Communicated by Richard Lippmann

A Spherical Basis Function Neural Network for Modeling Auditory Space

Rick L. Jenison
Kate Fissell
Department of Psychology, University of Wisconsin, Madison, WI 53706 USA

This paper describes a neural network for approximation problems on the sphere. The von Mises basis function is introduced, whose activation depends on polar rather than Cartesian input coordinates. The architecture of the von Mises Basis Function (VMBF) neural network is presented along with the corresponding gradient-descent learning rules. The VMBF neural network is used to solve a particular spherical problem of approximating acoustic parameters used to model perceptual auditory space. This model ultimately serves as a signal processing engine to synthesize a virtual auditory environment under headphone listening conditions. Advantages of the VMBF over standard planar Radial Basis Functions (RBFs) are discussed.

1 Introduction

Artificial neural networks and approximation techniques typically have been applied to problems conforming to an orthogonal Cartesian input space. In this paper we present a neural network operating on a problem in acoustics whose input space is best represented in spherical (or polar), rather than Cartesian coordinates. The neural network employs a novel basis function, the von Mises function, which is well adapted to spherical input, within the standard Radial Basis Function (RBF) architecture. The primary advantage of a basis on a sphere, rather than on a plane, is the natural constraint of periodicity and singularity at the poles. The RBF network architecture using a single layer of locally tuned units (basis functions) covering a multidimensional space is now well known (e.g., Broomhead and Lowe 1988; Moody and Darken 1988, 1989; Poggio and Girosi 1990). Gradient-descent learning rules akin to backpropagation that move and shape the basis functions to minimize the output error have also been proposed (Poggio and Girosi 1990; Hartman and Keeler 1991).

Our network was constructed to synthesize a continuous map of acoustic parameters used to simulate the virtual experience of free-field spatial hearing under headphones. The actual signal processing details

used to synthesize auditory space will for now be deferred, with the main focus being on the description of the spherical neural network and corresponding learning rules. It is anticipated that this neural network will have general application to functional approximation problems on the sphere, such as inverse kinematics of spherical mechanisms (Chiang 1988) and global weather prediction (Wahba and Wendelberger 1980; Ghil *et al.* 1981).

2 von Mises Basis Function (VMBF) Network

The network basis function is based on a spherical probability density function (p.d.f)[1] that has been used to model line directions distributed unimodally with rotational symmetry. The function is well-known in the spherical inferential statistics literature and commonly referred to as either the von Mises–Arnold–Fisher or Fisher distribution [see Mardia (1972) or Fisher *et al.* (1987)]. The kernel form of the p.d.f. was first introduced by Arnold (1941) in his unpublished doctoral dissertation. Historically, the function has served as an assumed parametric distribution from which spherical data are sampled; but we are not aware of its use as a basis function in the approximation theory literature. The expression for the von Mises basis function, dropping the constant of proportionality and elevational weighting factor from the p.d.f., is

$$\mathrm{VM}(\theta, \phi, \alpha, \beta, \kappa) = e^{\kappa[\sin\phi \sin\beta \cos(\theta-\alpha)+\cos\phi\cos\beta]} \tag{2.1}$$

where the input parameters correspond to a sample location in azimuth and elevation (θ, ϕ), a centroid in azimuth and elevation (α, β), and the concentration parameter κ. Application of the von Mises function requires an azimuthal range in radians from 0 to 2π and elevational range from 0 to π. Any sample (θ, ϕ) on the sphere will induce an output from each VMBF proportional to the solid angle between the sample and the centroid of the VMBF (α, β). The azimuthal periodicity of the basis function is driven by the $\cos(\theta - \alpha)$ term, which will be maximal when $\theta = \alpha$. The $(\sin\phi \sin\beta)$ term modulates the azimuthal term in the elevational plane, hence the requirement that ϕ range from 0 to π. As the sample elevation or the centroid elevation approaches either pole (0 or π), the multiplicative effect of $(\sin\phi \sin\beta)$ progressively eliminates the contribution of azimuthal variation and the $(\cos\phi\cos\beta)$ term dominates. κ is a shape parameter called the concentration parameter, where the larger the value the narrower the function width after transformation by the expansive function e. While other spherical functions have been proposed for approximation on the sphere [e.g., thin-plate pseudosplines (Wahba 1981)], the VMBF serves as a convenient spherical analogue of the well-known multidimensional gaussian on a plane (see Fig. 1). It behaves in a similar fashion to the planar gaussian with the centroid corresponding

[1] The p.d.f. is given by $f(\theta, \phi, \alpha, \beta, \kappa) = (\kappa/4\pi \sinh\kappa)e^{\kappa[\sin\phi \sin\beta \cos(\theta-\alpha)+\cos\phi\cos\beta]} \sin\phi$.

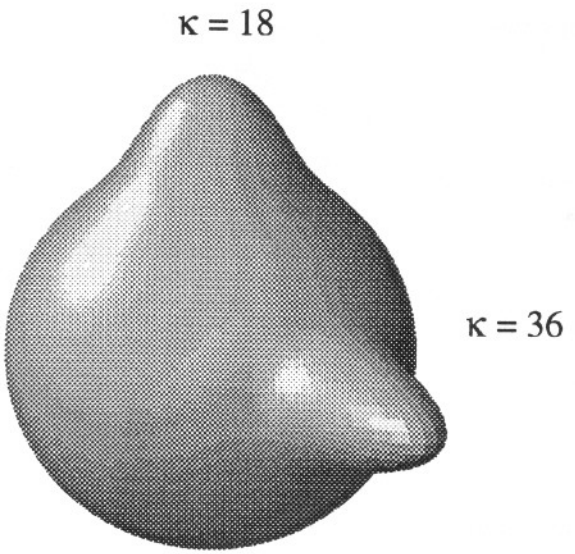

Figure 1: Three-dimensional rendering of two von Mises basis functions positioned on the unit sphere. The basis functions are free to move about the sphere, changing adaptively in position, width, and magnitude.

to the mean and $1/\kappa$ corresponding to the standard deviation. It differs from the thin-plate spline in that it has a parameter for controlling the width or concentration of the basis function, which allows the VMBF to focus resolution optimally.

The von Mises basis function serves as the activation function of the hidden layer units conforming to the RBF architecture as shown in Figure 2. The output of the ith output node, $f_i(\theta, \phi)$, when spherical coordinates are presented as input, is given by

$$f_i(\theta, \phi) = \sum_{j}^{J} w_{ij} \, \mathrm{VM}(\theta, \phi, \alpha_j, \beta_j, \kappa_j) \tag{2.2}$$

where $\mathrm{VM}(\theta, \phi, \alpha_j, \beta_j, \kappa_j)$ is the output of the jth von Mises basis function and w_{ij} is the weight connecting the jth basis function with the ith output node.

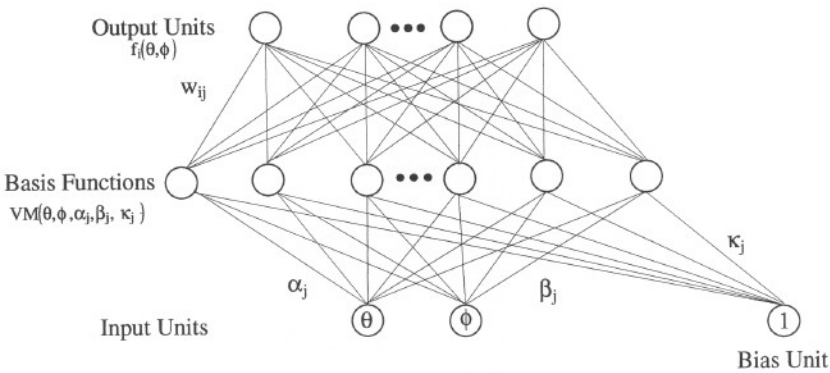

Figure 2: Architecture of the von Mises Basis Function (VMBF) neural network.

3 Parameter Learning

To optimize the approximation function given a fixed number of basis functions we apply a gradient-descent method on an error function to update the parameters of the network. In our case we require the sum-of-squared-error to be minimized. This technique has been applied successfully to gaussian RBF neural networks (Moody and Darken 1989; Poggio and Girosi 1990; Hartman and Keeler 1991) and we derive the analogous equations for the von Mises basis here. The error function for the pth M-dimensional training pattern is

$$E_p = \frac{1}{2}\sum_{i}^{M}[t_{ip} - f_{ip}(\theta, \phi, \Omega_i)]^2 \tag{3.1}$$

where t_{ip} is the ith element of the pth training pattern. The notation of the network output $f_{ip}(\theta, \phi, \Omega_i)$ includes Ω_i , which is a vector of changing network parameters

$$\Omega_i \equiv [\mathbf{w}_i^T, \alpha^T, \beta^T, \kappa^T] \tag{3.2}$$

Parameter values are learned through successive presentation of input–output pairs using the well-known update rule

$$\Omega_i^{\text{new}} = \Omega_i^{\text{old}} + \eta\ \Delta\Omega_i \tag{3.3}$$

where η corresponds to the learning rate, which can also be updated as learning proceeds. Obtaining $\Delta\Omega_i$ involves computing the error gradient by differentiating the error function with respect to each free parameter in the network. These derivations are relatively straightforward using

the chain rule and algebra, however are rather involved with respect to the bookkeeping of indices. For convenience Ω_i is omitted from the following derivations. The update expression for w_{ij} is just the Widrow–Hoff learning rule

$$\Delta w_{ij} = -\frac{\partial E}{\partial w_{ij}} = [t_i - f_i(\theta, \phi)]\,\mathrm{VM}(\theta, \phi, \alpha_j, \beta_j, \kappa_j) \tag{3.4}$$

The update expression for the azimuthal movement of the basis center can be derived in a similar fashion

$$\begin{aligned}\Delta\alpha_j = -\frac{\partial E}{\partial \alpha_j} &= \kappa_j[\sin\phi \sin\beta_j \sin(\theta - \alpha_j)] \\ &\times \sum_i^M \{[t_i - f_i(\theta, \phi)]w_{ij}\}\,\mathrm{VM}(\theta, \phi, \alpha_j, \beta_j, \kappa_j)\end{aligned} \tag{3.5}$$

as can the update expression for elevational movement

$$\begin{aligned}\Delta\beta_j = -\frac{\partial E}{\partial \beta_j} &= \kappa_j[\sin\phi \cos\beta_j \cos(\theta - \alpha_j) - \cos\phi \sin\beta_j] \\ &\times \sum_i^M \{[t_i - f_i(\theta, \phi)]w_{ij}\}\,\mathrm{VM}(\theta, \phi, \alpha_j, \beta_j, \kappa_j)\end{aligned} \tag{3.6}$$

and finally the concentration parameter

$$\begin{aligned}\Delta\kappa_j = -\frac{\partial E}{\partial \kappa_j} &= \left[\sin\phi \sin\beta_j \cos(\theta - \alpha_j) + \cos\phi \cos\beta_j\right] \\ &\times \sum_i^M \left\{[t_i - f_i(\theta, \phi)]\, w_{ij}\right\}\mathrm{VM}(\theta, \phi, \alpha_j, \beta_j, \kappa_j)\end{aligned} \tag{3.7}$$

The degree of improvement realized by the gradient-descent optimization of the parameter vector Ω_i over a direct matrix pseudoinverse solution of $\mathbf{w}^{\mathrm{T}}{}_i$ alone depends on the particular problem as well as the number of basis functions available. We have typically observed about a 2-fold improvement in the final root-mean-square (RMS) error for networks with a small number of basis functions, which will be discussed for our specific application in the following section.

4 Approximating Auditory Space

The VMBF neural network is ideally suited to the problem of learning a continuous mapping from spherical coordinates to acoustic parameters that specify sound source direction. The spatial position of a sound source in a listener's environment is specified by a number of factors related to the interaction of sound pressure waves with the pinna (external ear), head, and upper body. These interactions can be described mathematically in the form of a linear transformation, commonly referred to

as a "head-related transfer function," or HRTF. HRTFs are measured by sampling the air pressure variation in time near the eardrum as a function of known sound source location. Generally, HRTFs are visualized in the frequency (spectral) domain rather than in the Fourier equivalent time domain, because it is a more meaningful way of characterizing acoustic processing by the auditory system.[2] Henceforth, an n-dimensional vector will represent the HRTF, where each element of the vector corresponds to a discrete sample in frequency.

Figure 3 illustrates a typical spherical grid of speaker placements used to position sound sources, preferably within an anechoic environment. Studies of human sound localization most often use a coordinate system in degrees with the origin (0°, 0°) located at the intersection of the horizon and the medial saggital plane directly in front of the listener. We adhere to this convention for the remaining discussion, aware that coordinates must be converted from degrees to radians under the range constraints imposed by the VMBF (equation 2.1). Enforcement of these constraints is accomplished by mapping $-180° \leq \theta < +180°$ to $0 \leq \theta < 2\pi$ and $-90° \leq \phi \leq +90°$ to $0 \leq \phi \leq \pi$.

HRTFs change in complicated, but systematic ways as a function of sound direction relative to each ear. The HRTFs completely specify the sound source location because the measurements are made near the point where sound is transduced by the auditory system. The measurement and analysis techniques are well developed and have been detailed in the psychoacoustic literature (Wightman and Kistler 1989; Middlebrooks *et al.* 1989; Pralong and Carlile 1994). One practical use of these measurements is to create the sense of a sound coming from any "virtual" location under a headphone listening condition. This is accomplished by convolving an HRTF with any digitally recorded sound, effectively simulating the actual spectral filtering of the individual's external ear, and inducing an apparent location of the sound. Using the VMBF neural network to create the parameters for a virtual environment affords the ability to synthesize a set of HRTFs (for each ear) for any location in space, not just the spherical locations where measurements were obtained. Furthermore, the continuous modeled environment affords smooth auditory motion (sound trajectory). Others have recently applied standard regularization techniques to the problem of HRTF approximation (Chen *et al.* 1993), using a basis of planar thin-plate splines rather than spherical basis functions, and without benefit of gradient-descent learning. Their approximation performance is consistent with that of the standard class of planar RBF networks, hence subject to the distortions addressed by this paper.

From both a practical as well as a theoretical standpoint, the measured HRTFs are of much higher dimensionality than necessary for modeling purposes. The redundancy inherent in the measurement can be statis-

[2]Vision, on the other hand, is more naturally considered in spatial terms due to the retinotopic projection.

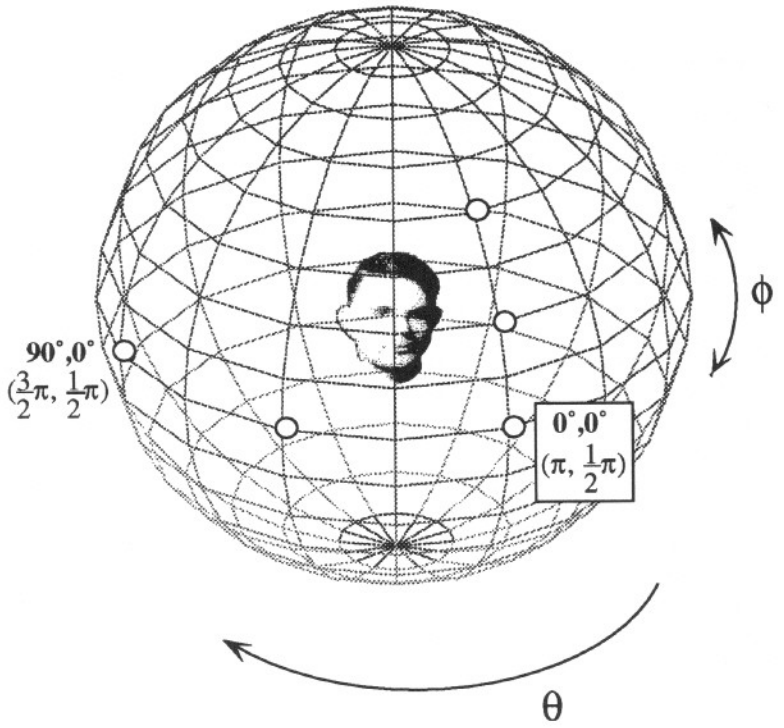

Figure 3: Coordinate system for sampling head-related transfer functions (HRTFs). Measurements are recorded near the eardrum of a subject placed in the center of a spherical speaker array. Azimuth is denoted by θ (ranging periodically from $-180°$ to $+180°$ or 0 to 2π) and elevation by ϕ (ranging from $-90°$ to $+90°$ or 0 to π). The origin (0°, 0°) of the coordinate system is directly in front of the listener.

tically removed via the technique of principal component analysis (see Kistler and Wightman 1992). Each HRTF magnitude spectrum consists of 256 real-valued frequency components for each ear and each sound source location in space, hence, a 256-dimensional vector. From the set of all measurements of sound source locations, we are able to reduce the dimensionality of the frequency components to six principal components (for each ear and location) that account for 98% of the spectral variance. This linear operation allows the use of six output units in the network rather than 256 frequency components while losing only a negligible

amount of information. Reconstruction to the original dimensionality of the HRTFs is a straightforward linear inverse transformation.

5 Learning Auditory Space

450 HRTF measurements from a single human ear served as the database for training a VMBF neural network.[3] These measurements, corresponding to discrete locations on the sphere, ranged from -170° to 180° in azimuth and -50° to 90° in elevation in 10° steps. At higher elevations the solid angle spanned is less, thus requiring fewer samples. The six principal components described above are the M elements of the training pattern t_{ip}. The total number of training patterns was 400, with 50 random patterns reserved for test data (i.e., data fed forward through the network without parameter updating). The network parameters were initialized by positioning the basis functions uniformly on the input space with a small degree of relative overlap and solving the output weights w_{ij} with a single pseudoinverse. Iterative training then proceeded with the successive presentation of training patterns, evaluation of gradient-descent equations, and update of free parameters. Figure 4 shows the training and intermediate testing history for 2500 epochs for a VMBF network with 9 basis functions.

The initial point of the learning curves corresponds ot the RMS error immediately following the initialization of the VMBF network parameters. Improvement due to learned basis function positions and shapes is clearly evident in Figure 4. The RMS testing error is only slightly worse than that of the training data, which demonstrate reasonable generalization to novel data. Good generalization is desirable since the training data themselves are somewhat noisy. Figure 5 illustrates two views of the final positions and widths of the nine von Mises basis functions following gradient-descent learning. This particular network was trained on measurements taken from the right ear. Due to the acoustical interaction of the head with the radiating sound pressure wave, most of the variability, hence available information, in the HRTFs occurs when the sound source is on the same side of the head. This asymmetry is reflected in the final positioning of the basis function as shown in Figure 5, and demonstrates the directness with which we can interpret the learned hidden unit weights (positions and widths).

Planar gaussian RBF networks applied to this spherical problem generally perform as well as the VMBF network in regions near the center of the planar projected input space, but perform less well near the edges. As learning proceeds, the centers of the gaussian basis functions move inward due to the artificially absent data beyond the edges. The periodic VMBF network is immune to this bias due to its intrinsic lack of edges.

[3]HRTFs were obtained from Drs. Wightman and Kistler, Hearing Development Research Laboratory, University of Wisconsin—Madison.

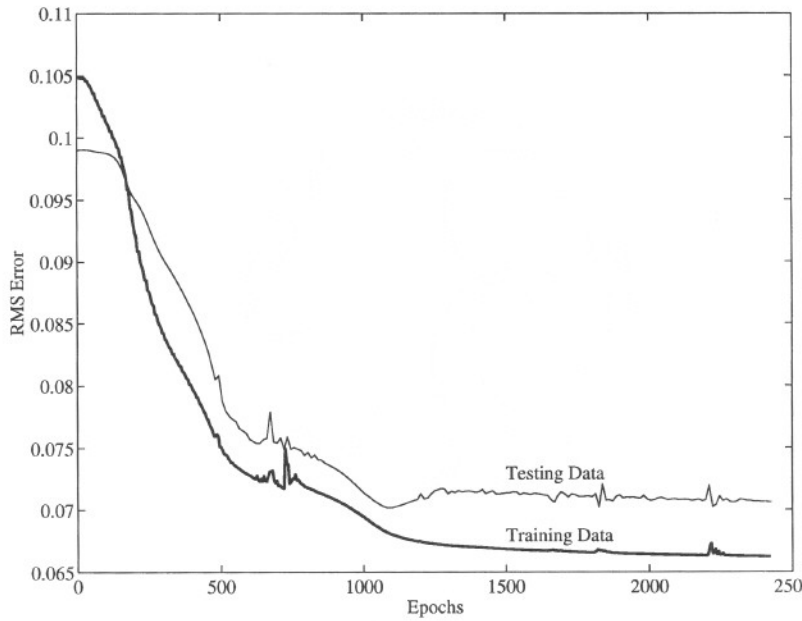

Figure 4: Root-mean-squared error for training and testing data as a function of training epochs.

To demonstrate this immunity, 40 input–output pairs were systematically split off from the original database of 450 in the region near an arbitrarily defined edge ($\pm 180^\circ$ azimuth) for use as novel probe data (in contrast with a distribution of randomly selected test data.) The average magnitudes of the probe data set and the training data set were equalized. The top panel of Figure 6 shows the progressive decline in RMS error as training progresses for a gaussian RBF network and a VMBF network with the remaining 410 input–output pairs. Both networks have nine basis functions and the appropriate gradient-descent update rules. The bottom panel shows the initial decline in RMS error of the testing data. As the gaussian RBF network learns the training set, the RMS error of the probe data rises dramatically . In contrast, the VMBF network training generalizes well to the probe data due to its intrinsic periodicity. For this particular set of novel probe data, the gaussian RBF network must extrapolate beyond the artificial edge of the training data, while the VMBF network performs a spherically constrained interpolation.

Figure 7 illustrates the principal component surfaces approximated by

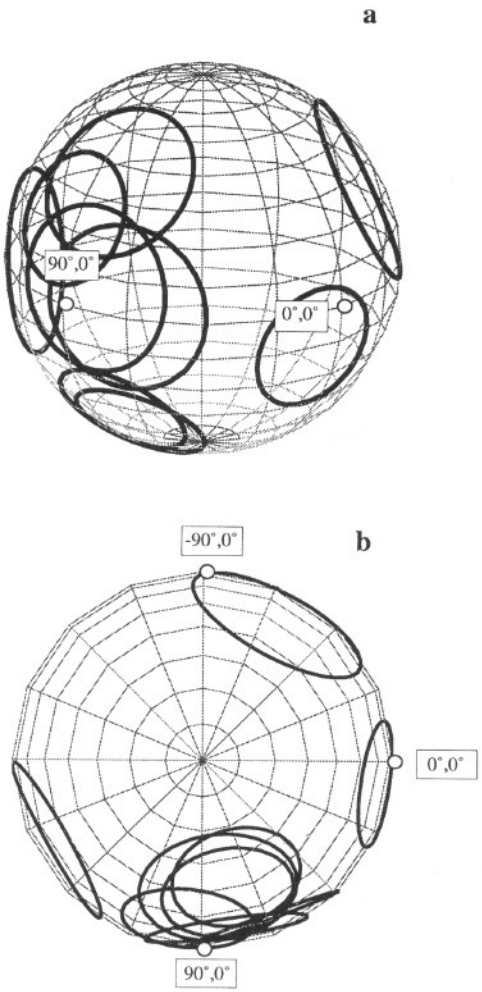

Figure 5: Positions and relative widths of the von Mises basis functions from a 9-basis function VMBF network shown from two viewpoints: (a) 45° to the right of the median line and (b) directly overhead. The displayed width of each basis function is determined by a cross-sectional cut 25% down from the peak of each basis function, which by definition is located at the centroid.

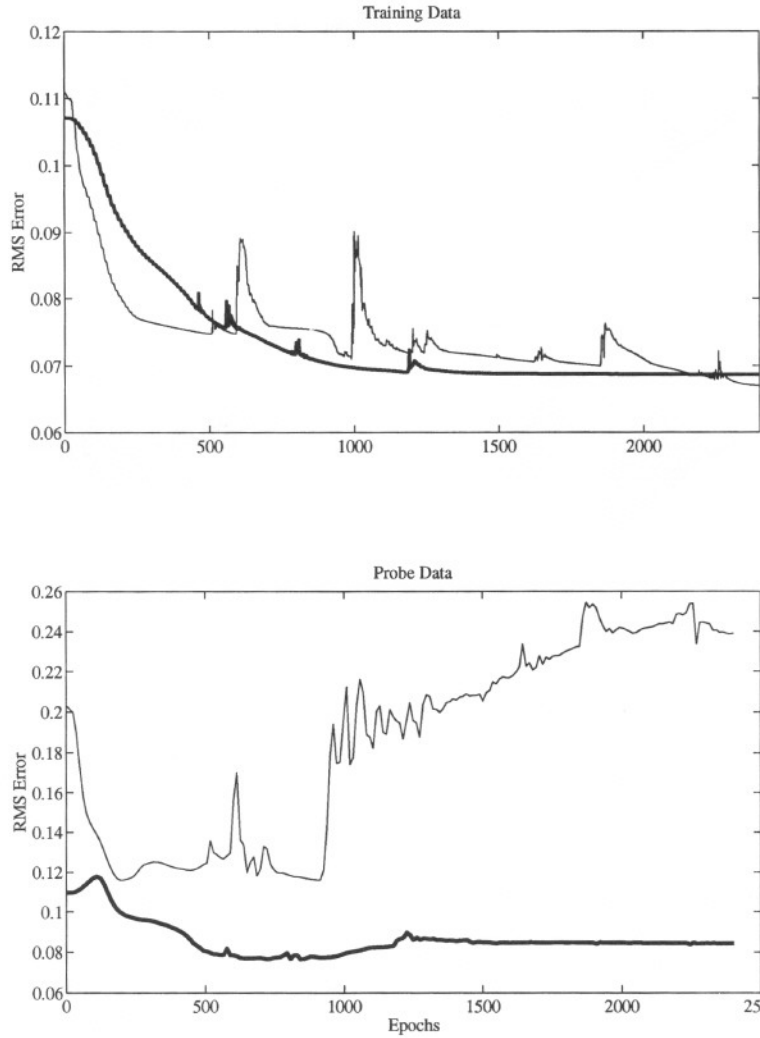

Figure 6: Comparison of gaussian RBF network and VMBF network generalization to novel probe data selected in the region near an arbitrary edge ($\pm 180^{\circ}$ azimuth). Gaussian RBF error is denoted by the fine line and VMBF error is denoted by the bold line.

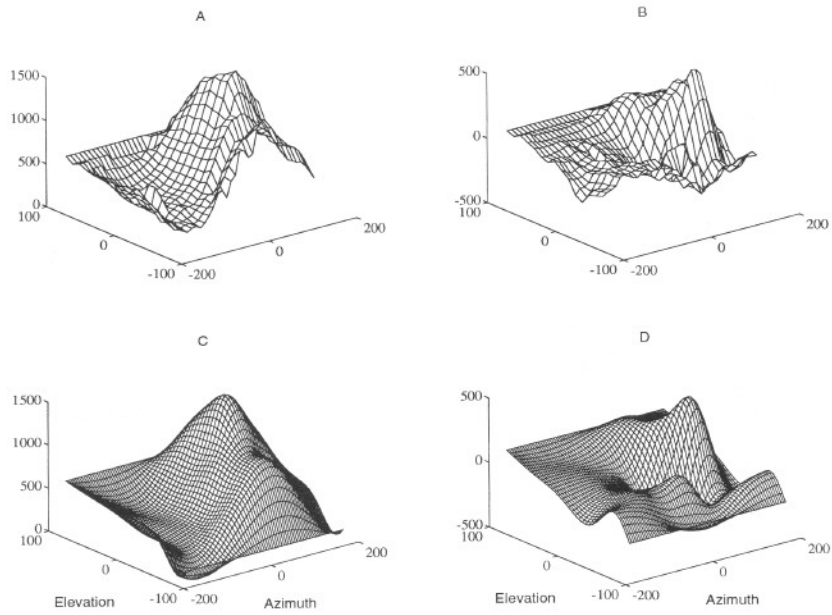

Figure 7: Final approximation surfaces for a 25-basis function VMBF. (A, B) The database of HRTF principal components I (A) and II (B) as a function of direction (azimuth and elevation) in steps of 10°. The actual network contained 6 output units (i.e., four more in addition to the 2 shown) corresponding to the 6 principal components derived from the total 450 256-dimensional HRTFs. (C, D) The results of VMBF network training. The predicted principal components I (C) and II (D) are shown in increments of 5°.

a 25-basis function VMBF network. Figures 7A and 7B show the known first and second principal components (I and II), respectively, as a function of azimuth and elevation location, which together account for 93% of the total variance of the measured 450 HRTFs. Note that measurements were not taken below −50° elevation due to technical constraints in obtaining those samples. Figures 7C and 7D show the principal components predicted by the VMBF network as a function of spherical input for the entire range of positions. The data are plotted on a two-dimensional Cartesian grid with edges, rather than on the more difficult to visualize spherical grid. Therefore, the graphic representation of samples near the poles will naturally distort in the fashion well-known to cartographers. For example, all of the samples at −90° (south pole) are the same measurement regardless of azimuth, hence an isomorphic line of output from

the network should, and does, occur at this elevation. This singularity can be observed in Figures 7C and 7D. Due to the spherical topology of the basis function, this constraint is fundamental to the VMBF network; in contrast, the gaussian RBF network operating in Cartesian space would not enforce this constraint. Smoothing of the measured data as a consequence of the well-trained approximation function is also evident in Figures 7C and 7D.

6 Conclusions

The model of auditory space represents a good example of tailoring a particular neural network architecture (or basis function) to the appropriate input representation, in this case a spherical representation. Because of the periodic nature of the spherical basis, there are no edge effects that arise when using the multidimensional gaussian for approximation in Cartesian space. Well-behaved networks are obtained as a result of this spherical constraint.

Work is ongoing to better characterize the hidden-layer subspace and learning dynamics. These algorithms currently serve as the foundation for ongoing research into implementing auditory virtual environments. We are particularly interested in how a human listener could be integrated into the network learning loop for individual tuning of an auditory space model trained on another person's set of HRTFs, thereby eliminating the need for technically demanding acoustic measurements.

Acknowledgments

This work was supported in part by the Wisconsin Alumni Research Foundation and the Office of Naval Research. We greatly appreciate the database of HRTFs provided by Fred Wightman and Doris Kistler.

References

Arnold, K. J. 1941. On spherical probability distributions. Unpublished Ph.D. Thesis, Massachusetts Institute of Technology.

Broomhead, D. S., and Lowe, D. 1988. Multivariable functional interpolation and adaptive networks. *Complex Syst.* **2**, 321–355.

Chen, J., Van Veen, B. D., and Hecox, K. E. 1993. Synthesis of 3D virtual auditory space via a spatial feature extraction and regularization model. In *IEEE Virtual Reality Annu. Int. Symp.* 188–193.

Chiang, C. H. 1988. *Kinematics of Spherical Mechanisms.* Cambridge University Press, Cambridge.

Fisher, N. I., Lewis, T., and Embleton, B. J. J. 1987. *Statistical Analysis of Spherical Data.* Cambridge University Press, Cambridge.

Ghil, M., Cohn, S., Tavantzis, J., Bube, K., and Isaacson, E. 1981. Applications of estimation theory to numerical weather prediction. In *Dynamic Meteorology: Data Assimilation Methods*, L. Bengtsson, M. Ghil, and E. Kallen, eds., pp. 139–284. Springer-Verlag, New York.

Hartman, E. J., and Keeler, J. D. 1991. Predicting the future: Advantages of semilocal units. *Neural Comp.* **3**, 566–578.

Kistler, D. J., and Wightman, F. L. 1992. A model of head-related transfer functions based on principal component analysis and minimum phase reconstruction. *J. Acoust. Soc. Am.* **91**, 1637–1647.

Mardia, K. V. 1972. *Statistics of Directional Data.* Academic Press, London.

Middlebrooks, J. C., Makous, J. C., and Green, D. M. 1989. Directional sensitivity of sound-pressure levels in the human ear canal. *J. Acoust. Soc. Am.* **86**, 89–108.

Moody, J., and Darken, C. 1988. Learning with localized receptive fields. In *Connectionist Models Summer School* 1–11.

Moody, J., and Darken, C. J. 1989. Fast learning in networks of locally-tuned processing units. *Neural Comp.* **1**, 281–294.

Poggio, T., and Girosi, F. 1990. Networks for approximation and learning. *Proc. IEEE* **78**, 1481–1496.

Powell, M. J. D. 1987. Radial basis functions for multivariable interpolation: A review. In *Algorithms for Approximation*, J. C. Mason and M. G. Cox, eds., pp. 143–166. Clarendon Press, Oxford.

Pralong, D., and Carlile, S. 1994. Measuring the human head-related transfer functions: A novel method for the construction and calibration of a miniature "in-ear" recording system. *J. Acoust. Soc. Am.* **95**, 3435–3444.

Wahba, G., and Wendelberger, J. 1980. Some new mathematical methods for variational objective analysis using splines and cross-validation. *Monthly Weather Rev.* **108**, 1122–1145.

Wahba, G. 1981. Spline interpolation and smoothing on the sphere. *SIAM J. Sci. Stat. Comput.* **2**, 5–16.

Wightman, F. L., and Kistler, D. J. 1989. Headphone simulation of free-field listening I: Stimulus synthesis. *J. Acoust. Soc. Am.* **85**, 858–867.

Received July 29, 1994; accepted March 20, 1995.