

THE MANIFOLDS OF SPATIAL HEARING

Ramani Duraiswami and Vikas C. Raykar

Perceptual Interfaces and Reality Lab., Institute for Advanced Computer Studies,
Department of Computer Science, University of Maryland, College Park

ABSTRACT

We present exploratory studies on learning the non-linear manifold structure, in Head Related Impulse Responses (HRIRs). We use the recently popular Locally Linear Embedding [1] technique. The lower dimensional manifold encodes the perceptual information in the HRIRs, namely the direction of the sound source. Based on this we propose a new method for HRIR interpolation. We also propose that the distance between two HRIRs of an individual be taken as the geodesic distance on the learned manifold.

1. HUMAN SPATIAL HEARING

Humans have an amazing ability to localize a sound source, i.e., determine the range, elevation and azimuth angles of the direction of the sound source. The major mechanisms responsible for the directional capability of the human hearing system have been fairly well understood though not completely [2, 3]. One of the primary cues responsible for localization of the sound source are the Interaural Time Difference (ITD) and the Interaural Level Difference (ILD) cues. However ITD and ILD cues alone do not completely explain the source localization mechanism. For example, for all points lying on the "cone of confusion" (on half of the hyperboloid of revolution with vertex as the center of the head), the ITD and ILD cues are essentially the same. Yet we have the ability to localize the sound source in the vertical plane. Perceptual experiments done with virtual sources rendered using just ITD and ILD cues show that while the lateral placement of the source is correct, the perceived range and elevation are not. Additional important acoustic cues arise from the scattering of the sound by the head, torso and the pinna. The combined effects of different scatterers can be encapsulated in terms of the total spectral filtering provided by the torso, head and the pinna. This filtering can be described by a complex frequency response function called the Head Related Transfer Function (HRTF). The corresponding impulse response is called the Head Related Impulse Response (HRIR), which can be experimentally measured. The spectral features in the HRTF due to pinna diffraction and scattering are known to provide cues for vertical localization [4]. By manipulating the cues responsible for the directional hearing capability a virtual audio system that places the sound at any given location can be built by using just a pair of headphones or only two cross-talk canceled loudspeakers.

2. MANIFOLD REPRESENTATION

A HRIR of N samples can be considered as a point in N dimensional space. Consider all HRIRs in the mid-sagittal plane as shown in Figure 1. Each HRIR (corresponding to one elevation and azimuth) is a point in the higher dimensional space. As the elevation is varied smoothly, the points essentially trace out a one-dimensional manifold in N -dimensional space. Manifolds arise naturally whenever there is a smooth variation of parameters,

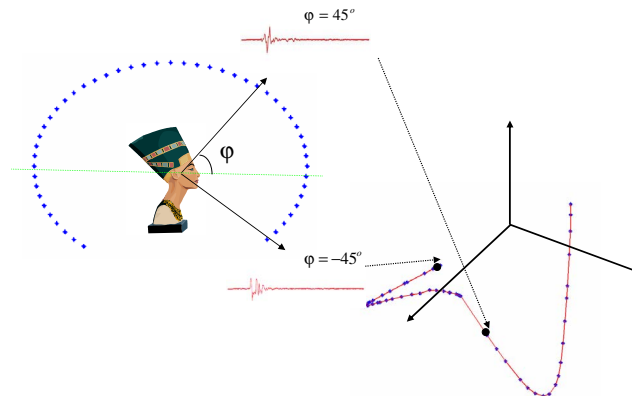


Fig. 1. Conceptual diagram of a one-dimensional manifold embedded in a higher dimensional space.

like the elevation angle in our case. Manifolds encode the perceptual information in a given signal. For all the HRIRs in the median plane the dominant perceptual information is the elevation of the source. The natural order is preserved in the low-dimensional manifold. In the N dimensional Euclidean space of the original HRIRs, two HRIRs corresponding to far apart elevations may still be very close to each other. However on the one-dimensional manifold, where we measure the distance between two points as the length of the geodesic on the manifold, they are far apart. If we can unfold this low-dimensional manifold we have a good perceptual representation of the signal. Manifolds could prove to be crucial for understanding how perception of the direction arises from the dynamics of neural networks in the brain [5]. Nonlinear manifold techniques essentially help to unfold the manifold giving a low dimensional representation [5].

3. NONLINEAR MANIFOLD LEARNING

Let Y be a d dimensional domain contained in a Euclidean space R^d . Let $f : Y \rightarrow R^D$ be a smooth embedding for some $D > d$. The goal of manifold learning is to recover Y and f given N points in R^D . Isomap [6] and Locally Linear Embedding (LLE) [1] are two techniques which provide implicit description of the mapping f . Given $X = \{x_i \in R^D \mid i = 1 \dots N\}$ find $Y = \{y_i \in R^d \mid i = 1 \dots N\}$ such that $\{x_i = f(y_i) \mid i = 1 \dots N\}$. Note that we are implicitly inverting the generative model without explicit parametrization of the generative function f .

Without imposing any restrictions on f the problem is ill-posed. The simplest case is a linear isometry i.e. f is a linear mapping from $R^d \rightarrow R^D$. In this case Principal Component Analysis (PCA) recovers the d significant dimensions of the observed data. Two other possibilities are considered in [7]: f can be either a isometric embedding or a conformal embedding. An isometric embedding preserves infinitesimal lengths and angles while a conformal embedding preserves only infinitesimal angles. The Isomap

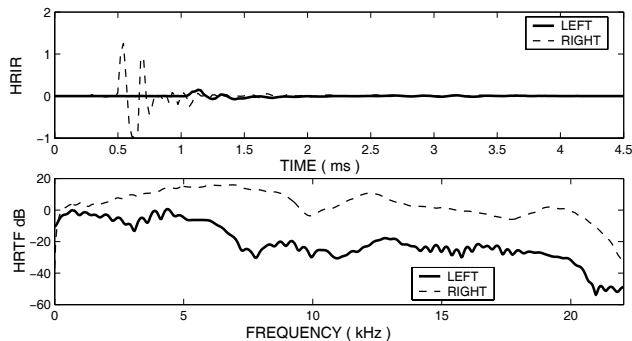


Fig. 2. HRIR and HRTF for the left and the right ear when the source is directly in front of the right ear at a distance of 1m from the center of the head.

algorithm can recover an isometric embedding. LLE can recover both isometric and conformal embeddings. In our case since we do not know the nature of our embedding we use the LLE, since it has a good representational capacity and does not make any assumptions regarding manifold structure. Also LLE is computationally more efficient since it uses sparse matrices.

3.1. Locally Linear Embedding

LLE models local neighborhoods as linear patches and then embeds them in a lower dimensional manifold [1]. The LLE algorithm is summarized below (For more details see [7]). For each data point X_i find its K nearest neighbors. We expect each data point and its neighbors to lie on or close to a locally linear patch of the manifold. Each point can be written as a linear combination of its neighbors. Compute the weights W_{ij} that best linearly reconstructs X_i from its neighbors. If the data lie on or near a smooth nonlinear manifold of lower dimensionality then there exists a linear mapping (consisting of a translation, rotation and rescaling) that maps the higher dimensional coordinates of each neighborhood to global internal coordinates on the manifold. By design, the weights that minimize the reconstruction errors are invariant to rotation, rescaling and translation of the data points. Hence the same weights that reconstruct the data points in D dimensions should reconstruct it in the manifold in d dimensions. The weights characterize the intrinsic geometric properties of each neighborhood. Compute the lower dimensional embedding vectors Y_i best reconstructed by W_{ij} .

4. THE HRIR MANIFOLD

We use the public-domain CIPIC HRTF database [8]. The coordinate system followed is the head centered interaural polar coordinate system. The interaural axis is the line passing through the center of the left and the right ears. The origin of this spherical coordinate system is the interaural midpoint which is exactly the midpoint of the line joining the two ears. The azimuth angle θ is the angle between a vector to the sound source and the vertical median plane or the midsagittal plane and varies from -90° to $+90^\circ$. The elevation ϕ is the angle from the horizontal plane to the projection of the source into the midsagittal plane, and varies from -90° to 270° .

The database contains HRIRs sampled at 1250 points around the head for 45 subjects. Azimuth is sampled from -80° to 80° and elevation from -45° to $+230.625^\circ$. The temporal sampling frequency is $f_s = 44.1kHz$. Each HRIR is 200 samples long corresponding to a duration of about 4.5ms. Figure 2 shows a typical

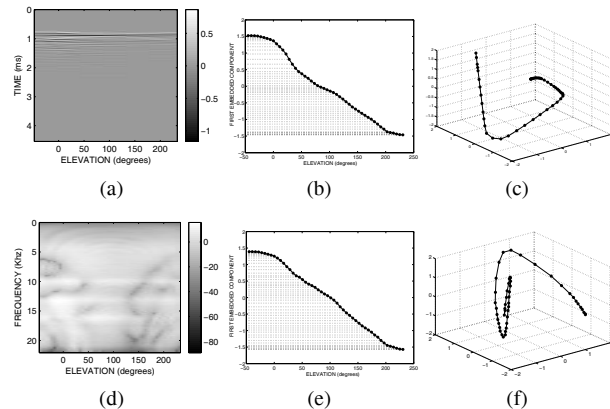


Fig. 3. (a) HRIR and (d) HRTF as a function of elevation for azimuth 0° . (b) and (e) The one-dimensional manifold recovered by the LLE technique using $K = 2$ neighbors for the HRIR and HRTF respectively. (c) and (f) shows the same manifold embedded in three dimensions recovered using $K = 4$ neighbors.

HRIR and the magnitude of the HRTF for both the left and the right ear for a person when the source is directly in front of the right ear at a distance of 1 m from the center of the head. The ITD can be clearly seen by comparing the HRIR for the right and the left ear. The difference in the spectral response for the right and the left ear are due to scattering by the head, torso, and the pinnae.

4.1. The HRIR manifold

Consider all the HRIRs lying in the midsagittal plane (i.e. 0° azimuth and all elevations) as shown in Figure 1. Figure 3(a) shows the HRIR for a particular subject for azimuth 0° as a function of elevation. The HRIR is displayed as a gray scale image with the grayscale corresponding to the amplitude of the HRIR. Figure 3(b) shows the one-dimensional manifold recovered using the LLE algorithm. $K = 2$ neighbors were used. The plot shows the distance on the one-dimensional manifold as a function of elevation. The corresponding manifold is shown embedded in three dimensions in Figure 3(c). Figure 3(b) is unrolled version of a manifold as in Figure 3(c), but in a high dimensional space. One interesting observation is that even though the elevation is sampled uniformly the points are not uniformly distributed in the one-dimensional manifold. In Figure 3(b) it can be seen that the points are clustered closely for negative elevations and elevations above 180° .

4.2. The HRTF manifold

In the strictest sense the HRIR is a not a minimum-phase system, because of the multiple-transmission paths associated with diffraction/reflections from different parts. However for simplicity we use only the magnitude spectrum. We ran the same algorithm on the HRTF spectrum magnitude instead of the HRIR. We obtained similar results. Figure 3(d) shows the HRTF magnitude spectrum in dB. Figure 3(e) and Figure 3(f) show the corresponding one-dimensional manifold.

4.3. Choice of K

The only free parameter that needs to be selected in the LLE algorithm is the number of neighbors K . First the algorithm can only recover embeddings with dimensionality strictly less than K . K is closely related to the intrinsic dimensionality of the data. This step is vulnerable to short-circuit errors if the neighborhood is too

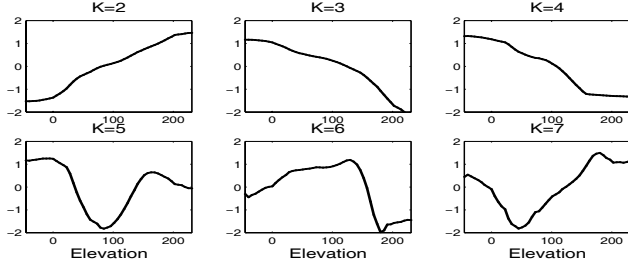


Fig. 4. The manifold recovered for different values of K .

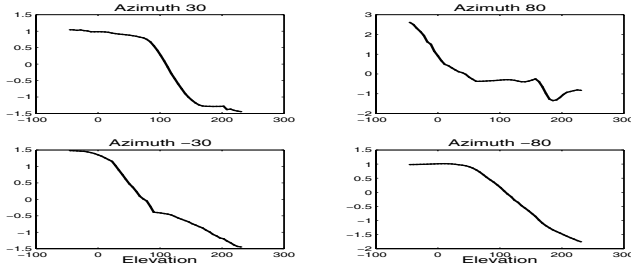


Fig. 5. The manifold recovered for different azimuth angles.

large with respect to the folds in the manifold on which the data points lie or if noise in the data moves the points slightly off the manifold. Even a single short-circuit error can lead to a drastically different (and incorrect) low-dimensional embedding [9]. In such cases appropriately selecting K is very essential for the algorithm. Choosing a very small neighborhood is not satisfactory, as this can fragment the manifold into a large number of disconnected regions. Figure 4 shows the manifold recovered for different values of K . For $K \leq 4$ the manifold is recovered. However as K increases the folding behavior can be seen. The reason for this is that the number of neighbors decides the the boundary of the linear patch. If the manifold is curved very closely then there may be problems of short-circuit. Figure 5 shows the manifold recovered for different azimuth angles. Figure 6 shows the manifold recovered for different subjects.

5. A NEW DISTANCE METRIC

One of the problem that frequently arises is how to compare any two given HRIRs i.e. how to formulate a distance metric in the space of HRIRs. Suppose we have modelled/interpolated a HRIR and often we have to find out how good is the modelled to the actual HRIR. The distance metric has to be perceptually inspired. The absolute justification however is to do psychoacoustical tests. In the absence of any good perceptual error metric the most commonly used one is the squared log-magnitude error of the spectrum of the HRIRs. If $H(\omega, \phi_1)$ and $H(\omega, \phi_2)$ are the frequency response of two HRIRs $h(n, \phi_1)$ and $h(n, \phi_2)$ corresponding to two elevation angles ϕ_1 and ϕ_2 then one measure of distance between the two HRIRs which is widely used is,

$$\epsilon(\phi_1, \phi_2) = \sqrt{\frac{1}{2\pi} \int_{-\pi}^{\pi} [20 \lg |H(\omega, \phi_1)| - 20 \lg |H(\omega, \phi_2)|]^2 d\omega} \quad (1)$$

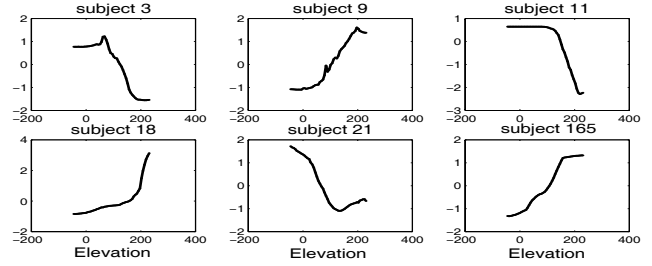


Fig. 6. The manifold recovered for different subjects.

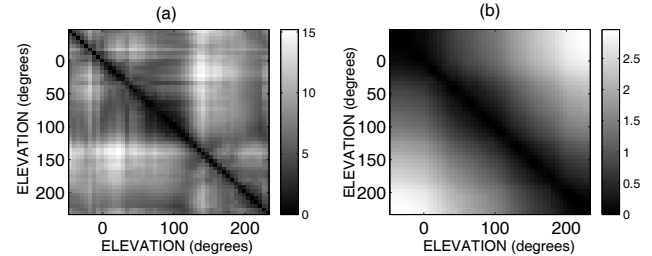


Fig. 7. (a) The distance matrix using the metric defined in Equation 1 (b) using the distance on the manifold.

It is tough to decide what aspects of a given signal are perceptually relevant. For our case of all HRIRs for different elevation angles, the obvious perceptual information to be extracted is the elevation of the source. A natural measure of distance would be the distance on the extracted one dimensional manifold. Figure 7(a) shows the distance (according to Equation 1) between each HRIR and all other HRIRs, for all elevations. Note that all elements along the diagonal must be zero. Figure 7(b) shows the same with the distance as measured on the manifold. It can be seen that Figure 7(b) is a better distance metric. The distance between two HRIRs is proportional to how far they are in elevation.

6. HRIR INTERPOLATION

As of now we have a implicit mapping from the high dimensional HRTF space to the one-dimensional elevation manifold. However it is also possible to go from the manifold to the signal representation. To do this we use the same property which was used by the LLE algorithm. The weights that reconstruct the data points in higher dimensions should reconstruct it in the lower dimensional embedded manifold. The weights characterize the intrinsic geometric properties of each neighborhood. If we want the HRTF for a new elevation ϕ we find the value of the lower-dimensional manifold at the required angle ϕ . The one-dimensional manifold can be linearly interpolated or some other higher order interpolation can be used to fit the lower dimensional manifold. Once we know the value on the manifold it can be written as a linear combination of its neighbors and compute the weights that best linearly reconstructs it from its neighbors. The same weights reconstruct the HRTF in the higher dimensional space. Figure 8 shows the original and the reconstructed HRTF for elevation $\phi = 0^\circ$. Figure 9 shows the same for all elevations. In Figure 9(b) each HRTF is obtained by reconstructing it from its $K = 2$ neighbors. Note that even though we are writing the HRTF as a linear combination of

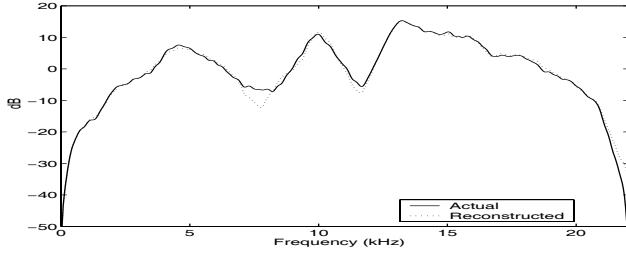


Fig. 8. The actual and the reconstructed HRTF for elevation 0° and azimuth 0° .

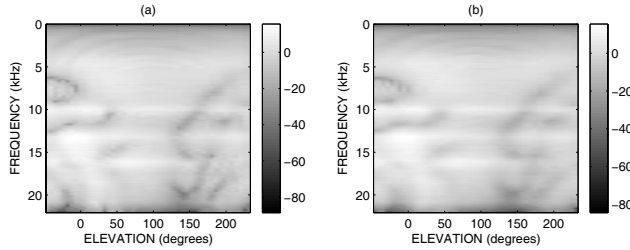


Fig. 9. The (a) actual and the (b) reconstructed HRTF for all elevations and azimuth zero.

its neighbors, it is not the same as linearly interpolating from the neighbors. We first learn the manifold, the interpolate in the manifold representation, and then go back to the original HRTF representation by exploiting the local linearity in the lower-dimensional manifold. To evaluate the proposed method of interpolation we first ran the LLE algorithm on all the elevation except one. Using this manifold we then generated the HRIR for the excluded elevation. The same was repeated for all elevations. Figure 10(a) shows the error between the actual and the interpolated HRTF using the error metric in equation 1 for different elevations. Figure 10(b) shows the same but using the distance on the manifold as a error metric. As pointed out in the previous section the manifold distance is a more reliable metric. Note that the error is large for lower elevations and elevations behind the head. The HRTF is more complicated in those elevations and the one-dimensional manifold is not able to capture all the details.

7. THE COMPLETE MANIFOLD

Until now we were concerned with all the HRTFs in the vertical plane. The same results can be extended to HRTF of all elevations and azimuths. Now we have a two dimensional manifold. Figure 11 shows the manifold for all elevations and azimuth from 0° to 45° . $K = 4$ neighbors were used to unfold the manifold. We would like to comment that the algorithm is not very stable when using the complete data. We got better results when considering only the manifold of elevations.

8. CONCLUSIONS

We presented a new representation for the HRTFs in terms of the elevation manifold they lie on. We also proposed a new distance metric and a new scheme for HRIR interpolation. Future work would include enforcing the intrinsic dimensionality in the LLE procedure and evaluation with other interpolation methods.

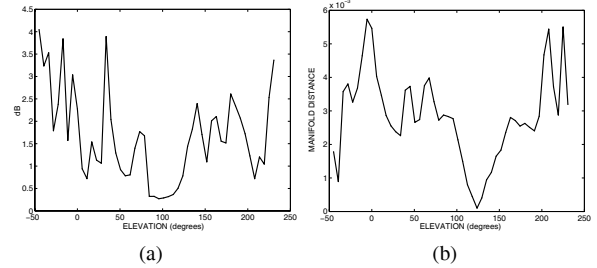


Fig. 10. Error between the actual and the interpolated one using the (a) error metric in Eq.1 and (b) the distance on the manifold as a error metric for different elevations.

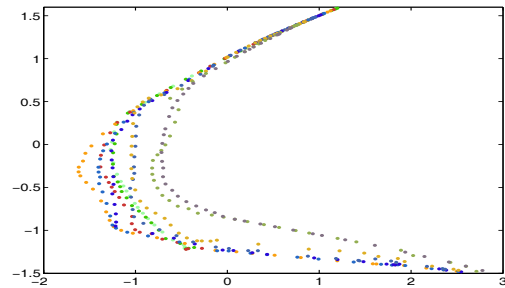


Fig. 11. The complete two dimensional manifold for all elevations and azimuths from 0° to 45° .

9. REFERENCES

- [1] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323–2326, december 2000.
- [2] J.P. Blauert, *Spatial Hearing (Revised Edition)*, MIT Press, Cambridge, MA, 1997.
- [3] J. C. Middlebrooks and D. M. Green, "Sound localization by human listeners," *Annual Review of Psychology*, vol. 42, pp. 135–159, 1991.
- [4] F. L. Wightman and D. J. Kistler, "Monaural sound localization revisited," *Journal of the Acoustical Society of America*, vol. 101, no. 2, pp. 1050–1063, Feb. 1997.
- [5] H. S. Seung and D. D. Lee, "The manifold ways of perception," *Science*, vol. 290, pp. 2268, december 2000.
- [6] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, pp. 2319–2323, december 2000.
- [7] V. de Silva and J. B. Tenenbaum, "Local versus global methods for nonlinear dimensionality reduction," *Advances in Neural Information Processing Systems*, vol. 15, 2003.
- [8] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," *Proc. 2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, Mohonk Mountain House, New Paltz, NY*, pp. 99–102, October 2001.
- [9] M. Balasubramanian, E. L. Schwartz, J. B. Tenenbaum, V. de Silva, and J. C. Langford, "The isomap algorithm and topological stability," *Science*, vol. 295, 2002.