INTERPOLATION OF HEAD-RELATED TRANSFER FUNCTIONS (HRTFS): A MULTI-SOURCE APPROACH

Fábio P. Freeland, Luiz W. P. Biscainho, Paulo S. R. Diniz

PEE/COPPE/UFRJ, LPS-DEL/POLI
Caixa Postal 68504 - 21945-970 Rio de Janeiro, RJ – Brazil
{freeland, wagner, diniz}@lps.ufrj.br

ABSTRACT

This paper proposes a new method for the interpolation of Head-Related Transfer Functions (HRTFs) applied to the generation of 3-D binaural sound, especially when dealing with moving sound sources indoors. The method combines a modified linear interpolation strategy with a representation of the auditory space based on spatial characteristic functions (SCFs), previously known from the literature. The main idea here is to associate the low complexity the SCF-based representation yields in the multi-source case with the inherent simplicity of the linear interpolation. Complexity issues are discussed. The performance of the proposed method is evaluated against the direct bilinear interpolation of HRTFs, using Spatial Frequency Response Surfaces (SFRSs).

1. INTRODUCTION

The three-dimensional sound generation through head-phones can be achieved by filtering a monaural sound by Head-Related Transfer Functions (HRTFs). These functions describe the paths between a sound source and each ear of the listener.

The placement of a virtual sound source at an arbitrary position around the listener requires an adequate interpolation procedure, since those functions have usually been measured only for a finite set of positions [1]. There are several ways to perform interpolation [2, 3, 4, 5], one of the most popular being the direct bilinear interpolation of HRTFs.

On the other hand, for multiple sound sources, which may include the modeling of early reverberation through image sound sources, some alternative representation strategies have been devised to reduce the computational complexity of 3-D sound systems [6, 7, 8]. One of them is the so-called Spatial Feature Extraction and Regularization (SFER) model, a continuous representation of the auditory space based on Spatial Characteristic Functions (SCFs).

The present work proposes the application of a modified linear interpolation method to the SCFs, aiming at further reduction of the computational requirements of the SFER method.

The next section reviews the bilinear interpolation method and describes its triangular version. Then, in Section 3, after the KLT-based Spatial Feature Extraction is reviewed, the proposed method for interpolation of the SCFs is described, and complexity issues are addressed. In Section 4, the results attained by the procedure described in Section 3 are compared with those of the direct bilinear interpolation of HRTFs. Section 5 presents the conclusions.

2. BILINEAR INTERPOLATION

Given a known set of measured HRTFs, or their time counterparts, the HRIRs (Head-Related Impulse Responses), a simple way to interpolate non-measured functions is the bilinear [3, 4] method. The HRIR associated to a desired position is computed by weighting the contributions of the HRIRs associated to the nearer surrounding positions with measured HRIRs, as Figure 1 illustrates.

The authors wish to thank CNPq, a Brazilian research council, for supporting this work.

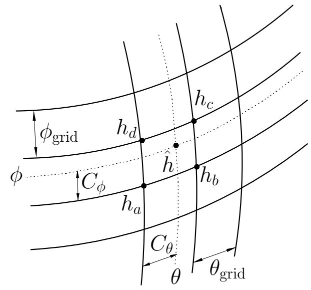

Figure 1: Bilinear interpolation.

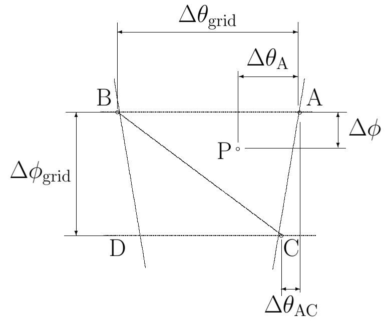

Figure 2: Detail of triangular regions over the reference sphere.

Assume that all measured HRIRs refer to points located over a reference sphere centered in the middle point between the ears, along fixed angular steps $\theta_{\rm grid}$ and $\phi_{\rm grid}$ for azimuth and elevation, respectively. Referring to Figure 1, the approximation for the HRIR $h(k)$ associated to the desired point would be

$$\begin{aligned}\hat{h}(k) \quad &= \quad (1-c_\theta)(1-c_\phi)h_a(k)+c_\theta(1-c_\phi)h_b(k)+\\ &\quad\quad c_\theta c_\phi h_c(k)+(1-c_\theta)c_\phi h_d(k),\end{aligned}$$

where $h_a(k)$, $h_b(k)$, $h_c(k)$ and $h_d(k)$ are the HRIRs for the nearer positions related to the desired one. The coefficients c_θ and c_ϕ can be calculated as follows:

$$c_\theta = \frac{C_\theta}{\theta_{\rm grid}} = \frac{\theta \bmod \theta_{\rm grid}}{\theta_{\rm grid}} \text{ and } c_\phi = \frac{C_\phi}{\phi_{\rm grid}} = \frac{\phi \bmod \phi_{\rm grid}}{\phi_{\rm grid}},$$

where C_θ and C_ϕ are the indicated relative positions.

For practical reasons, a typical set of measured HRIRs (as those measured by Gardner and Martin [1], which are employed along this work) would refer to points more or less homogeneously distributed around the sphere, which implies variable angular grids. A possible generalization of the bilinear method for this situation employs triangular regions [5], as in Figure 2. Now, the interpolation of the HRIR for point P can be performed as

$$\hat{h}_{\rm P}(k) = w_{\rm A}h_{\rm A}(k)+w_{\rm B}h_{\rm B}(k)+w_{\rm C}h_{\rm C}(k), \quad (1)$$

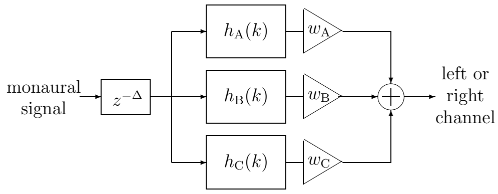

Figure 3: Bilinear interpolation structure.

where

$$w_C = \frac{\Delta\phi}{\Delta\phi_{\text{grid}}}, \quad w_B = \frac{1}{\Delta\theta_{\text{grid}}}\left(\Delta\theta_A - w_C \Delta\theta_{AC}\right),$$
$$\text{and} \quad w_A = 1 - w_B - w_C, \tag{2}$$

with the angular distances defined as

$$\Delta\theta_{\text{grid}} = \theta_B - \theta_A, \quad \Delta\theta_A = \theta_P - \theta_A, \quad \Delta\theta_{AC} = \theta_C - \theta_A,$$
$$\Delta\phi_{\text{grid}} = \phi_C - \phi_A, \quad \text{and} \quad \Delta\phi = \phi_P - \phi_A.$$

It is assumed, without loss of generality, that the points A and B have the same elevation.

In order to avoid undesired destructive interference, the interpolation procedure given by equation (1) should be performed on the minimum-phase versions of the measured HRIRs [9]. In fact, the excess of phase of each original HRIR over its minimum-phase counterpart amounts approximately to a pure delay [10]. To preserve the effect of these inherent delays [11], they should be estimated and combined into an interpolated delay Δ.

Figure 3 shows the block diagram of the interpolation procedure described above, for a single channel (left or right) of the binaural system.

3. SCF INTERPOLATION

The HRTFs can be considered as functions of three variables: two spherical coordinates of position (assuming constant radius) and the frequency. The frequency dependence can be separated from the dependence of the angular coordinates azimuth θ and elevation ϕ [6], in such a way that

$$H(\theta,\phi,f) = \sum_{i=1}^{N} \omega_i(\theta,\phi)\Gamma_i(f),$$

where $\omega_i(\theta,\phi)$ and $\Gamma_i(f)$ are the spatial and frequency dependent components, respectively.

In [7, 8], the authors propose the so-called Spatial Feature Extraction and Regularization (SFER) model, that represents the HRTFs via the Karhunen-Loève Transform (KLT) [12], by using its base functions as the component $\Gamma_i(f)$. An analogous procedure can be adopted for the HRIRs, and this is the chosen formulation hereafter.

3.1 Spatial Features Extraction

Considering a matrix $\mathbf{H}$ containing measured order-N HRIRs in its rows, the eigenvectors $[\psi_j(0)\cdots\psi_j(N-1)]^T$ of the HRIR autocovariance matrix $\mathbf{C}$ can be obtained as the columns of Ψ in

$$\mathbf{C} = (\mathbf{H}-\overline{\mathbf{H}})^T(\mathbf{H}-\overline{\mathbf{H}}) = \Psi\Lambda\Psi^T, \quad \text{with} \quad \Psi^T\Psi = \mathbf{I}, \tag{3}$$

where Ψ is the KLT matrix, Λ is the diagonal matrix whose diagonal elements are the eigenvalues of $\mathbf{C}$, and $\overline{\mathbf{H}}$ contains in every row the mean HRIR, $\overline{h}(k)$.

The goal is to find an approximation for the HRIR $h(\theta,\phi,k)$ associated to any point (θ,ϕ) over the reference sphere in the form

$$\hat{h}(\theta,\phi,k) = \overline{h}(k) + \sum_{j=1}^{N} \omega_j(\theta,\phi)\psi_j(k),$$

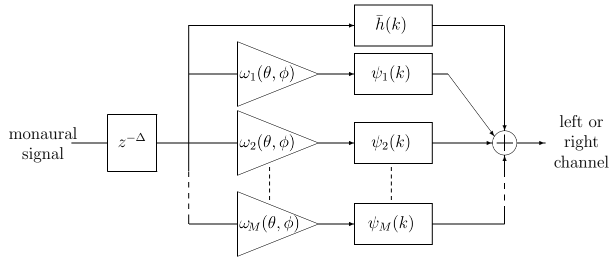

Figure 4: SFER structure.

where $\omega_j(\theta,\phi)$ are the so-called Spatial Characteristic Functions (SCFs). For the previously known HRIRs $h(\tilde{\theta},\tilde{\phi},k)$, measured at points $(\tilde{\theta},\tilde{\phi})$, the KLT coefficients $\omega_j(\tilde{\theta},\tilde{\phi})$ can be determined, thus yielding $\hat{h}(\tilde{\theta},\tilde{\phi},k) = h(\tilde{\theta},\tilde{\phi},k)$. The KLT coefficients related to any other HRIR should be computed from $\omega_j(\tilde{\theta},\tilde{\phi})$.

The use of the KLT allows a significant complexity reduction: Instead of using all eigenvectors found through Equation (3) [7, 8] to represent the HRIRs, one can use only M of them (with $M < N$) by selecting those with larger eigenvalues. This work uses the eigenvectors corresponding to the $M = 32$ largest eigenvalues, which concentrates 99,9% of the total energy.

Figure 4 shows the block diagram of the approximated HRIR for a single channel (left or right) of the binaural system

Notice that the SFER model is expected to yield reduced complexity when compared to the usual bilinear interpolation because in the former only the number of necessary SCFs increases proportionally to the number of sources (the number of $\psi_j(k)$ remains the same), while in the latter the overall calculations must be independently done for each sound source

3.2 Approximation of the SCFs

In [8, 7, 13], the SCFs $\omega_j(\theta,\phi)$ for any position (θ,ϕ) are computed from their known samples $\omega_j(\tilde{\theta},\tilde{\phi})$ through spline interpolation. Despite its good performance, this approach may require a large number of operations for spline evaluation, depending on the spline employed. In the following, a simpler alternative to the spline is proposed.

If a virtual source is inside a triangular region ABC, as in Figure 2, supposing that an initial estimate has been previously computed, the estimation of SCFs is done in an incremental form. The updating equation is given by

$$\omega_j(\theta_l,\phi_l) = \omega_j(\theta_{l-1},\phi_{l-1}) + \Delta\omega_{j,l-1}, \tag{4}$$

where l is the angular position index. The increment applied to weight ω_j from position $l-1$ to position l can be computed as

$$\begin{aligned}\Delta\omega_{j,l-1} \quad = \quad & (\theta_l - \theta_{l-1}) \left.\frac{\partial\omega_j(\theta,\phi)}{\partial\theta}\right|_{\substack{\theta=\theta_{l-1}\\ \phi=\phi_{l-1}}} + \\ & +(\phi_l - \phi_{l-1}) \left.\frac{\partial\omega_j(\theta,\phi)}{\partial\phi}\right|_{\substack{\theta=\theta_{l-1}\\ \phi=\phi_{l-1}}}.\end{aligned}$$

Of course, the partial derivatives cannot be obtained analytically. However, each SCF $\omega_j(\theta,\phi)$ can be approximated by triangular faces whose vertices are its known samples $\omega_j(\tilde{\theta},\tilde{\phi})$. Then, those derivatives can be estimated by the slopes of the face containing the point $(\theta_{l-1},\phi_{l-1})$. Their values should be computed a priori and stored in a look-up table.

One should take care with the cumulative error that might develop along the updating of $\omega_j(\theta,\phi)$. This error can originate mainly when the sound source crosses the boundary between two faces, since the plane derivatives are discontinuous on the boundaries. This problem is easily circumvented simply by recalculating an initial estimate for each new visited region. A possible way to perform this estimation is to apply the bilinear method directly to the SCFs associated to the vertices A, B, and C:

$$\begin{aligned} \omega_j(\theta,\phi) \quad &= \quad w_{\mathrm{A}}\omega_j(\theta_{\mathrm{A}},\phi_{\mathrm{A}})+w_{\mathrm{B}}\omega_j(\theta_{\mathrm{B}},\phi_{\mathrm{B}})+ \\ & \quad\quad w_{\mathrm{C}}\omega_j(\theta_{\mathrm{C}},\phi_{\mathrm{C}}), \end{aligned} \tag{5}$$

where w_{A}, w_{B}, and w_{C} are computed according to Equation (2).

The algorithm for obtaining the SCFs along the virtual sound source path can be summarized as follows:

1. For a new triangular region, compute the first SCFs according to Equation (5).
2. If the next position is inside the same region, update the SCFs through Equation (4).
3. Otherwise, return to the first step.
4. Return to the second step.

The advantage of the incremental method over the spline interpolation, without considering the operations related to the fixed filters shown in Figure 4, which are the same in both cases, can be easily verified. The former needs only $3M$ multiplications plus $5M$ additions per channel per sample per source, while the complexity of the latter is proportional to M times the logarithm of the total number of measured HRIRs [14] (over 700 in the present case) per channel per sample per source.

It can also be shown that if the number of sound sources processed exceeds $\frac{(M+1)N}{3(N-M+2)}$, the proposed incremental method yields lower complexity than the direct bilinear interpolation of HRTFs. For the present case ($N = 128$ and $M = 32$), this limit number equals 16, which, in the context of image sources modeling early reflections, is not too large a number.

4. PERFORMANCE COMPARISON

Figures 5 and 6 provide some comparisons between the direct bilinear interpolation of HRTFs (left columns) and the incremental interpolation of the SCFs (right columns), using the so-called Spacial Frequency Response Surfaces (SFRSs) [15]. Each plot shows the magnitude response of the approximated HRTFs versus angular positions, for a given frequency.

There is no noticeable difference between the SFRSs of the two methods. In order to provide a more precise evaluation, Figure 7 shows a 100-bin histogram for the relative error

$$\xi = 20\log_{10}\left(\frac{|\hat{H}(\theta,\phi,f)|}{|H(\theta,\phi,f)|}\right)$$

computed over a dense set of angles and frequencies (θ, ϕ, f). $|H(\theta,\phi,f)|$ and $|\hat{H}(\theta,\phi,f)|$ are samples of the SFRSs shown in Figures 5 and 6 for the direct bilinear and incremental interpolations, respectively. This histogram indicates that more than 98% of the calculated results exibits an error lower than 0.729 dB, thus confirming that the proposed method is a computationally advantageous alternative for the multi-source implementation of 3-D binaural sound.

5. CONCLUSIONS

The present paper proposed an hybrid way to interpolate HRTFs by applying the principle of bilinear interpolation to the spatial features extracted from a set of HRIRs. It exhibits lower computational complexity than previous methods for the multi-source case, thus being especially suited for modeling moving sound sources indoors. The quantitative comparison between the proposed incremental method and the direct bilinear interpolation of HRTFs indicates very similar performances.

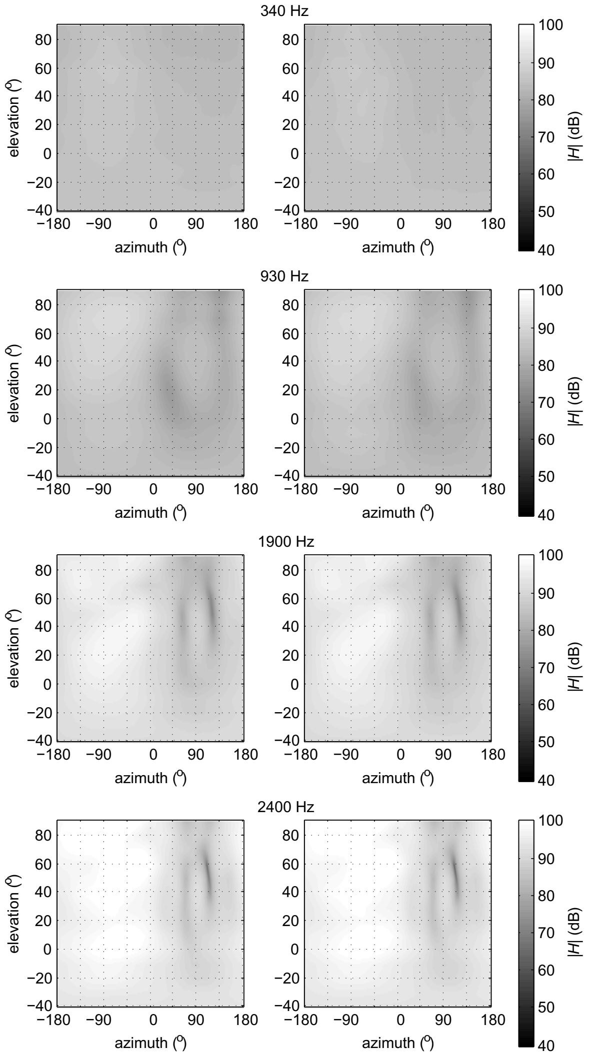

Figure 5: Direct-bilinear (left) versus incremental-SCF (right) method: 340, 930, 1900, 2400 Hz.

REFERENCES

[1] B. Gardner and K. Martin, "HRTF measurements of a KEMAR dummy-head microphone," Technical Report 280, MIT Media Lab., Cambridge, MA, USA, May 1994.

[2] K. Hartung, J. Braasch, and S. J. Sterbing, "Comparison of different methods for the interpolation of head-related transfer function," in *The Proceedings of the AES 16th International Conference*, Rovaniemi, Finland, Apr. 1999, pp. 319–329.

[3] D. R. Begault, *3D Sound for Virtual Reality and Multimedia*, Academic Press, Cambridge, MA, USA, 1994.

[4] L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen, "Creating interactive virtual acoustic environments," *J. Audio Eng. Soc.*, vol. 47, no. 9, pp. 675–705, Sep. 1999.

[5] L. W. P. Biscainho, F. P. Freeland, and P. S. R. Diniz, "Using inter-positional transfer functions in 3D-sound," in *IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Orlando, Florida, USA, May 2002.

[6] J.-M. Jot, S. Wardle, and V. Larcher, "Approaches to binaural synthesis," in *Proceedings of AES 105th Convention*, California, USA, Sep. 1998, AES, preprint 4861.

[7] J. Chen, B. D. Van Veen, and K. E. Hecox, "Synthesis of

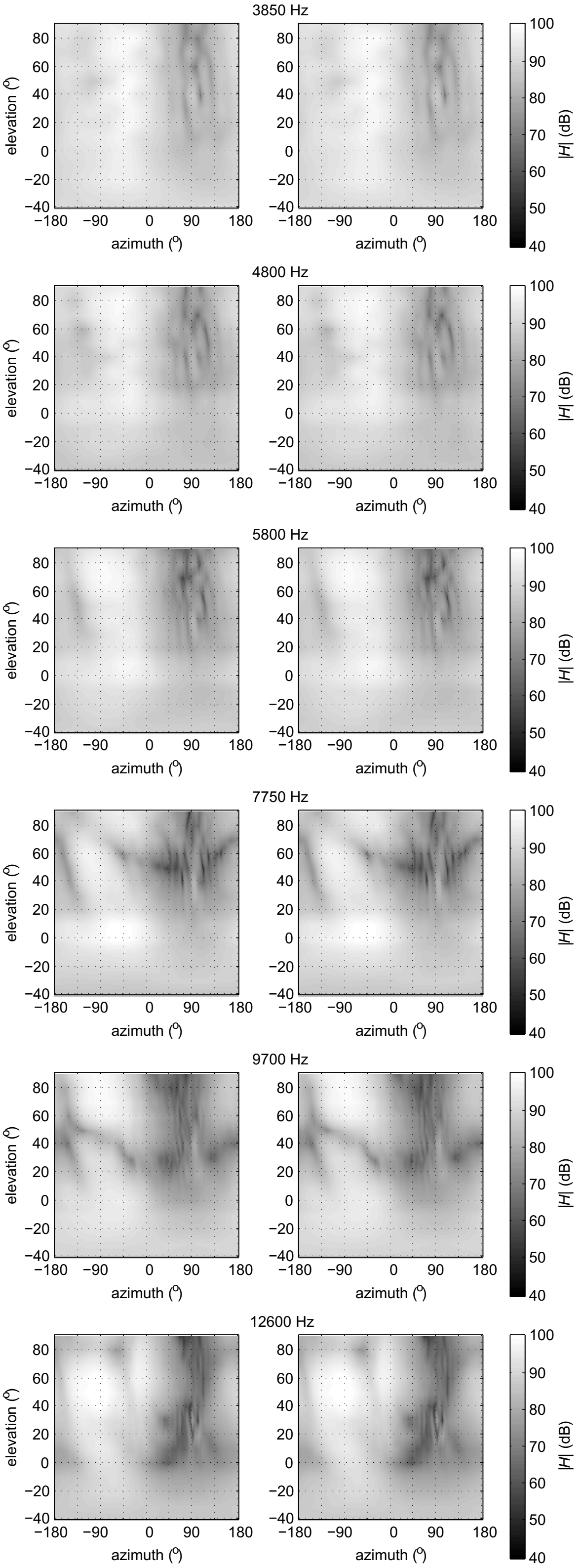

Figure 6: Direct-bilinear (left) versus incremental-SCF (right) method: 3850, 4800, 5800, 7750, 9700 and 12600 Hz.

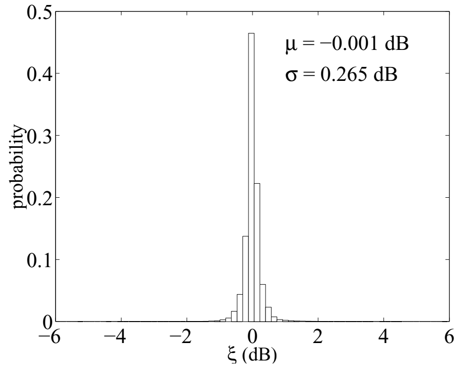

Figure 7: Histogram for the error ξ.

3d virtual auditory space via a spatial feature extraction and regularization model," in *IEEE Virtual Reality Annual International Symposium*, Sep. 1993, pp. 188—193.

[8] J. Chen, B. D. Van Veen, and K. E. Hecox, "A spatial feature extraction and regularization model for virtual auditory display," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Apr. 1993, vol. 1, pp. 129–132.

[9] A. Kulkarni, S. K. Isabelle, and H. S. Colburn, "On the minimum-phase approximation of head-related transfer functions," in *IEEE Proc. WASPAA'1995*, New Paltz, NY, USA, Oct. 1995.

[10] J.-M. Jot, V. Larcher, and O. Warusfel, "Digital signal processing issues in the context of binaural and transaural stereophony," in *Proc. 98th AES Convention*, Paris, France, Feb. 1995, AES, (preprint 3980).

[11] J. Huopaniemi, N. Zacharov, and M. Karjalainen, "Objective and subjective evaluation of head-related transfer function filter design," *J. Audio Eng. Soc.*, vol. 47, no. 4, pp. 218–239, Apr. 1999.

[12] A. K. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall, Englewood Cliffs, NJ, USA, 1989.

[13] T. Nishino, S. Kajita, K. Takeda, and F. Itakura, "Interpolating head related transfer function in the median plane," in *Proc. IEEE WASPAA'1999*, New Paltz, NY, USA, Oct. 1999, pp. 167–170.

[14] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, Cambridge, 2nd edition, 1992.

[15] C. I. Cheng and G. H. Wakefield, "Spatial frequency response surfaces (SFRS'S): An alternative visualization and interpolation thechnique for head-related transfer functions (HRTF's)," in *Proc. AES 16th Int. Conf.*, Rovaniemi, Finland, Apr. 1999, pp. 147–159.