply closer to each other (which would imply greater difficulty in distinguishing source locations near to each other). Assuming that the uniform spatial sampling is dense enough, however, it is possible to compare the local distances of the extended-tangent vectors and determine if this region contains points that are relatively closer to each other (i.e., more difficult to distinguish neighboring source locations). Spatial distributions of the mean local distances of each point to their local neighbors (neighborhoods are obtained using K-nearest neighborhood method where $K = 8$) are given in Figures 3a and 3b. Mean local distances for both subjects show a steady decrease with increasing elevation, confirming that at poles, the extended-tangent vectors are relatively close to each other. In Figures 3c and 3d, we have given the mean local distance maps of the underlying HRTF for each of the subjects. Notice that for both subjects, local HRTF distances are low above 40 degrees elevation. Minimum local distance values are obtained at 90 degrees elevation. In order to determine the robustness of these results, with increased sampling density we doubled the number of points and repeated the analysis and obtained quantitatively similar results.

Based on these results, we conclude that the high-density region in the global coordinate maps for all subjects have not only higher local curvature values but also consist of extended-tangent vectors that are similar to each other. The similarity of the HRTFs corresponding to the same spatial positions suggests that acoustic inputs vary less per degree in these regions. These results predict that subjects should experience more difficulty resolving sound-source locations in these regions (depending on the level of the minimum detectable change in the acoustic input level at different frequencies). Notice that similarity of the HRTF is based on a metric defined on the neural representations of the acoustic inputs, which is a result of ear shape and neural coding and not related to the metric associated with the acoustic space.

5 Demonstration 2: Obtaining Spatial Coordinates by Echolocating Bats

We also test our computational scheme with a different species, an echolocating bat *(Eptesicus fuscus)*, to stress that the sensorimotor approach provides a general approach that can capture spatial relations between points in auditory space for other animals that use a very different frequency range for auditory localization. Echolocating bats produce ultrasound vocalizations and listen to the echoes reflected from the objects around them to monitor their environment. Bats rely on active spatial hearing (biosonar) to

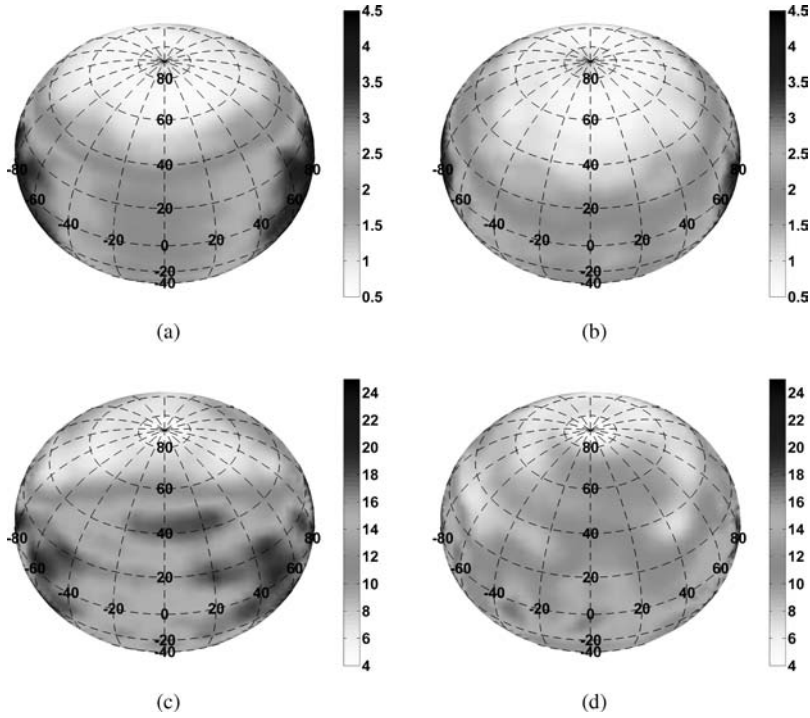


Figure 3: (a, b) Mean local distances of learned extended-tangent vectors. (c, d) Mean local distances of underlying head-related transfer functions (HRTFs). Subject 147 in a and c; subject 131 in b and d. Mean distances are determined within each local neighborhood of uniformly distributed 1748 spatial positions on the hemisphere (a and b) with K-nearest neighborhood criteria ($K = 8$). The local mean distance for each subject decreases with elevation and reaches its minimum value around the north pole. This property was common across all the subjects. Subject 147 (lowest total local dissimilarity) shows larger local distances below 60 degree elevation in comparison to subject 131 (moderately high total local dissimilarity) for both types of local distances.

detect, track, and hunt insects (Griffin, 1958). Thus, accurate sound localization is very important for these animals' survival.

The HRTFs of three bats were measured in the frontal hemisphere (Aytekin et al., 2004). Each HTRF spans 180 degrees azimuth and 180 degrees elevation, centered at 40 degrees, 30 degrees and 22 degrees elevation for subjects EF1, EF2, and EF3, respectively. The HRTFs were measured from 20 kHz to 100 kHz, in 68 logarithmic steps, and smoothed with a rectangular window with a constant quality factor of 20. Phase information (i.e., ITD information) was discarded, and magnitude is represented in dB units.

As in the human simulations, the sound-source locations were selected using uniform-spiral sampling. The simulated bat performed the same head movements as the human subjects (in particular, without pinna movement) to generate the extended-tangent vectors. Using the adaptive-LTSA algorithm, the global parameters for sound-source directions are obtained. In Figure 4 global parameters of two echolocating bats (subjects EF2, and EF3) are shown. In the inset of each figure, azimuth and elevation coordinates of these sound-source locations are given. For all three global coordinate maps, the local relationships between the sound-source directions in the head reference frame are preserved, showing that the sensorimotor approach can be applied successfully to echolocating bats.

The bat global parameter maps (see Figures 4a and 4b)—also show similar characteristics to the ones we obtained for humans. Sound positions corresponding to high elevations show a denser distribution than the midelevation regions. Similar to the human, local curvature values for bats are also higher at and around the north pole. Mean local curvatures below 60 degrees elevation were 0.40 ± 0.13, for subject EF1, 0.42 ± 0.19 for subject EF2, and 0.40 ± 0.10 for subject EF3 and above 60 degrees elevation were 0.59 ± 0.16, 0.82 ± 0.37, and 0.60 ± 0.27, respectively. All of these values were lower than those of the human subjects.

Figures 5a and 5b show the spatial distribution of the local distances of both the extended-tangent vectors and Figures 5c and 5d the underlying HRTFs. For both sets of local distance distributions, the spatial areas corresponding to the positions at and around the north pole give the minimum local distances. Notice also that larger distances are observed at the midelevation ranges, where the acoustic axes of the HRTFs are observed (Aytekin et al., 2004). Sound locations that give larger local distances generally correspond to the HRTF regions with the largest gain. The direction and frequency-dependent nature of the acoustic axes contribute to the larger distance values in the mid elevation region. This effect is especially clear for subject EF2's HRTF local distance distribution (see Figure 5d).

Based on these observations, we conclude that denser distributions for global parameters should be expected at high elevations. This effect might result in lower perceptual resolution of sound-source locations for the subjects at these sound-source locations.

6 Discussion

We have demonstrated that the acquisition of the ability to localize sounds can be achieved based on sensorimotor interactions. Unlike standard sound localization models, our model provides an unsupervised computational process, based on the sensorimotor experience of a mobile listener. The viability of the model is shown by simulations using HRTFs of the human and the echolocating bat. The spatial neighborhood relationships between

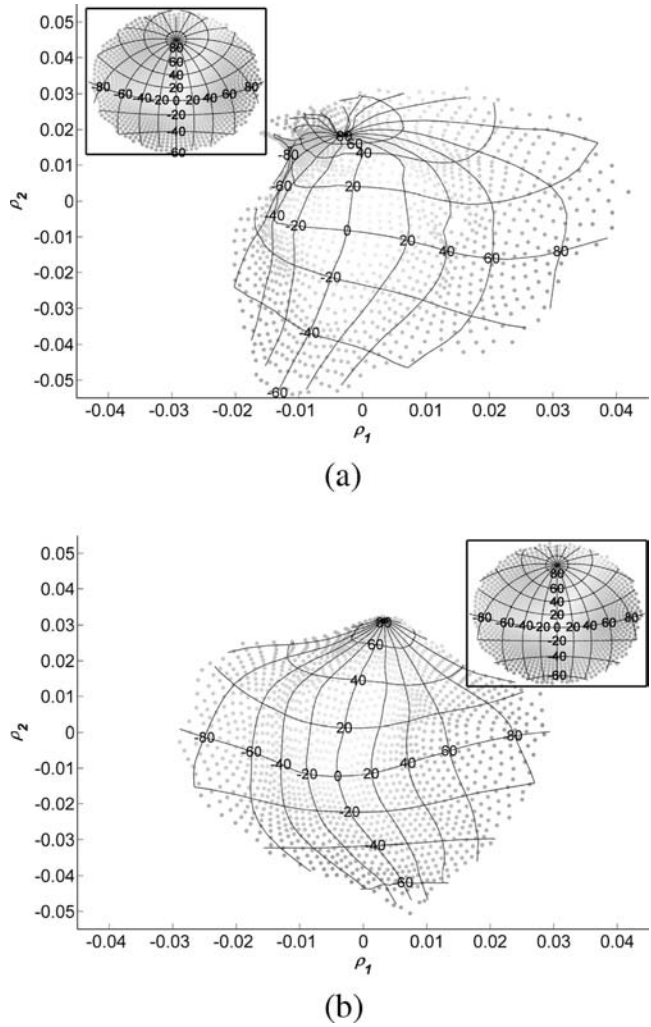


Figure 4: Global coordinates of two echolocating bats: (a) subject EF2 and (b) subject EF3. Similar to human subjects, global maps show increased density near the north pole. Sampled spatial positions (1707 points uniformly sampled on the hemisphere) are given in the insets of each figure. Global parameters obtained from both subjects preserved the topology of the sound-source locations given in the insets of each figure.

sound-source locations were successfully learned for both species' HRTFs. We found that the learned global parameters used to identify sound location reflect important features of the HRTF, for example, relative indistinguishability of the HRTF at high elevations resulted in smaller local

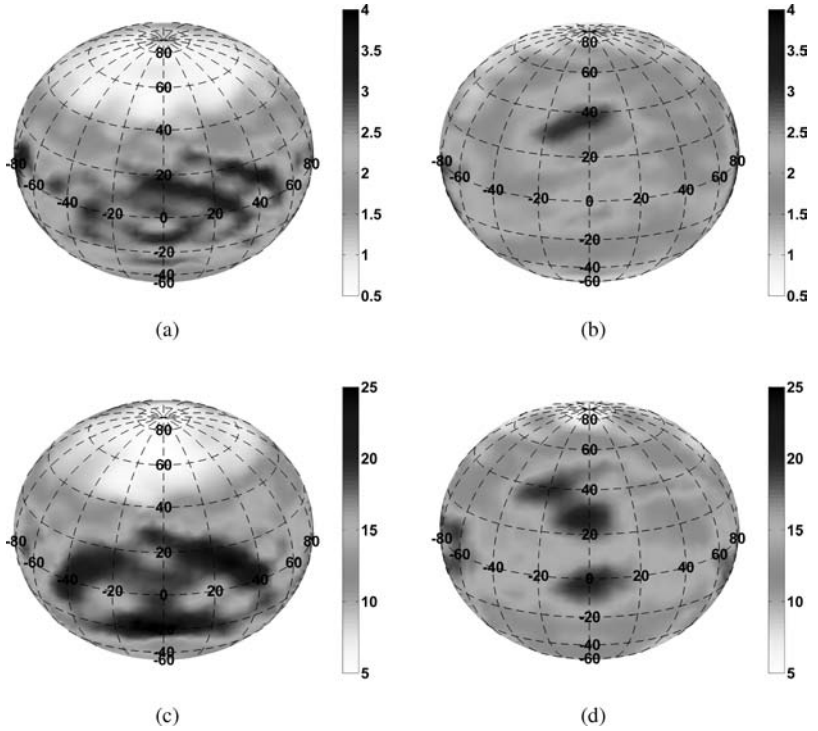


Figure 5: Mean local distances for two echolocating bats. (a, b) Mean local distances of learned extended-tangent vectors. (c, d) Mean local distances of underlying HRTFs. Subject EF2 in *a* and *c*; subject EF3 in *b* and *d*. Mean distances are determined within each local neighborhood of uniformly distributed 1707 spatial positions on the hemisphere (*a* and *b*) with K-nearest neighborhood criteria ($K = 8$). The local mean distance for each subject decreases with elevation and reaches its minimum value near the north pole, similar to human subjects, and near the south pole (data for which was unavailable for human subjects).

distances between the learned global parameters corresponding to those points.

This sensorimotor approach is based on three important assumptions. We first allow the assumption that organisms are initially naive to the spatial properties of sound sources. Second, we limit the external sensory information to auditory signals only. Third, we postulate an interaction between the auditory system and the organism's motor state, that is, proprioception and motor actions. The first two assumptions may be viewed as a worst-case scenario for sound localization, ignoring any potential mechanisms of sound localization that might be hardwired in the brain (e.g., as a set of initial

conditions to be later modified by plasticity). But they do not significantly constrain the approach. The third assumption, in contrast, is crucial to the proposed computational scheme.

With an organism's observations of the sensory consequences of its self-generated motions, there is sufficient information to capture spatial properties of sounds. We stress that the solution provided is not to find a way to match acoustic inputs to corresponding spatial parameters but rather to show how the animal could learn that acoustic inputs have spatial properties.

6.1 Geometry of the Auditory Space. In this work we have shown that for a given motor state, m_o (0 degree azimuth and 0 degree elevation, for instance), it could be possible for the nervous system to learn the manifold of sound-source locations using small head movements. However, we have not shown whether the nervous system could identify the manifolds obtained at different motor states, m_k, as different images of a single entity—auditory space. In other words, it remains to be determined how the system learns the equivalence of points on two different manifolds in terms of representing the same relative location in auditory space.

What identifies a spatial point is the unique relationship between the acoustic-input changes and the known set of head movements that generate them. Hence, both sensory and motor inputs are necessary for the learning of the spatial points. In general, identical head movements at different starting positions (e.g., all the 30 degree leftward head turns), result in different proprioceptive inputs. Furthermore, acoustic input changes associated with these head movements could also be different at different initial head positions. As a result, spatial correspondence of the points representing sound-source locations on a given manifold, to the points on other manifolds obtained for different internal states, is not obvious to the organism.

In order to complete the representation of auditory space, the nervous system must also be able to maintain the identification of the spatial positions of sound sources despite head movements. To illustrate this, take two points, A and B, on the manifold at m_o, where the location corresponding to B is 1 degree to the left of the spatial location corresponding to A. A 1 degree leftward head movement maps these points, to $\hat{A}$ and $\hat{B}$, respectively, on a second manifold at a new motor state m_k. We have not yet addressed how the organism might establish the equality of the identical spatial locations corresponding to B and $\hat{A}$, since they are on different manifolds. This can be accomplished if two conditions are fulfilled. First, the organism needs to be able to generate the same head movements at any motor state. Then the nervous system may identify a mapping between two different manifolds via that movement (e.g., the movement that changes the motor state from m_o to m_k), allowing the pairs A and $\hat{A}$, and B and $\hat{B}$, to be compared to each other. But this knowledge is not sufficient to conclude the equality of the B and $\hat{A}$.

The second condition is that the sound-source-invariant features must be independent of motor state changes for at least one subset of motor actions: rigid movements. These movements should leave the related compensable external sensory input changes invariant. For instance, movements that rotate the body without changing relative positions of its parts would protect these invariant features. Similarly, head movements also qualify if their effect on HRTF is negligible. As a counterexample, pinna movements change the HRTF (Rice, May, Spirou, & Young, 1992; Young, Rice, & Tong, 1996), and hence the extended-tangent vectors, giving different sound-source-invariant features at different pinna orientations.

With these two conditions, it is then possible to unify the manifolds obtained at different motor states as one entity: auditory space. The same head movements at different motor states should compensate the same external changes or transformations of points on the learned representation of space. An equivalence class of different proprioceptive inputs with identical spatial displacements can be labeled as transformations, for example, a 1 degree leftward head movement, that results in spatial displacement of the sound-source location. In other words, the organism can now generate similar head movements at different motor states voluntarily. Vuillemin (1972) argues that construction of these transformations requires continuity of space as an idealization. Space continuity implies that a spatial displacement of sound-source location caused by an organism's motion can be iterated infinitely many times.

In addition, since the organism could establish the equality of the spatial locations corresponding to $\hat{A}$ and B, it is now possible to study the metric properties of auditory space. When the organism has a unified representation of relative spatial points in auditory space through rigid movements, the distance between any two points on the unified representation of space has a spatial meaning. The equality of the distance of any two points can be established if the transformations realized by the rigid head movements are commutative, meaning the order in which any two transformation is applied does not alter the mapping. A subgroup of rigid movements that commute will result in metric-preserving mappings of points in auditory space, that is, an isometry. The question of how an organism can learn these spatial movements has recently been addressed by Philipona et al. (2004) (though only for the local group properties), who provide a mathematical foundation to study the metric properties of sensory space from the perspective of the organism based on sensorimotor contingencies.

6.1.1 On the Neural Representations of the Invariant Features. We have shown that in order for the organism to identify spatial points, it has to access the auditory-motor features that are independent of sound-source spectrum. The nature of these invariant features is dependent on the neural representations of the acoustic inputs. Logarithmic coding makes them

readily identifiable, since in an ideal logarithmic coding scheme, the spatial displacements result in equal changes in the representations, independent of the sound spectra. Although there is an approximate linear relationship between spectral logarithmic magnitude changes and an auditory nerve discharge rates, a complete representation of the magnitude of the sound spectra requires combining the rate information from different sets of auditory nerve fibers, each of whose dynamic range is limited to different sound intensity levels (May & Huang, 1997). Hence, same amount of spectral magnitude change of an acoustic signal at different intensities may not be represented in similar ways by the auditory nerve responses. This complicates the computation of invariant features, for it cannot simply be determined by the identical neural response changes.

At this point, we do not have a firm idea on how these invariant features might be determined empirically by an organism. We might hypothesize the existence of a neural mechanism that processes the relationships between the relevant movements through proprioception or efferent motor command signals and their sensory consequences, and hence represent the invariant features. Evidence for the existence of neural processes of this kind, also known as internal models, has been provided in the cerebellum (Imamizu, Kuroda, Miyauchi, Yoshioka, & Kawato, 2003; Kawato et al., 2003). Internal models are thought to be involved in estimating sensory consequences of motor actions (Blakemore, Frith, & Wolpert, 2001). A cerebellum-like circuitry has also been shown, for instance, in the dorsal cochlear nucleus (DCN), a low-level auditory processing area thought to be involved in sound localization (Oertel & Young, 2004). The DCN receives somatosensory, proprioceptive signals in addition to auditory inputs (see references in Oertel & Young, 2004), all necessary components of an internal model. Oertel and Young (2004) have, for instance, proposed that cerebellum-like architecture in DCN could function to correct sound localization cues in relation to head and pinna movements in a similar way. If internal model–like structures do exist in the auditory system, these structures could capture the spatial invariant properties. For now, however, these possibilities remain as hypotheses.

6.2 Role of Sensorimotor Experience in Auditory Spatial Perception. Evidence is accumulating for the importance of sensory experience and the role of voluntary movements on the development of the exteroceptive senses, such as vision and hearing (Grubb & Thompson, 2004; Wexler & van Boxtel, 2005). Experiments on vision and hearing show that active movement is essential for the development of sensory perception and to adapt to changes that might occur after its acquisition (Held, 1955; Held & Hein, 1963; Hein et al., 1970; Muir & Hains, 2004). Recent parallel findings in human infants stress the importance of self-generated actions in the development of spatial perception (Campos et al., 2000). Studies investigating the effect of signal properties on sensory information processing also

reveal that normal development of the neural circuitry in the visual and auditory systems depends on the properties of the sensory inputs (White, Coppola, & Fitzpatrick, 2001; Chang & Merzenich, 2003). The maturation of the auditory space map in the superior colliculus (SC), a sensorimotor nucleus involved in orientation behavior, is selectively affected in guinea pigs raised in an omnidirectional noise environment (Withington-Wray, Binns, Dhanjal, Brickley, & Keating, 1990). Organization of primary auditory cortex in rat is shaped by salient acoustic inputs (Chang & Merzenich, 2003). Alteration of auditory spectral-spatial features also disrupts the development of the topographic representation of acoustic space in the SC (Schnupp, King, & Carlile, 1998; King et al., 2000). This body of evidence suggests that normal sensory development requires exposure to relevant sensory inputs.

Experience-dependent plasticity and adaptation is not limited to early postnatal development. Adult humans and other animals have been shown to adapt to altered acoustic experience. Adult humans listening to sounds in the environment through modified HRTFs can reacquire the ability to localize sound with altered cues (Hofman et al., 1998). More recently, Kacelnik et al. (2006) have shown that adult ferrets that are subjected to altered acoustic spatial cues can relearn to localize sounds only when they are required to use the new cues in a behaviorally relevant task. In both studies, subjects were allowed the opportunity to experience and learn the altered auditory-motor relations; however, the potential role of the sensorimotor learning was not systematically studied.

Plasticity in the spatial properties of the auditory SC maps has been demonstrated during adulthood and facilitated by behaviors that require spatial hearing for ferret (King et al., 2001) and barn owl (Bergan, Ro, Ro, & Knudsen, 2005). Thus, a computational theory of sound localization should include mechanisms that can recalibrate after changes in sensory experience. Since the sensorimotor approach is inherently experience driven, it can easily capture this observed plasticity.

More evidence comes from sensory substitution experiments performed with blind human subjects (Bach-y-Rita & Kercel, 2003; Bach-y-Rita, 2004). There, providing spatial information through a different sensory modality (tactile stimulation) was enough to allow subjects to perceive spatial information, and the effect diminished if subjects were not allowed to interact with the environment. This phenomenon can be explained by the fact that sensorimotor contingencies were similar between the visual and substituting tactile inputs (O'Regan & Noë, 2001).

An interesting example also comes from studies of blind human infants equipped with sonar aids. These sonar devices emit ultrasound pulses and use received echo parameters to modulate a tone signal that is played to the ears (the direction information of the echo is encoded by creating an intensity difference in the tone signals at the two ears, and the distance information is represented by the frequency of the tone). These studies

have shown that the infants use the device as a new sensory organ to monitor their environment (Bower, 1989). What makes these sonar studies especially interesting is that the incoming acoustic information does not include most of the natural spatial cues that humans use to localize sound sources. Thus, specialized auditory circuits that process particular spatial cues (e.g., ITD or ILD) could not have contributed to the localization of the echo sources. Moreover, neural circuits that could interpret such artificial information could not have been innately hardwired as such. Yet the subjects were able to interpret and use the spatial information that was available through these devices. This again can be explained by the theory asserting that the brain monitors sensorimotor contingencies to interpret space rather than relying on innately placed circuitry that is specifically designed to serve a particular function, such as ILD and ITD processing.

Using a virtual sound-source localization experiment in which subjects could interact with the acoustic environment, Loomis et al. (1990) showed that in the absence of a spectral-shaping effect of the pinnae with limited and unnatural spatial cues, subjects can still externalize and localize sound sources. These findings suggest that spatial perception may not be purely innate, requiring voluntary actions to develop and maintain it.

We cannot ignore, however, innate (nonnaive) components of exteroceptive sensory processing. Organisms might plausibly use genetically wired information to lay the foundations of the computations we have been examining, and it is the experience gained by active monitoring of the environment that shapes, tunes, and calibrates these structures to generate meaningful interpretation of the sensory signals (Clifton, 1992; Muir & Hains, 2004). Behavioral studies on young animals demonstrate that immediately after birth, or coinciding with the onset of hearing, they show the ability to orient toward sound sources (Kelly & Potash, 1986). This behavior is slow and not as accurate compared to that of adults. Moreover, during development, accuracy of orientation to sound sources shows a U-shape function, such that accuracy of the orientation behavior initially decreases with age, for example, two to three in human infants (Muir et al., 1989) and 24 to 27 days in gerbils (Kelly & Potash, 1986), and then slowly increases in accuracy and finally reaches adult levels. It has been suggested that the initial acoustic orienting response might be diminished or disappear as the development of the forebrain progresses. Thus, later emergence of the orientation toward a sound-source location observed in the young animals might reflect localization behavior controlled by the midbrain and forebrain structures (Muir & Hains, 2004). Muir and Hains (2004) proposed that orienting to sound sources in early infancy is an example of a reflex that disappears later in development. The advantage of such reflexes in the development and learning of more complex behavior in robots has been proposed recently by Metta (2000).

6.3 Multisensory Nature of Spatial Hearing. The examples presented in this study are limited to hearing as the only exteroceptive sense. For this restriction, the HRTF must satisfy certain conditions that allow the organism to identify points in space without the need for a reference provided by another exteroceptive sense. This reference is not (though it might have been) essential for the organism to define a point as the set of acoustic inputs from which the same motor actions generate the reference sensory input. We have argued that a point in space can be identified by the organism based on the observation that different acoustic inputs originated from the same relative spatial location show similar changes to motor actions. We employ this assumption because there is considerable evidence provided by studies on blind subjects suggesting that sound localization can develop without the help of another distal exteroceptive sense such as vision (Ashmead et al., 1998; Zwiers, van Opstal, & Cruysberg, 2001b; Lewald, 2002). However, we fully recognize the potential importance of vision, when available. The influence of vision over the auditory system has been demonstrated in barn owl (Knudsen & Knudsen, 1989) and in human; particularly under noisy conditions, vision is thought to be involved in the calibration of sound-source elevation cues (Zwiers, van Opstal, & Cruysberg, 2001a). But the fact that sound localization can develop in the absence of vision suggests that visual influence is not required, and so a computational model of sound localization should not require supplemental sensory information for calibration (Kacelnik et al., 2006).

6.4 Role of the Motor State in Sound Localization. One of the important aspects of the sensorimotor approach is the organism's ability to monitor its motor states and associate them with its changing acoustic inputs. This requires the motor system, proprioception, and the auditory system to interact with each other. Recent studies provide evidence suggesting that these relations do exist. Influence of proprioception on sound localization has been shown in relation to eye (Lewald, 1997) and head positions (Lewald et al., 2000), direction of gravity (DiZio et al., 2001), and whether the head is free to move (Tollin, Populin, Moore, Ruhland, & Yin, 2005). Computation of sound localization has also been shown to be influenced by the vestibular system (Lewald & Karnath, 2000). Vliegen, van Grootel, and van Opstal (2004) proposed that head position signals interact with the processing of the acoustic spatial cues by way of modulating each frequency channel in a frequency specific manner. Electrophysiological findings by Kanold and Young (2001) in cats have shown that ear movements influence neurons DCN. This body of evidence suggests that computation of sound localization does not solely depend on the acoustic inputs.

6.4.1 Reference Frame of Sound Location. We have shown that, for a given motor state, an organism can capture a set of global parameters that represents the spatial locations of sound sources. These parameters can be different for different motor states (e.g., for an animal with mobile pinnae).

A change in pinna position induces changes in the HRTF that may not be accounted for by a simple rotation of the HRTF before the position change (Young et al., 1996). The motor state of the pinnae will determine the operating HRTF at every instant. With our method, one can produce a family of global parameters associated with the different pinna states. These parameters represent sound locations in a pinna-related reference system. However, representations of the sound locations in different reference frames that are attached to the head, body, or the exterior space are important for the behavior of an organism. One psychoacoustical study demonstrates that the sound-source locations are represented in a body-centered reference frame (Goossens & van Opstal, 1999). In order for the auditory system to use a body-centered reference frame, proprioceptive information from the pinna and the head should be taken into consideration by the system. Vliegen et al. (2004) suggest that the auditory system processes dynamically varying acoustic information caused by self-generated head movements in such a way that a stable representation of the sound-source location is constructed. From an animal's point of view, the environment surrounding the animal is stable as it moves. This requires the ability to distinguish sensory input changes caused by self-generated movements from those that are the result of changes in the environment. This can be achieved by using proprioceptive information, plus the ability to predict sensory consequences of the organism's actions, that is, sensorimotor expertise. It has been shown that human subjects represent visual information in an allocentric reference frame if the sensory consequences of their actions are predictable. When the sensory consequences of the movements are not predictable or the movements are involuntary, the representation is shown to be in egocentric reference frame (eye-centered) (Wexler, 2003). Our computational scheme can be extended to create a body-centered or allocentric representation of sound-source location. As mentioned earlier, different global parameters obtain at different motor states are related to each other by a coordinate transformation (Vuillemin, 1972).

6.5 Localization of Sound Sources Without Head Movements. We have shown that it is possible for a naive organism to obtain the spatial parameters, of sound-source directions using voluntary head movements as a tool to explore the sensory input space. In fact, knowledge of the auditory consequences of voluntary movements has been shown to be very effective to estimate both azimuth and elevation of a sound source even for a spherical head with no pinnae (Handzel & Krishnaprasad, 2002). However, it is well known from common psychoacoustic studies that localization of a sound source does not require head movements. Rather, a subject can localize acoustic signals based on the acoustic information received at the ears. How then can the sensorimotor approach account for the sound localization without motion? We assert that sensorimotor early experience, during development, is necessary for accurate sound localization.

Sensorimotor interactions give the organism the means to identify and parameterize the points in space without requiring prior knowledge of the spatial parameters or the organism-environment geometry. We posit that it is through head movements that the organism learns this geometry and the necessary parameters. With this knowledge, it is then possible to further investigate the properties of the acoustic signals to determine a relationship between these inputs and their corresponding spatial parameters in the absence of head movements. We have also discussed that invariant features, obtained from sensorimotor contingencies and independent of sound spectra, should be unique to a particular location for it to be localized unambiguously. Thus, all that the organism needs is to discover the invariant feature associated with each point in space and recognize it in the acoustic signals received at the ears. The access to these features is possible through motor actions. Once the organism is able to learn these features, it is possible to explore a function that maps the acoustic inputs to the internal coordinates of the spatial positions. Note that because of the smoothness of the orbits, the invariant feature will also change smoothly across space. This property then will allow interpolation at the unexperienced acoustic sensory inputs.

6.6 Neurophysiological Implications. Most neurophysiological studies investigating spatial properties of auditory neurons are limited to cases in which animals are prevented from moving their bodies, heads, and ears. Involvement of anesthesia may also limit capturing the auditory system's normal function. For well-studied animals like bats and cats, localization cues are subject to change in relation to ear position. One would expect to see the effects of different motor states on the processing of spatial information, which may not be accessible when animals are limited in their ability to move. Thus, it would be informative to study the effect of movement of the head and ears on the auditory nuclei that are thought to be involved in spatial information processing. The proposition of the distributed effect of head motion across frequency channels suggests that this can happen in different parts of the auditory system (Vliegen et al., 2004). There is also evidence of somatosensory influence on the DCN in cats (Kanold & Young, 2001) and proprioceptive influence on the auditory cortex (Alexeenko & Verderevskaya, 1976). Effects of eye position on auditory processing have been reported in the inferior colliculus (IC) (Groh, Trause, Underhill, Clark, & Inati, 2001; Zwiers, Versnel, & van Opstal, 2004) and superior colliculus (SC) (Jay & Sparks, 1984) in monkey. These findings suggest that the auditory system receives information from many sensory modalities, which may contribute to spatial processing.

6.7 Applications to Robotics. Sensorimotor theory offers important applications in the field of robotics and design of intelligent systems. Even if the tabula rasa assumption may not be completely valid for young living

organisms, it is typically valid for artificial systems designed to interact with the surroundings. The importance of autonomous behavior in robots has been long recognized: a system cannot be programmed to handle every possible scenario of organism-environment interaction, so it is important that robots have the ability to learn and adapt. When designing sensors, it is critically important that the sensory information not deviate from the tolerable limits of its parameters, since the robot's interpretation of its sensory inputs is dependent on how reliable they are. The sensorimotor approach allows flexibility to the design: instead of hard-coding the properties of the sensors and the interpretation of the information provided by them, flexible algorithms can be designed, using sensorimotor principles, to allow a robot to calibrate its sensors and choose the information that is useful for a given task. This provides the freedom to the designer determining how much of the hard-coding should go in the system.

7 Conclusion

In this article, we have proposed a computational method for the learning of the auditory space using sensorimotor theory (O'Regan & Noë, 2001; Poincaré, 1929), an unexplored issue of the problem of sound localization. We have argued that a computational theory of sound localization should be able to explain the experience-dependent nature of the computation as well as its dependence on other sensory inputs. This computational method provides a framework under which integration of experience-dependent plasticity and multisensory information processing aspects of sound localization can be achieved. By way of examples from humans and bats, we have shown that a naive organism can learn to localize sound based solely on dynamic acoustic inputs and their relation to motor state.

Acknowledgments

We thank P. S. Krishnaprasad, Kaushik Ghose, and Benjamin Falk for their valuable comments and and Fei Sha for providing assistance and software for conformal component analysis for manifold learning. This work was supported by NSF grant IBN-0111973, NIH research grants R01-MH056366, R01-EB004750 (C. F. Moss, PI), and the Comparative and Evolutionary Biology of Hearing Center grant P30-DC004664 (R. J. Dooling, PI).

References

Alexeenko, N. Y., & Verderevskaya, N. N. (1976, Dec). Proprioceptive effects on evoked responses to sounds in the cat auditory cortex. *Exp. Brain. Res.*, *26*(5), 495–508.

Algazi, V. R., Duda, R. O., Morrison, R. P., & Thompson, D. M. (2001). The CIPIC HRTF database. In *Proceedings of the 2001 IEEE Workshop on Applications of Signal*

Processing to Audio and Acoustics (pp. 99–102). New Paltz, NY: Mohonk Mountain House.

Ashmead, D. H., Wall, R. S., Ebinger, K. A., Eaton, S. B., Snook-Hill, M. M., & Yang, X. (1998). Spatial hearing in children with visual disabilities. *Perception*, *27*(1), 105–122.

Aytekin, M., Grassi, E., Sahota, M., & Moss, C. (2004). The bat head-related transfer function reveals binaural cues for sound localization in azimuth and elevation. *J. Acoust. Soc. Am.*, *116*(6), 3594–3605.

Bach-y-Rita, P. (2004). Tactile sensory substitution studies. *Ann. N.Y. Acad. Sci.*, *1013*, 83–91.

Bach-y-Rita, P., & Kercel, S. W. (2003). Sensory substitution and the human-machine interface. *Trends. Cogn. Sci.*, *7*(12), 541–546.

Belkin, M., & Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, *15*(6), 1373–1396.

Bergan, J. F., Ro, P., Ro, D., & Knudsen, E. I. (2005). Hunting increases adaptive auditory map plasticity in adult barn owls. *J. Neurosci.*, *25*(42), 9816–9820.

Blakemore, S. J., Frith, C. D., & Wolpert, D. M. (2001). The cerebellum is involved in predicting the sensory consequences of action. *Neuroreport*, *12*(9), 1879–1884.

Blauert, J. (1997). *Spatial hearing*. Cambridge, MA: MIT Press.

Bower, T. G. R. (1989). In R. C. Atkinson, G. Lindzey, & R. F. Thompson (Eds). *Rational infant: Learning in infancy*. New York: Freeman.

Campos, J., Anderson, D., Barbu-Roth, M., Hubbard, E., Hertenstein, M., & Witherington, D. (2000). Travel broadens the mind. *Infancy*, *1*, 149–219.

Chang, E. F., & Merzenich, M. M. (2003). Environmental noise retards auditory cortical development. *Science*, *300*(5618), 498–502.

Clifton, R. K. (1992). The development of spatial hearing in human infants. In A. Werner & E. W. Rubel (Eds.). *Developmental psychoacoustics* (pp. 135–157). Washington, DC: American Psychological Association.

Clifton, R. K., Gwiazda, J., Bauer, J. A., Clarkson, M. G., & Held, R. M. (1988). Growth in head size during infancy: Implications for sound localization. *Developmental Psychology*, *24*(4), 477–483.

Colburn, H. S., & Kulkarni, A. (2005). Models of sound localization. In A. N. Popper & R. R. Fay (Eds.). *Sound source localization* (Vol. 25, pp. 272–316). New York: Springer.

DiZio, P., Held, R., Lackner, J. R., Shinn-Cunningham, B., & Durlach, N. (2001). Gravitoinertial force magnitude and direction influence head-centric auditory localization. *J. Neurophysiol.*, *85*(6), 2455–2460.

Evans, M. J., Angus, J. A. S., & Tew, A. I. (1998). Analyzing head-related transfer function measurements using surface spherical harmonics. *J. Acoust. Soc. Am.*, *104*, 2400–2411.

Fermuller, C., & Aloimonos, Y. (1994, June). *Vision and action* (Tech. Rep. Nos. CAR-TR-722,CS-TR-3305). College Park, MA: Computer Vision Laboratory Center for Automation Research, University of Maryland.

Getzmann, S. (2002). The effect of eye position and background noise on vertical sound localization. *Hear. Res.*, *169*(1–2), 130–139.

Goossens, H. H., & van Opstal, A. J. (1999). Influence of head position on the spatial representation of acoustic targets. *J. Neurophysiol.*, *81*(6), 2720–2736.

Griffin, D. (1958). *Listening in the dark*. New Haven, CT: Yale University Press.
Groh, J. M., Trause, A. S., Underhill, A. M., Clark, K. R., & Inati, S. (2001). Eye position influences auditory responses in primate inferior colliculus. *Neuron*, *29*(2), 509–518.
Grubb, M. S., & Thompson, I. D. (2004). The influence of early experience on the development of sensory systems. *Curr. Opin. Neurobiol.*, *14*(4), 503–512.
Handzel, A. A., & Krishnaprasad, P. S. (2002). Biomimetic sound-source localization. *IEEE Sensors Journal*, *2*(6), 607–616.
Hein, A., Held, R., & Gower, E. C. (1970). Development and segmentation of visually controlled movement by selective exposure during rearing. *J. Comp. Physiol. Psychol.*, *73*(2), 181–187.
Held, R. (1955). Shifts in binaural localization after prolonged exposures to atypical combinations of stimuli. *Am. J. Psychol.*, *68*, 526–548.
Held, R., & Hein, A. (1963). Movement-produced stimulation in the development of visually guided behavior. *J. Comp. Physiol. Psychol.*, *56*(5), 872–876.
Hofman, P. M., van Riswick, J. G., & van Opstal, A. J. (1998). Relearning sound localization with new ears. *Nat. Neurosci.*, *1*(5), 417–421.
Imamizu, H., Kuroda, T., Miyauchi, S., Yoshioka, T., & Kawato, M. (2003). Modular organization of internal models of tools in the human cerebellum. *Proc. Natl. Acad. Sci. U.S.A.*, *100*(9), 5461–5466.
Javer, A. R., & Schwarz, D. W. (1995). Plasticity in human directional hearing. *J. Otolaryngol.*, *24*(2), 111–117.
Jay, M. F., & Sparks, D. L. (1984). Auditory receptive fields in primate superior colliculus shift with changes in eye position. *Nature*, *309*(5966), 345–347.
Kacelnik, O., Nodal, F. R., Parsons, C. H., & King, A. J. (2006). Training-induced plasticity of auditory localization in adult mammals. *PLoS Biol.*, *4*(4), e71.
Kanold, P. O., & Young, E. D. (2001). Proprioceptive information from the pinna provides somatosensory input to cat dorsal cochlear nucleus. *J. Neurosci.*, *21*(19), 7848–7858.
Kawato, M., Kuroda, T., Imamizu, H., Nakano, E., Miyauchi, S., & Yoshioka, T. (2003). Internal forward models in the cerebellum: fMRI study on grip force and load force coupling. *Prog. Brain. Res.*, *142*, 171–188.
Kelly, J. B., & Potash, M. (1986). Directional responses to sounds in young gerbils *(Meriones unguiculatus)*. *J. Comp. Psychol.*, *100*(1), 37–45.
King, A. J., Kacelnik, O., Mrsic-Flogel, T. D., Schnupp, J. W., Parsons, C. H., & Moore, D. R. (2001). How plastic is spatial hearing? *Audiol. Neurootol.*, *6*(4), 182–186.
King, A. J., Parsons, C. H., & Moore, D. R. (2000). Plasticity in the neural coding of auditory space in the mammalian brain. *Proc. Natl. Acad. Sci. U.S.A.*, *97*(22), 11821–11828.
Knudsen, E. I. (1982). Early auditory experience modifies sound localization in barn owls. *Nature*, *295*, 238–240.
Knudsen, E. I., & Knudsen, P. F. (1985). Vision guides the adjustment of auditory localization in young barn owls. *Science*, *230*(4725), 545–548.
Knudsen, E. I., & Knudsen, P. F. (1989). Vision calibrates sound localization in developing barn owls. *J. Neurosci.*, *9*(9), 3306-3313.
Lewald, J. (1997). Eye-position effects in directional hearing. *Behav. Brain. Res.*, *87*(1), 35–48.

Lewald, J. (2002). Vertical sound localization in blind humans. *Neuropsychologia*, *40*(12), 1868–1872.

Lewald, J., Dörrscheidt, G., & Ehrenstein, W. (2000). Sound localization with eccentric head position. *Behav. Brain. Res.*, *108*(2), 105–125.

Lewald, J., & Ehrenstein, W. H. (1998). Influence of head-to-trunk position on sound lateralization. *Exp. Brain. Res.*, *121*(3), 230–238.

Lewald, J., & Karnath, H. (2000). Vestibular influence on human auditory space perception. *J. Neurophysiol.*, *84*(2), 1107–1111.

Loomis, J. M., Hebert, C., & Cicinelli, J. G. (1990). Active localization of virtual sounds. *J. Acoust. Soc. Am.*, *88*(4), 1757–1764.

May, B. J., & Huang, A. Y. (1997). Spectral cues for sound localization in cats: A model for discharge rate representations in the auditory nerve. *J. Acoust. Soc. Am.*, *101*(5 Pt. 1), 2705–2719.

Metta, G. (2000). *Babyrobot: A study on sensori-motor devolopment*. Unpublished doctoral dissertation, University of Genova.

Moore, D. R., & Irvine, D. R. (1979). A developmental study of the sound pressure transformation by the head of the cat. *Acta. Otolaryngol.*, *87*(5–6), 434–440.

Muir, D. W., Clifton, R. K., & Clarkson, M. G. (1989). The development of a human auditory localization response: A U-shaped function. *Can. J. Psychol.*, *43*(2), 199–216.

Muir, D., & Hains, S. (2004). The U-shaped developmental function for auditory localization. *Journal of Cognition and Development*, *5*(1), 123–130.

Oertel, D., & Young, E. D. (2004). What's a cerebellar circuit doing in the auditory system? *Trends Neurosci.*, *27*(2), 104–110.

O'Regan, J., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behav. Brain. Sci.*, *24*(5), 939–973; discussion 973–1031.

Philipona, D., O'Regan, J. K., & Nadal, J.-P. (2003). Is there something out there? Inferring space from sensorimotor dependencies. *Neural Comput.*, *15*(9), 2029–2049.

Philipona, D., O'Regan, J. K., Nadal, J.-P., & Coenen, O. J.-M. (2004). Perception of the structure of the physical world using multimodal unknown sensors and effectors. In S. Becker, S. Thrün, & K. Obermayer (Eds.), *Advances in neural information processing systems*, *15*. Cambridge, MA: MIT Press.

Poincaré, H. (1929). *The foundations of science; Science and hypothesis, the value of science, science and method*. New York: Science Press. (G. B. Halsted, trans. of *La valeur de la science*, 1905)

Poincaré, H. (2001). *The value of science: Essential writings of Henri Poincaré.* (Ed. S. J. Gould). New York: Modern Library.

Prieur, J.-M., Bourdin, C., Vercher, J.-L., Sarès, F., Blouin, J., & Gauthier, G. (2005). Accuracy of spatial localization depending on head posture in a perturbed gravitoinertial force field. *Exp. Brain. Res.*, *161*(4), 432–440.

Rice, J., May, B., Spirou, G., & Young, E. (1992). Pinna-based spectral cues for sound localization in cat. *Hear. Res.*, *58*(2), 132–152.

Roweis, S. T., & Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, *290*(5500), 2323–2326.

Saff, E. B., & Kuijlaars, A. B. J. (1997). Distributing many points on a sphere. *Mathematical Intelligencer*, *19*(1), 5–11.

Saul, L. K., Weinberger, K. Q., Ham, J. H., Sha, F., & Lee, D. D. (2006). Spectral methods for dimensionality reduction. In O. Chapelle, B. Schölkopf, & A. Zien (Eds.), *Semisupervised learning*. Cambridge, MA: MIT Press.

Schnupp, J., King, A., & Carlile, S. (1998). Altered spectral localization cues disrupt the development of the auditory space map in the superior colliculus of the ferret. *J. Neurophysiol.*, *79*(2), 1053–1069.

Sha, F., & Saul, L. K. (2005). Analysis and extension of spectral methods for nonlinear dimensionality reduction. In *ICML '05: Proceedings of the 22nd International Conference on Machine Learning* (pp. 784–791). New York: ACM Press.

Sparks, D. L. (2005). An argument for using ethologically "natural" behaviors as estimates of unobservable sensory processes: Focus on "Sound localization performance in the cat: the effect of restraining the head." *J. Neurophysiol.*, *93*(3), 1136–1137.

Stevens, S. S., & Newman, E. B. (1936). The localization of actual sources of sound. *Am. J. Psychol.*, *48*(2), 297–306.

Tollin, D. J., Populin, L. C., Moore, J. M., Ruhland, J. L., & Yin, T. C. T. (2005). Sound-localization performance in the cat: The effect of restraining the head. *J. Neurophysiol.*, *93*(3), 1223–1234.

van Wanrooij, M. M., & van Opstal, A. J. (2005). Relearning sound localization with a new ear. *J. Neurosci.*, *25*(22), 5413–5424.

Vliegen, J., van Grootel, T. J., & van Opstal, A. J. (2004). Dynamic sound localization during rapid eye-head gaze shifts. *J. Neurosci.*, *24*(42), 9291–9302.

Vuillemin, J. (1972). Poincarés philosophy of space. *Synthèse*, *24*(1–2), 161–179.

Wang, J., Zhang, Z., & Zha, H. (2005). Adaptive manifold learning. In L. K. Saul, Y. Weiss & L. Bottou (Eds.), *Advances in neural information processing systems, 17* (pp. 1473–1480). Cambridge, MA: MIT Press.

Wexler, M. (2003). Voluntary head movement and allocentric perception of space. *Psychol. Sci.*, *14*(4), 340–346.

Wexler, M., & van Boxtel, J. J. A. (2005). Depth perception by the active observer. *Trends Cogn. Sci.*, *9*(9), 431–438.

White, L. E., Coppola, D. M., & Fitzpatrick, D. (2001). The contribution of sensory experience to the maturation of orientation selectivity in ferret visual cortex. *Nature*, *411*(6841), 1049–1052.

Wilmington, D., Gray, L., & Jahrsdoerfer, R. (1994). Binaural processing after corrected congenital unilateral conductive hearing loss. *Hear Res.*, *74*(1–2), 99–114.

Withington-Wray, D. J., Binns, K. E., Dhanjal, S. S., Brickley, S. G., & Keating, M. J. (1990). The maturation of the superior collicular map of auditory space in the guinea pig is disrupted by developmental auditory deprivation. *Eur. J. Neurosci.*, *2*(8), 693–703.

Young, E. D., Rice, J. J., & Tong, S. C. (1996). Effects of pinna position on head-related transfer functions in the cat. *J. Acoust. Soc. Am.*, *99*(5), 3064–3076.

Zha, H., & Zhang, Z. (2005). Spectral analysis of alignment in manifold learning. In *Acoustics, Speech, and Signal Processing, 2005. Proceedings (ICASSP '05), IEEE International Conference on* (Vol. 5, pp. 1069–1072). Piscataway, NJ: IEEE Press.

Zhang, Z., & Zha, H. (2004). Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. *SIAM Journal on Scientific Computing*, *26*(1), 313–338.

Zwiers, M. P., van Opstal, A. J., & Cruysberg, J. R. (2001a). A spatial hearing deficit in early-blind humans. *J. Neurosci.*, *21*(9), RC142, 1–5.

Zwiers, M. P., van Opstal, A. J., & Cruysberg, J. R. (2001b). Two-dimensional sound-localization behavior of early-blind humans. *Exp. Brain Res.*, *140*(2), 206–222.

Zwiers, M. P., Versnel, H., & van Opstal, A. J. V. (2004). Involvement of monkey inferior colliculus in spatial hearing. *J. Neurosci.*, *24*(17), 4145–4156.

Received December 20, 2005; accepted May 22, 2007.