A Spherical Basis Function Neural Network for Pole-Zero Modeling of Head-Related Transfer Functions

Rick L. Jenison

Department of Psychology

University of Wisconsin, Madison WI 53706

ABSTRACT

This paper describes a neural network for approximating the parameters of a pole-zero model of the head-related transfer function (HRTF). The von Mises basis function (VMBF) is described, whose response depends on spherical rather than Cartesian input coordinates. 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. A method for computing the common poles of a set of HRTFs is also discussed.

1. INTRODUCTION

Artificial neural networks and approximation techniques typically have been applied to problems conforming to a multidimensional Cartesian input space. However, approximation problems that exist on a unit sphere are most accurately framed within a spherical or polar coordinate system, rather than projected implicitly onto a two-dimensional Cartesian system. Spherical data arise in many areas of science such as geophysics and meteorology, and projection geometry has been the focus of cartographic interest for problems involving a transformation of the spherical earth into a planar map. A well-known problem to cartographers is that area, direction, and shape cannot be preserved simultaneously in these projections, hence some degree of spatial distortion must be incurred. In this paper a neural network for approximating parameters of a pole-zero model of the head-related transfer function (HRTF) is presented. Since the HRTF is primarily directional-dependent, it is most suited to spherical, rather than Cartesian, approximation/interpolation techniques.

2. VON MISES BASIS FUNCTION (VMBF) NEURAL NETWORK

The now well-known Radial Basis Function (RBF) neural network can be viewed as expanding an approximate input-output mapping (or surface) into a linear combination of weighted nonlinear basis functions (see Poggio and Girosi, 1990). We have extended the use of planar radially-symmetric functions to that of a spherical rotationally-symmetric function (Jenison, 1995; Jenison and Fissell, 1995ab). This basis function is based on a spherical probability density function(p.d.f) 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 is commonly referred to as either the von Mises-Arnold-Fisher or Fisher distribution (see Fisher, Lewis, and Embleton (1987)). The expression for the von Mises basis function, dropping the constant of proportionality and elevational weighting factor from the p.d.f., is

$$VM(\theta, \phi, \alpha, \beta, \kappa) = e^{\kappa\left(\sin\phi \sin\beta \cos(\theta-\alpha) + \cos\phi\cos\beta\right)} \quad (1)$$

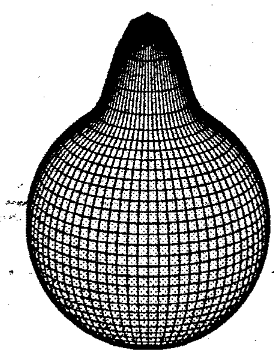

Figure 1. von Mises Basis Function on the unit sphere.

where the input parameters correspond to a sample direction 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). 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 problems on the sphere (e.g. thin-plate pseudo-splines (Wahba, 1981)), the VMBF serves as a spherical analogue of the well-known multidimensional Gaussian on a plane. It resembles a bump on a sphere and behaves in a similar fashion to the planar Gaussian with the centroid corresponding to the mean and $1/\kappa$ corresponding to the standard deviation (see Fig. 1). 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 local resolution optimally.

The architecture of the VMBF neural network (Fig. 2) is constructed as a weighted linear expansion of nonlinear basis functions. The approximation function $f_i(\theta,\phi)$, when spherical coordinates are presented as input, is given by

$$f_i(\theta,\phi) = \sum_{j}^{J} w_{ij} \ VM(\theta,\phi,\alpha_j,\beta_j,\kappa_j) \quad (2)$$

where $VM(\theta,\phi,\alpha_j,\beta_j,\kappa_j)$ is the output of the jth von Mises basis function and w_{ij} is the weight on the jth basis function. For a multivariate output, $f_i(\theta,\phi)$ is the ith element in the output vector.

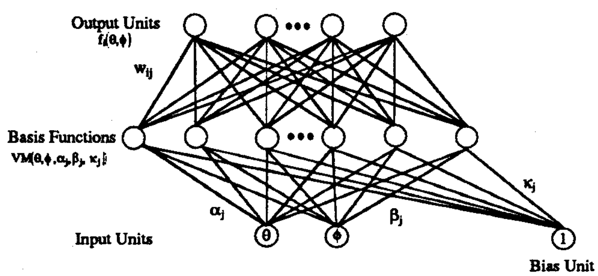

Figure 2. VMBF Neural Network Architecture

2.1. VMBF Neural Network Parameter Learning

The approximation parameters (α, β, κ, and w) for a fixed number of basis functions can be adaptively learned by applying a gradient-descent method to the desired cost-function for the training set of P input-output pairs. For this case we require the sum-of-squared-error to be minimized. The gradient-descent equations are straightforward and are derived in Jenison and Fissell (1995ab). The expression of the cost-function for the pth M dimensional the teaching pattern is

$$E_p = \frac{1}{2} \sum_{i}^{M} (t_{ip} - f_{ip}(\theta,\phi))^2 \quad (3)$$

where t_{ip} is the ith element of the pth teaching pattern and $f_{ip}(\theta,\phi)$ is the pth approximated pattern. Parameter values are learned through successive presentation of P input-output teaching pairs and application of the specific update rules for each parameter of the von Mises basis function. The amount of change made to each parameter following successive presentation is based on the negative derivative of the error with respect to that parameter, which was derived analytically using the chain rule for partial differential equations. Over the course of iterative training, the centroid of each basis function will move on the surface of the sphere, and the concentration (width) of the basis function will change.

3. APPROXIMATING HRTF ZEROS WITH COMMON POLES

From both a practical as well as a theoretical standpoint, the measured HRTFs are of much higher dimensionality in the frequency domain than necessary for modeling purposes. The redundancy inherent in the measurement can be statistically removed via the technique of principal component analysis (Kistler and Wightman, 1992) or Karhunen-Loeve expansion (Chen et al., 1995).

In this paper, the dimensionality of the HRTF impulse response is reduced in the time domain using the well-known Prony method for the synthesis of recursive digital filters (or ARMA models). Others have also applied frequency domain techniques to pole-zero analysis/synthesis of HRTFs (Blommer and Wakefield, 1994). The general form of the Prony method is as follows (see Burrus and Parks, 1970; Parks and Burrus, 1987). The recursive filter in terms of the z transform is given by

$$H(z) = \frac{B(z)}{A(z)} = \frac{b_0 + b_1 z^{-1} + \dots + b_M z^{-M}}{1 + a_1 z^{-1} + \dots + a_M z^{-N}} \quad (4)$$

and the impulse response h(n) is related to H(z) by

$$H(z) = \sum_{n=0}^{\infty} h(n)\, z^{-n}. \quad (5)$$

Here, h(n) is the truncated impulse response of each HRTF as a function of spatial direction. Eq. 4 can then be written as the z transform version of convolution

$$B(z) = H(z)\, A(z) \quad (6)$$

This equation can be decoupled for the purpose of implementing a step-wise solution of the polynomial coefficients. First, the coefficients in the denominator are obtained, whose roots correspond to the poles of the transfer function. The partitioned convolution matrix becomes

$$\begin{bmatrix} \mathbf{b} \\ --- \\ 0 \end{bmatrix} = \begin{bmatrix} H_1 \\ ----- \\ \mathbf{h}_1 \mid H_2 \end{bmatrix} \begin{bmatrix} 1 \\ --- \\ \mathbf{a} \end{bmatrix} \quad (7)$$

Addressing just the lower partition gives

$$\mathbf{h}_1 = -H_2\, \mathbf{a} \quad (8)$$

affording a least-squares solution of **a** using the pseudo-inverse

$$-H_2^{+}\mathbf{h}_1 = \mathbf{a} \quad (9)$$

The coefficients of the denominator are then directly computed using the convolution matrix H_1 from the upper partition

$$\mathbf{b} = H_1 \begin{bmatrix} 1 \\ \mathbf{a} \end{bmatrix}. \tag{10}$$

In this analysis, 505 HRTFs recorded from one ear of a single human subject are represented by impulse responses with 256 coefficients (HRTFs were obtained from Drs. Wightman and Kistler, UW-Madison). These measurements, corresponding to discrete directions, ranged from -170° to 180° in azimuth and -50° to 90° in elevation in 10° steps. The impulse responses are minimum-phase, i.e. the group-delays, due to acoustic path delay, have been removed. After examining various orders of zero-pole models, it was observed that the transfer function poles remain relatively static across measured HRTFs from a single individual. Recently, Haneda, Makino, and Kaneda (1994) have modeled Room Transfer Functions using a common set of acoustical poles. Consistent with the observation of static poles, the general Prony least-squares solution of polynomial coefficients can easily be modified to compute a set of common poles based on the entire set of minimum-phase HRTFs, where now

$$H_2 = \begin{bmatrix} H_2(d_1) \\ \vdots \\ H_2(d_P) \end{bmatrix} \tag{11}$$

and

$$\mathbf{h}_1 = \begin{bmatrix} \mathbf{h}_1(d_1) \\ \vdots \\ \mathbf{h}_1(d_P) \end{bmatrix} \tag{12}$$

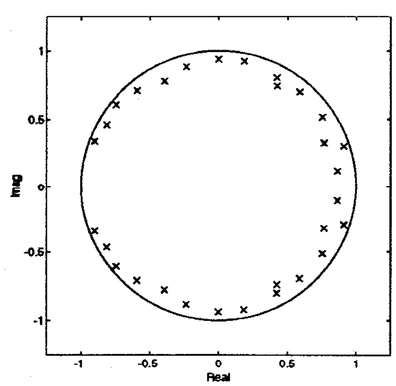

Figure 3. z-plane plot of 30 common poles for 505 HRTFs

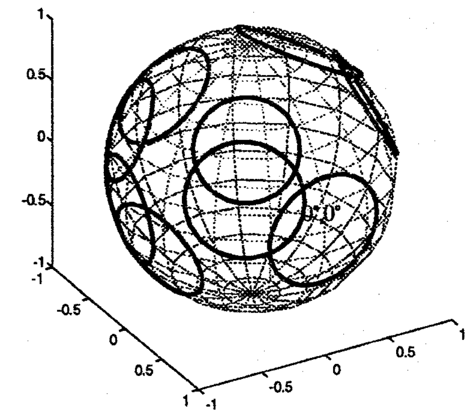

Figure 4. Placements and contours of the nine basis functions following gradient-descent training. Positive azimuths are to the the left of 0°.

where d_p denotes a spatial direction (θ, ϕ). The single set of denominator coefficients (or poles) are now common, in a least-squares sense, to all HRTFs. From this set of common poles, 505 sets of numerator coefficients **b** are generated using Eq. 10.

The VMBF neural network can now be "trained" to approximate the functional mapping of spatial direction in spherical coordinates to the numerator polynomial coefficients. To demonstrate the technique, 30 common poles were first generated using the least-squares Prony technique on the 505 HRTFs (result shown in Fig. 3). Then, 505 sets of eight numerator coefficients **b** were generated, using the second step of the Prony technique, for use as teaching patterns for training the VMBF neural network. A VMBF network with nine basis functions provides a reasonable approximation to the input-output mapping. Hence, the VMBF architecture illustrated in Fig. 2 has nine basis functions in the middle layer and 8 units in the output layer. The number of input-output patterns P would be 505.

The positions (α and β) and contours of the nine VMBFs following training to asymptotic performance (least-squared-error) are shown projected on unit sphere in Fig. 4. The initial conditions of the network at the onset of training is a uniform placement of the basis functions around the unit sphere. The resulting approximations to the first and third polynomial coefficients as a function of azimuth and elevation are shown in Figs. 5 and 6. A comparison of the magnitude spectrum from a selected location (40° azimuth, 60° elevation) for the measured impulse response (bold line) and the VMBF approximated response (fine line) is shown in Fig. 7.

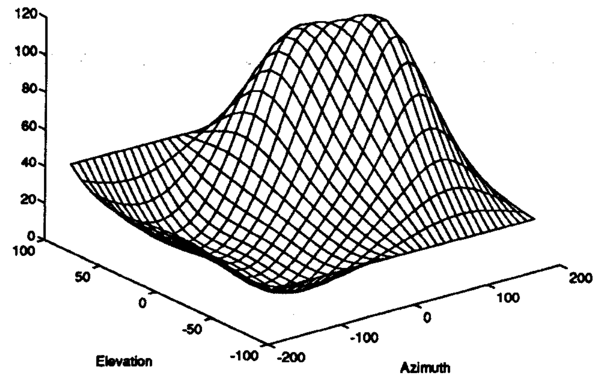

Figure 5. VMBF approximation of the first coefficient b(0) as a function of azimuth and elevation

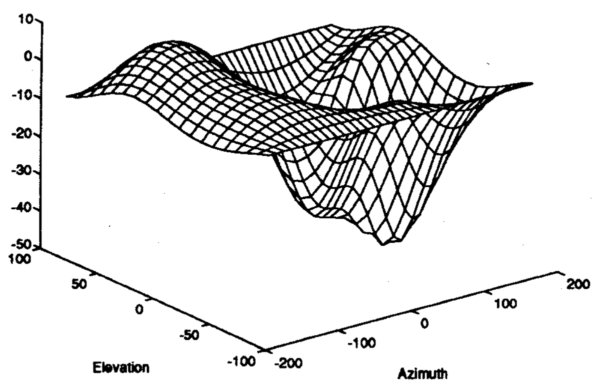

Figure 6. VMBF approximation of the third coefficient b(2) as a function of azimuth and elevation

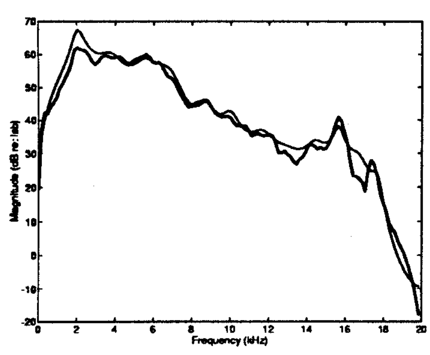

Figure 7. Magnitude spectrum for HRTF at 40° azimuth 60° elevation. Measured (bold line) and VMBF pole-zero approximated (fine line).

4. CONCLUSIONS

The VMBF neural network has been presented for approximating a functional mapping from spherical coordinates to the parameters of a pole-zero model that reduces the dimensionality of a measured set of HRTFs. The pole-zero model serves as an alternative to other techniques for dimensionality reduction such as principal component analysis. The advantage of the pole-zero model is that it can be directly implemented as a recursive filter without the intermediate step of eigenvector inverse-transformation for reconstructing nonrecursive filters. Furthermore, the common pole model described in this paper requires no update of the cascade of feedback coefficients as a function of change in sound source direction.

Acknowledgments

I'd like to thank Barry Van Veen for bringing to my attention the work on common poles of room transfer functions. The database of HRTFs provided by Fred Wightman and Doris Kistler is greatly appreciated.

REFERENCES

1. Blommer, M. A. and Wakefield, G. H. "On the design of pole-zero approximations using a logarithmic error measure," IEEE Trans. on Sig. Proc. 42, 3245-3248, 1994.
2. Burrus, C. S. and Parks, T. W. "Time domain design of recursive digital filters," IEEE Trans. Audio and Electro. 18, 137-141, 1970.
3. Chen, J., Van Veen, B. D. and Hecox, K. E. "A spatial feature extraction and regularization model for the head-related transfer function," J. Acoust. Soc. Am. 97, 439-452, 1995.
4. Fisher, N. I., Lewis, T. and Embleton, B. J. J. *Statistical Analysis of Spherical Data*, (Cambridge University Press, Cambridge), 1987.
5. Haneda, Y., Makino, S. and Kaneda, Y. "Common acoustical pole and zero modeling of room transfer functions," IEEE Trans. Speech Audio Proc. 2, 320-328, 1994.
6. Jenison, R. L. "von Mises Basis Function neural networks for spherical approximation," World Congress on Neural Networks, 1995.
7. Jenison, R. L. and Fissell, K. "A spherical basis function neural network for modeling auditory space," Neural Computation, in press.
8. Jenison, R. L. and Fissell, K. "A comparison of Gaussian and von Mises Basis Functions for approximating spherical acoustic scatter," IEEE Trans. Neural Networks, in press.
9. Kistler, D. J. and Wightman, F. L. "A model of head-related transfer functions based on principal component analysis and minimum phase reconstruction," J. Acous. Soc. Am. 91, 1637-1647, 1992.
10. Parks, T. W. and Burrus, C. S. Digital Filter Design, (Wiley, New York), 1987.
11. Poggio, T. and Girosi, F. "Networks for approximation and learning," Proc. IEEE 78, 1481-1496, 1990.
12. Wahba, G. "Spline interpolation and smoothing on the sphere," SIAM J. Sci. Stat. Comput., 2, 5-16, 1981.