A time domain binaural model based on spatial feature extraction for the head-related transfer function

Zhenyang Wu,[a] Francis H. Y. Chan,[b] and F. K. Lam
Department of Electrical and Electronic Engineering, University of Hong Kong, Pokfulam Road, Hong Kong

Joseph C. K. Chan
Division of Technology, City University of Hong Kong, Tat Chee Ave, Hong Kong

(Received 18 July 1996; revised 9 June 1997; accepted 11 June 1997)

A complex-valued head-related transfer function (HRTF) can be represented as a real-valued head-related impulse response (HRIR). The interaural time and level cues of HRIRs are extracted to derive the binaural model and also to normalize each measured HRIR. Using the Karhunen–Loeve expansion, normalized HRIRs are modeled as a weighted combination of a set of basis functions in a low-dimensional subspace. The basis functions and the space samples of the weights are obtained from the measured HRIR. A simple linear interpolation algorithm is employed to obtain the modeled binaural HRIRs. The modeled HRIRs are nearly identical to the measured HRIRs from an anesthetized live cat. Typical mean-square errors and cross-correlation coefficients between the 1816 measured and modeled HRIRs are 1% and 0.99, respectively. The real-valued operations and linear interpolating in the model are very effective for speeding up the model computation in real-time implementation. This approach has made it possible to simulate real free-field signals at the two eardrums of a cat via earphones and to study the neuronal responses to such a virtual acoustic space (VAR). © *1997 Acoustical Society of America.* [S0001-4966(97)02410-7]

PACS numbers: 43.64.Bt, 43.66.Ba, 43.64.Ha, 43.66.Qp [RDF]

INTRODUCTION

Sound arriving at listener's eardrums in a free-field sound signal varies with source direction (Shaw, 1974; Blauert, 1983). A direction-dependent transformation is referred to as a head-related transfer function (HRTF) to acknowledge the primary acoustical importance of head and pinnas. Each HRTF pair obtained in the two ears contains interaural cues in time and level, and spectral cues. The spectral cues of the HRTF are characterized by peaks and notches, some of which vary systematically with changes in sound source direction. New approaches based on the HRTFs have enabled researchers to synthesize directional stimuli for headphone presentation using measured HRTFs (Wightman and Kistler, 1989; Chan *et al.*, 1993; Brugge *et al.*, 1992; Wenzel, 1992). In these approaches, the stimulus waveforms appropriate for the right and left eardrums were computed using the pair of HRTFs corresponding to a sound-source direction and then delivered though compensated headphones. In a physiological study, we need more sophisticated stimuli for earphone presentation to understand how the neural system might use directional cues. Those allow complete and independent control of localization cues in synthesizing virtual auditory signals, that would be difficult to obtain with free-field stimuli, and present virtual auditory space to the subject with continuous spatial resolution.

Functional representations of the HRTFs were proposed by several authors (Batteau, 1967; Genuit, 1986; Chen *et al.*, 1992). They sought a mathematical model or equation that represents the HRTF as a function of frequency and direction, such that the models could provide explicit mathematical relationship between the HRTF and source location. One direction of research, for example, was based on Batteau's pinna-simulating model that saw the HRTF as being synthesized by a delay-and-add system with one direct path and two delayed paths. Functional approaches would reduce the storage requirement and represent the HRTF at an arbitrary direction. However, attempts to describe the HRTF in a simple mathematical equation have been of only limited success. Also, to maintain model accuracy it would be necessary to determine the delay time and the attenuation for the delayed path. Application of the functional model is thus limited to representation of external ear characteristics over whole auditory space.

Low-dimensional and orthogonal representations for sets of HRTFs have been generated by using the Karhunen–Loeve expansion of HRTFs, and employing an essentially identical mathematical operation, Kistler and Wightman (1992) established a model based on principal components analysis and minimum-phase reconstruction. They applied principal components analysis (PCA) to the logarithms of the HRTF magnitudes after the removal of direction-independent and subject-dependent spectral features. From the PCA representation, HRTFs for each direction were reconstructed using phase estimates based on the minimum phase characteristic corresponding to each magnitude function. These studies are available only for the measured directions. More recently, a spatial feature extraction and regularization model for the HRTFs was proposed (Chen *et al.*, 1995). In this model the HRTFs were expressed as weighted

[a]Current address: Department of Radio Engineering, Southeast University, Nanjing 210018, People's Republic of China.
[b]Electronic mail: fhychan@eee.hku.hk

combinations of a set of complex valued eigentransfer functions. The sample weights are determined by projecting all measured HRTFs onto the eigentransfer functions. A functional representation for weights is obtained by applying a thin plate generalized spline smoothing model to regularize the sample weights. This approach maintains the phase of the spectral components and model accuracy at a whole upper 3/4 sphere but deals with large amounts of complex valued computation in matrix-vector products and two-dimensional splines.

Both the PCA model and the spatial feature extraction and regularization model have focused on the frequency components. The time domain and frequency domain behaviors are related through the Fourier transform and are equivalent in their carrying of directional information. In the time domain, we can describe the HRTF as head-related impulse responses (HRIR). The HRIR, unlike HRTF, is a real-valued time series. The HRIR model need only deal with real valued computation. Using the HRIR model the actual phase information is maintained instead of having to make minimum phase approximations, and the extensive computation with complex-valued eigenfunctions is also avoided. The eardrum stimuli can be obtained by directly convolving the HRIR with the free-field signals in the time domain.

In this paper, a binaural HRIR model for the implementation of virtual acoustic space is developed. In this model the interaural time and level cues are extracted from the measured HRIRs of a cat. The HRIRs are normalized by removing the time differences and level differences. Karhunen–Loeve expansion is used to represent the normalized HRIR in a low-dimensional space. The normalized HRIRs, which contain the spectral cues, are expressed as weighted combinations of a set of basis functions. The basis functions are real-valued eigenvectors that are derived from a covariance matrix of the measured HRIRs. The weights, applied to each basis function and termed real spatial characteristic functions (RSCFs), define the relative contribution of each basis function to the HRIR and are real-valued functions of the spatial location. At an arbitrary spatial location, the estimates of the RSCFs, as well as the estimates of the interaural time and level cues can be obtained by interpolating the RSCFs and the interaural time and level cues at measurement locations. The modeled binaural HRIRs are reconstructed from these estimates. The modeled binaural HRIRs are used to synthesize stimuli that, when presented over earphones, simulate free-field sound signals transmitted to the eardrums of a cat. The work described in this article represents our attempt, through the development of a simple binaural model, to simulate realistic stimuli.

I. MODELING METHODS

The data employed in the evaluation of the model were derived from a detailed set of direction-dependent recordings made near the eardrum of an anesthetized cat by Musicant *et al.* (1990). In that study, the cat was secured in a holder in an experimental chamber with the head oriented. The frontal direction at ear level was defined as 0° azimuth and 0° elevation. The left and right sides at ear level were defined as $-90°$ and $+90°$ azimuth, respectively. The back direction at ear level was defined as $\mp 180°$ azimuth. Direction on the median plane spanned the range of $\pm 90°$ elevation. The free-field loudspeaker was moved in the spherical coordinate system. The loudspeaker signal was a digitally generated 10-μs rectangular pulse. The output of the preamplifier of the microphone, whether in the free-field or from the ear canal, was sampled by a 12-bit A/D converter at a rate of 160 kHz. Sampling was for a duration of 6.4 ms with a total of 1024 sampling points. Two recordings (left and right ears) were made for each position at which the loudspeaker was located. In this study, the data on a 9° grid are used to develop the model and the whole set of data on a 4.5° grid is used for comparisons with the model.

A. Data preprocessing

The free-field recording and eardrum recordings were contaminated by random noise, echoes, and dc levels that were produced by microphone, reflecting objects, and amplifier, respectively. Data preprocessing eliminates these effects as far as possible. Figure 1(a)–(f) is a depiction of the data preprocessing employed to eliminate the unwanted components of raw recordings. Careful examination of the eardrum recording [Fig. 1(a)] and free-field recording [Fig. 1(b)] reveals that random noise is dominant in the low (<1.5 kHz) and high (>30 kHz) frequency regions where the speaker has low responses and these recordings contain one or two echo components following the primary response peak. These unwanted components can cause very large variances in the estimated HRTF [Fig. 1(e)].

A Hamming window and linear phase finite-impulse-response (FIR) filter were chosen to preprocess these recording signals [Fig. 1(c) and (d)]. The distance between the primary response peak and the first echo is about 2.5 ms. So, we selected the window width as 450 sample points, corresponding to a width of about 2.8 ms. The window center is located at the primary response peaks of the recordings. A linear phase FIR filter was used to eliminate the random noise in free-field recording and eardrum recordings. The filter is a bandpass filter with 0.5-dB ripple for 1–30-kHz passband and 65-dB attenuation for 40–80-kHz stop band. The additional delay of the filter output is a constant with no phase distortion in the passband. After filtering, the signal was downsampled by a factor of 2 (to 80 kHz), thereby reducing the number of sampled points to 512.

Differences in propagation path lengths between sound source and eardrum lead to different delays in eardrum recordings. Onset times of each recording were detected. Details of detecting method were described by Chen *et al.* (1995). Each eardrum recording was then shifted backward or forward to a common time origin. Interaural time difference (ITD) is a major cue influencing the identification of the sound source direction. We calculated the ITD by subtracting the onset time of ipsilateral recording from the onset time of contralateral recording at each direction. The ITDs were used as a delay parameter for every reconstructed HRIR in a binaural model.

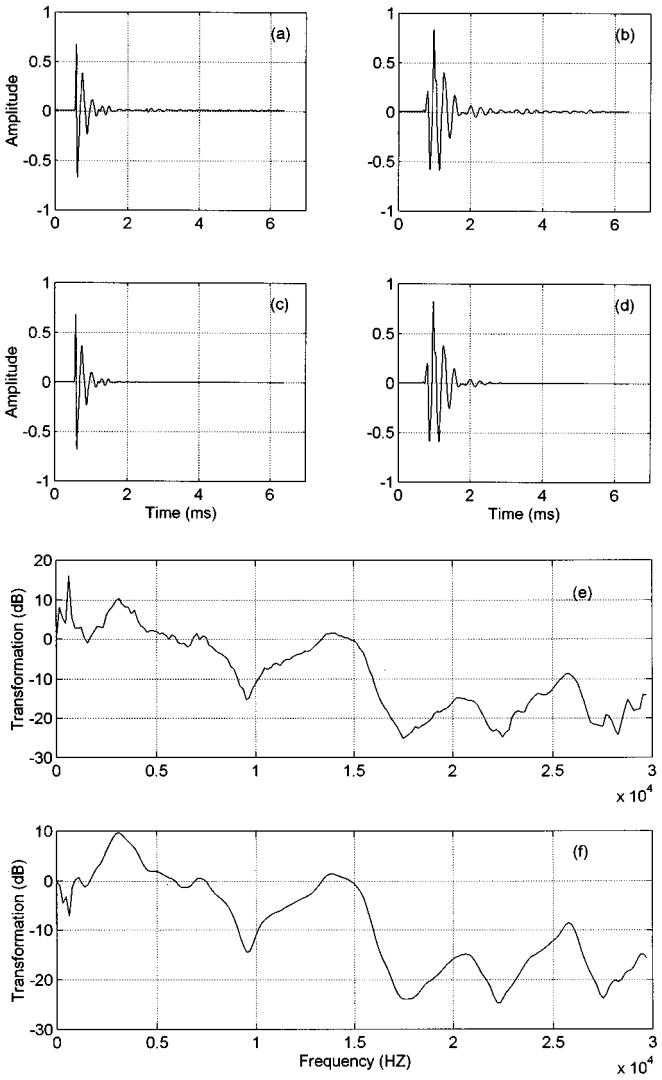

FIG. 1. Comparison between the HRTFs using the raw data in free field (a) and eardrum (b) and preprocessed data (c) and (d). Panel (e) is HRTF relative to (a) and (b); Panel (f) is HRTF relative to (c) and (d). Loudspeaker location was at 18° azimuth and 18° elevation.

B. Estimation and normalization of the HRIR

The HRIR, in most of the HRIR literature, is simply determined from the inverse Fourier transform of HRTF. The estimation of HRIR based on the fast Fourier transform (FFT) method is given by

$$H(\omega)=\frac{Y(\omega)+N(\omega)}{U(\omega)}, \tag{1}$$

where $Y(\omega)+N(\omega)$ and $U(\omega)$ are the Fourier transforms of eardrum recording $y(n)+\eta(n)$ and free-field recording $u(n)$, and $N(\omega)$ is Fourier transform of measurement noise $\eta(n)$. This method is termed empirical transfer function estimation. The estimate error in Eq. (1) is dependent on the measurement noise $\eta(n)$. Least-squares FIR filters could be applied to the HRIR estimation problem (Chen *et al.*, 1995; Wu *et al.*, 1996). The least-squares FIR filters were designed entirely in the time domain based on least-squares error criterion. Wu *et al.* (1996) compared the empirical estimation with the least-squares FIR filter estimation under different signal-to-noise ratio (SNR) conditions. The conclusion is that the least-squares FIR filter method is much better than the empirical method when the SNR of an eardrum recording is less than 40 dB. The SNR of an eardrum recording is difficult to define precisely because the level gain varies with azimuth and elevation. The maximum level gain difference over the sphere is about 25 dB. Some of the recordings have small peak values, that is only 30 dB above quantization noise of a 12-bit A/D converter. The level is generally low, furthermore, in the high-frequency region (above 15 kHz) where spectral notches occur in the eardrum recordings. For example, the noise peak is approximately 20 dB above the deepest notch at some directions. The improvement obtained with the least-squares FIR filter estimation is evident in this project.

Previous investigators (Hiranaka and Yamasaki, 1983) have reported that a 2-ms interval of time gating was enough to observe the human HRIR. The cat HRIRs have shorter duration because the cat external ear is smaller than the human's. Noting that the maximum cat ITD is about 0.26 ms, the duration of HRIR was selected longer than 2 ms. For the HRIR, the length N equals to the product of the duration (2×10^{-3} s) and the sampling rate (80×10^{3} Hz) and is corresponding exactly to the number of coefficients in the least-squares FIR filter $h(n)$ which is used to estimate the unknown HRIR system. The output of the FIR filter, $\hat{y}(n)=h(n)*u(n)$, where the asterisk denotes the convolution, is expected to be an approximation of $y(n)$. When the sum of square errors,

$$\epsilon=\sum_{n=0}^{N-1}[\hat{y}(n)-\{y(n)+\eta(n)\}]^2, \tag{2}$$

is minimized, the FIR filter coefficients $h(n)$, $n=0,1,...,N-1$, represent the best estimate of HRIR. This result is termed as least-squares FIR filter (Hakin, 1986).

Due to the ''head-shadow effect'', each HRIR has different energy which equals to the sum-square of $h(n)$. We define the square root of the energy as the level gain of the HRIR. All HRIRs used in our model were normalized by its level gain. The normalized HRIRs, having the same energy and onset time, would have different spectral characteristics. The normalizing procedures decrease the variances of HRIRs. Interaural level difference (ILD) is another major cue influencing the identification of the sound source direction. The ILD for broadband frequencies is the ratio of the ipsilateral level gain to the contralateral level gain. The ILD at a particular frequency or over a narrow frequency band depends on the ILD for broadband frequencies as well as the binaural HRIRs. The ILD for broadband frequencies was also used as an amplitude parameter for every reconstructed HRIR in a binaural model.

C. Low-dimensional representation of normalized HRIRs and linear interpolation

The low-dimensional representation of log-magnitude HRTFs was reported in (Kistler and Wightman, 1992). The low-dimensional representation of complex valued HRTFs

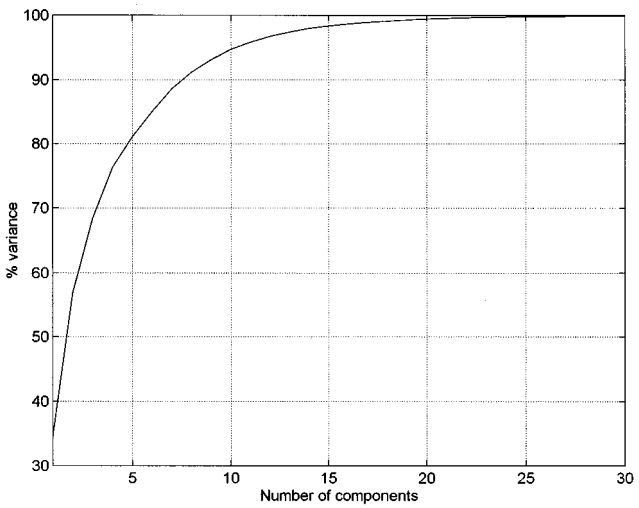

FIG. 2. The percentage of HRIR variation versus the number of components.

was described in detail by Chen *et al.* (1995). A brief description of our procedure for low-dimensional representation of normalized HRIR is given here.

We derived orthonormal basis functions from the normalized HRIRs. Prior to deriving the basis function, the space sample average was subtracted from each normalized HRIR to remove the direction-independent component. The space sample average is defined as

$$\mathbf{h}_{\text{av}}=\frac{1}{p}\sum_{j=1}^{P}\mathbf{h}_j, \tag{3}$$

where vector $\mathbf{h}_j$ with N elements represents the jth normalized HRIR from a set of P eardrum recordings. A time autocovariance matrix was then calculated by

$$\mathbf{R}_h=\frac{1}{P}\sum_{j=1}^{P}(\mathbf{h}_j-\mathbf{h}_{\text{av}})(\mathbf{h}_j-\mathbf{h}_{\text{av}})^T \tag{4}$$

which is a real valued matrix. The normalized HRIR on a 9° grid, which corresponds to $p=541$ measured samples, are used to determine the $\mathbf{R}_h$. The eigenvectors of $\mathbf{R}_h$ were chosen as the columns of an orthonormal transformation matrix $\mathbf{Q}=[\mathbf{q}_1 \quad \mathbf{q}_2 \quad \cdots \quad \mathbf{q}_N]$. For the Karhunen–Loeve expansion, a given normalized HRIR $\mathbf{h}_j$ is represented by

$$\mathbf{h}_j=\mathbf{Q}\mathbf{w}_j+\mathbf{h}_{\text{av}}=\sum_{i=1}^{N} w_{ij}\mathbf{q}_i+\mathbf{h}_{\text{av}}, \tag{5}$$

where the weight vector $\mathbf{w}_j$ is a set of orthonormal transform coefficients, given by

$$\mathbf{w}_j=\mathbf{Q}^T(\mathbf{h}_j-\mathbf{h}_{\text{av}}). \tag{6}$$

A small set of basis vector, that is discrete basis function, was then selected to represent the normalized HRIRs in a low-dimensional space. The Karhunen–Loeve expansion minimizes the mean-square error (MSE) for a given number of basis vector. Figure 2 shows the percentage of HRIR variation increases as a function of the number of components which are referred to weighted basis vectors. We noted that the first five components (Kistler and Wightman, 1992) represent 90% of human HRTF magnitude variation while the cat's HRIRs require eight components which account for the same amount of variation in Fig. 2. This is because that phase information is kept in the HRIRs and the HRIRs for the cat have a 30-kHz band width. We chose 20 components in this project, which can represent more than 99.9% of the variation in the normalized HRIRs.

We further define the ith coefficient subset $\{w_{ij},\ j=1,...,p\}$ as discrete samples of an underlying continuous function (Chen *et al.*, 1995), termed the ith real spatial characteristic function (RSCF). The RSCFs are functions of spatial coordinates (θ,φ) and are denoted as $w_i(\theta,\varphi)$, $i=1,2,...,20$. The normalized HRIR in a continuous space (θ,σ) is represented as

$$\hat{\mathbf{h}}(\theta,\phi)=\sum_{i=1}^{20}\hat{\omega}_i(\theta,\phi)\mathbf{q}_i+\mathbf{h}_{\text{av}}. \tag{7}$$

Using the linear space interpolation algorithm, for the arbitrary direction (θ,φ) over the upper 3/4 sphere (elevation from $-36°$ to 90°), we get the estimates of RSCFs

$$\begin{aligned}\omega_i(\theta,\phi)\approx{}&\frac{(\theta-\theta_2)(\varphi-\varphi_2)}{A}w_i(\theta_1,\varphi_1)\\&+\frac{(\theta-\theta_2)(\varphi-\varphi_1)}{A}w_i(\theta_1,\varphi_2)\\&+\frac{(\theta-\theta_1)(\varphi-\varphi_2)}{A}w_i(\theta_2,\varphi_1)\\&+\frac{(\theta-\theta_1)(\varphi-\varphi_1)}{A}w_i(\theta_2,\varphi_2),\end{aligned} \tag{8}$$

where $A=9\times 9$ is an area of the 9° grid, (θ_1,φ_1), (θ_1,φ_2), (θ_2,φ_1), and (θ_2,φ_2) are the four neighboring measurement locations of (θ,φ).

II. RESULTS

A. Interaural features

The ITD contour plot is presented in Fig. 3. This figure was constructed by subtracting the onset time of the right (near) ear from that of the left (far) ear for each sound source location. The plots, on a 4.5° grid, were determined through interpolation from the measured samples on a 9° grid. The numbers with a plus sign (+) are the time differences in ms. The maximum time difference, noted by an asterisk (*), is located at 108° azimuth, 0° elevation with 0.2625 ms. All contours are clustered around the maximum point. The minimum contour is located at the median plane (azimuth 0° or 180°) and the top of head (elevation 90°).

Figure 4 is an interaural representation (right minus left in dB) of level gain for broadband frequencies. The plots, on a 4.5° grid, also were determined through interpolation from the measured samples on a 9° grid. The numbers with a plus sign (+) are the level differences in dB. The maximum level difference, noted by an asterisk (*), is located at 72° azimuth, 9° elevation with 14.69 dB. The area up to 3-dB decrements is located in 45° to 95° azimuth and $-18°$ to 20° elevation. Minima, ranging from 140° to 170° azimuth and

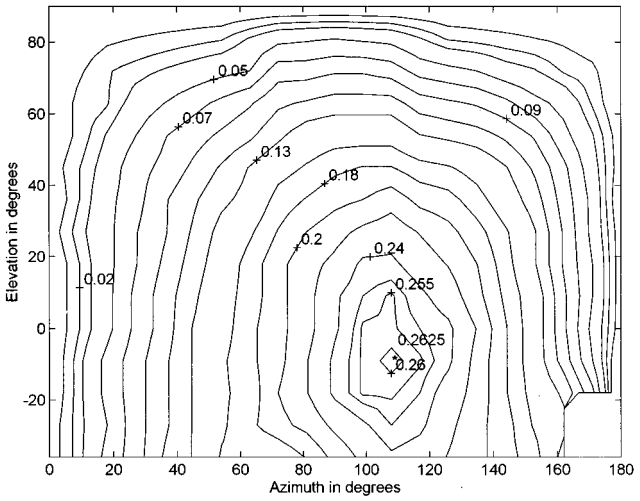

FIG. 3. Contour plot of interaural time difference (ITD) obtained by subtracting the onset time of ipsilateral recording from that of contralateral recording at any specified loudspeaker location.

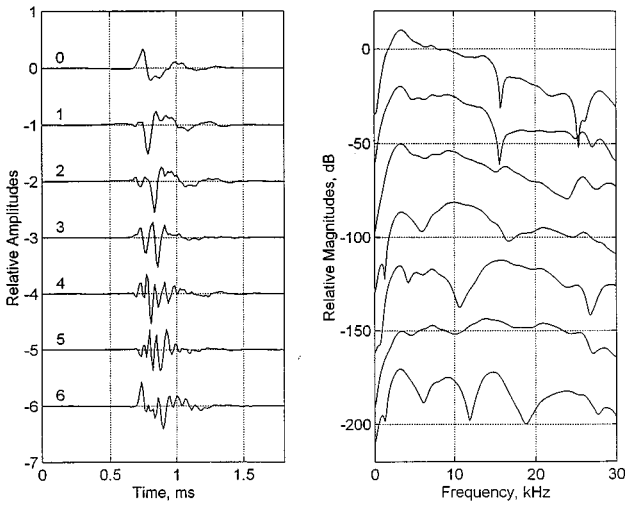

FIG. 5. The mean HRIR (numbered 0) and the first- through sixth-order basis functions (numbered 1–6). The left panel depicts amplitudes in time domain and the right panel magnitudes in frequency domain. The positions of the waveforms and spectra on the ordinate are arbitrary.

9° to 60° elevation, are negative ILDs. Those are most likely due to the fact that the lengths of two contralateral diffraction paths around the head are approximately equal in this region and the sounds excite more vibration modes on the left (far) ear. The values of ILD are near 0 dB at directly overhead and at the interaural axis.

B. Characteristics of basis functions and RSCFs

The space-sample average of normalized HRIR and six RSCF basis functions are plotted in Fig. 5. They represent time characteristics of auditory space described by the measured HRIRs. Each basis function is characterized by the number of the peaks and its envelope. Both the space sample average and the first-order basis function have broadband gain over the low- and middle-frequency ranges (up to 14 kHz) and a peak at 3 kHz. They contribute mostly to the first arriving wave that involves the resonances of the ear canal. The second-order basis function, with 40-μs lag, is very similar to the first-order basis function. This basis function seems to represent the second arriving wave produced by reflection or diffraction. It is difficult to directly relate specific basis function features to features in the HRIR waveform. However, more peaks are observed in the higher-order basis functions and they perhaps contribute to the details of high-frequency contents produced by reflection of pinna or diffraction around head.

A contour plot of the first-order RSCF of a cat's left ear is depicted in Fig. 6. The RSCF magnitudes have obvious features that are related to the geometry of the external ear. The maximum peak is 0.6114, noted by an asterisk, at azimuth $-54°$ and elevation 27° of the ipsilateral side. This is the opening direction of the cat's pinna. In general, the values in the first-order RSCF increase with small peaks and notches as the sound source moved from below to the top of the head. For a fixed azimuth, the positive maximum RSCF usually occurs in the region between 45° and 90° elevation. The zero contour in the front ipsilateral side is below the

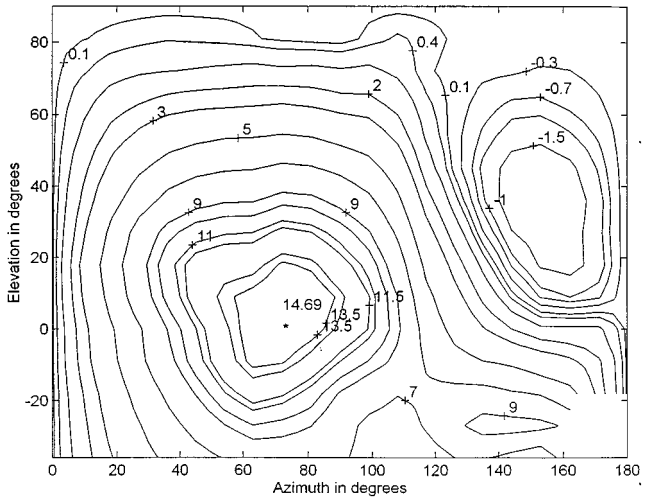

FIG. 4. Contour plot of interaural level difference (ILD) obtained by subtracting the level gain of contralateral recording from that of ipsilateral recording at any specified loudspeaker location.

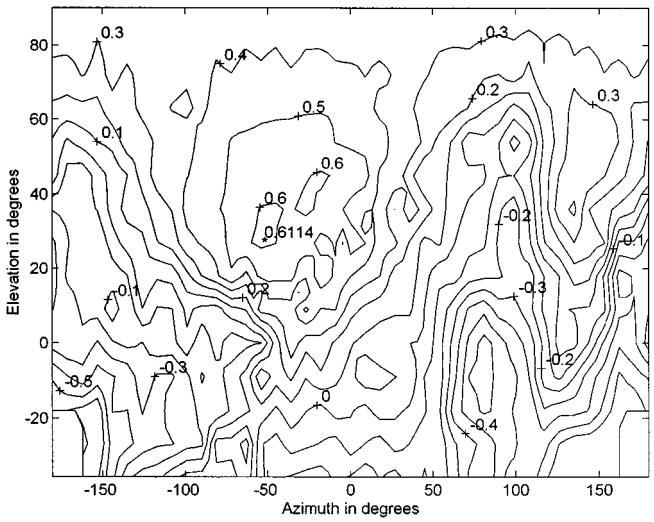

FIG. 6. Contour plot of the first-order RSCF.

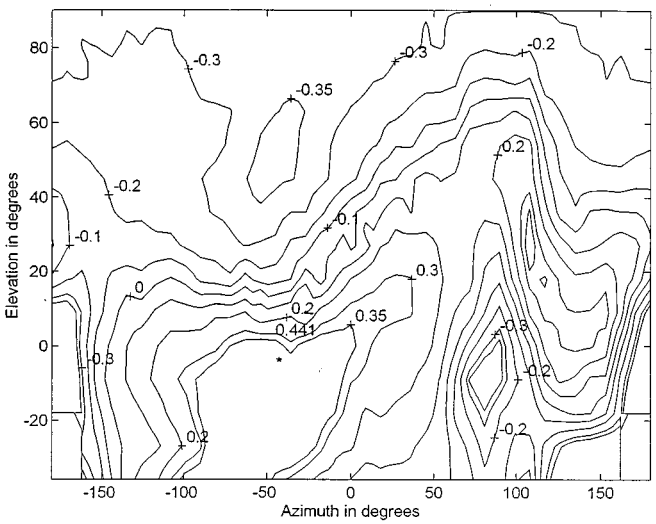

FIG. 7. Contour plot of the second-order RSCF.

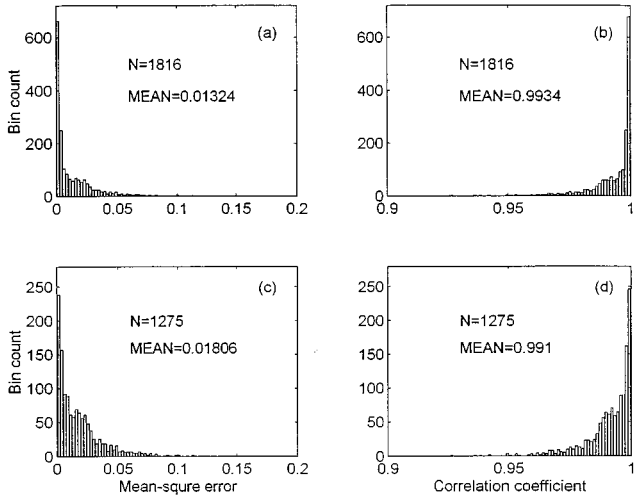

FIG. 8. Distributions of MSE (a) and cross-correlation coefficients (b) for all 1816 directions, and distributions of MSE (c) and cross-correlation coefficients (d) for 1275 directions which are not used to develop the model.

equator but then moves up in all other regions. Of interest is that RSCF has positive values in the direct sound regions and minus values in the head shadow regions. The head shadow also effects the low-elevation region of the ipsilateral side, as cat's ears are at the top of their head.

A contour plot of the second-order RSCF is depicted in Fig. 7. In contrast with the first-order RSCF, the maximum peak of the second-order RSCF positions at azimuth $-45°$ and elevation $-4.5°$ in the lower-elevation region. We also observe that both the first-order RSCF and the second-order RSCF have large absolute values in the region of elevation $-20°-0°$ and azimuth 70°–100°. The contralateral diffraction in the region contains many low and middle frequencies, according to the characteristics of related basis functions. The complexities of the RSCFs increase and the average absolute magnitudes decrease as the RSCF order increases.

C. Error between the measured and modeled HRIRs

In order to judge the suitability of the model, a comparison was made between measured and modeled HRIRs for the left ear of a cat. The HRIRs on a 9° grid (at 541 directions) were used to determine the model. The MSE and cross-correlation coefficient were examined by comparing the model and measured HRIRs on a 4.5° grid over a total of 1816 locations. The distributions of MSEs and cross-correlation coefficients for all 1816 directions are presented in Fig. 8(a) and (b), respectively. The ranges of MSEs and cross-correlation coefficients for this population are 0.03%–20% and 0.9–0.999, respectively. These indicate that some HRIR pairs are not very similar. However, the MSEs for 62% of 1816 HRIRs are less than 1% and only that for 0.5% of 1816 HRIRs are more than 10%. The mean MSE, 1.32%, is quite small and the mean cross-correlation coefficient, 0.993, is very high. HRIR model's interpolation capabilities are further illustrated in Fig. 8(c) and (d). In the figures, MSEs and cross-correlation coefficients are compared at the directions not used to develop the model. These locations are at the midpoints of those used to develop the model parameters and thus represent likely locations of maximum interpolation error. The mean MSE, 1.8%, is small and the mean cross-correlation coefficient, 0.991, is high for these locations. Those indicate that the interpolation errors are slightly greater than approximation errors. In order to compare with the thin-plate spline algorithm (Chen *et al.*, 1995), the average errors over subregions on the sphere are listed in Table I. Typical average errors are 1% with the range of 0.2%–2.9%. Larger average errors are obtained in the frontal ipsilateral region near the opening direction (at about azimuth $-40°$ and elevation 18° for the left ear). The result is very similar to thin-plate spline algorithm in error distribution except the rear lower elevation area where thin-plate spline algorithm has larger errors. There are two reasons for the error difference in this area. The first is that the preprocessing and least-squares method improve the quality of estimating the HRIRs in this project. The second reason is that the normalizing procedures decrease the variance of HRIRs. Table II lists the average cross-correlation coefficients for different subregions. The distribution of average cross-correlation coefficients is similar to the distribution of average errors. The data in these tables show that the error variances are small enough to maintain the model fidelity. This result suggests that the linear interpolating algorithm can adequately fit the HRIR at any direction over the upper 3/4 sphere.

TABLE I. Average error in 32 subareas.

Azimuth in degrees	Elevation in degrees			
	−36 to 0	0 to 30	30 to 60	60 to 90
−180 to −135	0.7934	1.0015	0.9482	0.6856
−135 to −90	0.4782	0.8931	0.8424	0.2196
−90 to −45	1.5357	1.7152	1.6458	1.2636
−45 to 0	1.2111	2.9081	2.1361	1.4011
0 to 45	1.0788	1.8800	2.2632	0.7944
45 to 90	0.9872	1.0852	1.1176	1.0721
90 to 135	0.4519	0.7684	0.8226	1.1576
135 to 180	0.7602	1.1165	0.5345	0.8938

TABLE II. Average cross-correlation coefficient in 32 subareas.

Azimuth in degrees	Elevation in degrees −36 to 0	0 to 30	30 to 60	60 to 90
−180 to −135	0.9960	0.9950	0.9953	0.9966
−135 to −90	0.9976	0.9955	0.9958	0.9989
−90 to −45	0.9923	0.9916	0.9917	0.9936
−45 to 0	0.9939	0.9855	0.9893	0.9929
0 to 45	0.9945	0.9906	0.9886	0.9960
45 to 90	0.9951	0.9945	0.9944	0.9947
90 to 135	0.9978	0.9962	0.9959	0.9941
135 to 180	0.9962	0.9944	0.9973	0.9955

D. The model implementation in the stimulus delivery system

The general features of our stimulus delivery system have been described elsewhere (Chan, 1993; Rhode, 1976; Chen *et al.*, 1994), so only the essential elements to implement the model are summarized here. Two parallel computers were used in this project. Stimuli for a given direction were synthesized in DEC Alpha. The stimuli were synthesized by compensating the acoustic channels, computing the modeled binaural HRIRs at a given direction, and convolving the HRIR with the compensated sound source. The acoustic volume and ear canal shape vary from subject to subject, and even from time to time for the same subject depending upon the fit of the transducer to the ear. To insure high-fidelity stimulation, the acoustic channels must be compensated. The least-squares FIR filters were used to equalize the acoustic channels that are for the left and right ears, respectively (Chen *et al.*, 1994). The sound source, the length of which could be infinite, is convolved with the pair of the FIR filters to compensate the left and right channels. The binaural HRIRs are modeled by computing the normalized HRIRs from interpolating RSCFs, adding the delays of ITD parameters as well as scaling the amplitudes with the level gain parameters. The compensated signals were convolved with the modeled binaural HRIRs at a 160-kHz sampling rate in real time. The synthesized stimuli were then directly stored on the memory of Micro VAX that presents the stimuli through a specially designed digital stimulus system (Rhode, 1976).

III. SUMMARY AND DISCUSSION

A simplified model for HRIR was developed and has been implemented in a DEC Alpha computer to simulate the sounds to the eardrum of the cat. The model avoids minimum phase approximation by directly representing the impulse response of HRTF. Furthermore, the only operations involved in reconstruction of the HRIR are real multiplication and real addition, which means the cost of computation is low. The linear interpolating algorithm was also used to speedup the model computing. Typical MSEs and cross-correlation coefficients between modeled HRIR and measured HRIR are 1% and 0.995, respectively.

Both the largest MSE and the lowest cross-correlation coefficient occur at interpolated points near the opening direction of the cat's pinna where the HRIRs have strong echoes reflected from the pinna. By checking the HRIRs in this area, we find that the HRIRs have large amplitude variances. These can be resolved by increasing the space-samples near the opening direction. Also, we can reduce the space-samples in the region of fewer errors to decrease the amount of space-samples. The space-samples, in fact, are not well-distributed in the sphere. There is high sample density in the region of higher elevation.

It is potentially important that measuring precision and estimation quality for HRIR affect the model fidelity. Generally, a broadband test signal is employed to make free-field recordings. Both the spectra of the test signal and transfer function for the acoustic transducer, however, are not flat within the concerned frequency range. This causes low SNR in some frequency ranges where the transducer does not emit much energy. To improve the test signal, new test signals, such as maximum-length sequences or Golay codes, have been used in the transfer function measurements (Gardner and Martin, 1995; Zhou *et al.*, 1992). In this project, we show that the least-squares FIR filter can be employed in the estimation of HRIRs with a high degree of fidelity.

Although our work in this project is restricted to the cat, we believe that our method could be at least a step toward applying the method to virtual acoustic space (VAS) implementation for human beings.

ACKNOWLEDGMENTS

Data used in this paper were collected in the Department of Neurophysiology at the University of Wisconsin-Madison, supported by NIH program project Grant No. NS12732 and by NIH/NRSA institutional fellowship Grant No. NS07026. The authors would like to thank Professor J. E. Hind and Professor J. F. Brugge of that Department, and A. D Musicant, now at Middle Tennessee State University, for providing the data. Dr. R. Reale tested reconstructed signals in experimental subjects. We would also like to thank Dr. J. Chen for his invaluable suggestions on the manuscript. Computing assistance of R. Kochhar and Zhengran Li are very much appreciated. The present work was supported in part by University of Hong Kong Research Grants and National Natural Science Foundation of China.

Batteau, D. W. (**1967**). ''The role of the pinna in human localization,'' Proc. R. Soc. London, Ser. B **168**, 158–180.

Blauert, J. (**1983**). *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT, Cambridge, MA).

Brugge, J. F., Chan, J. C. K., Hind, J. E., Musicant, A. D., Poon, P. W. F., and Reale, R. A. (**1992**). ''Neural coding of virtual acoustic space,'' J. Acoust. Soc. Am. **92**, 2333.

Chan, J. C. K., Musicant, A. D., and Hind, J. E. (**1993**). ''An insert earphone system for delivery of spectrally shaped signals for physiological studies,'' J. Acoust. Soc. Am. **93**, 1496–1501.

Chen, J., Van Veen, B. D., and Hecox, K. E. (**1992**). ''External ear transfer function modeling: A beamforming approach,'' J. Acoust. Soc. Am. **92**, 1933–1944.

Chen, J., Wu, Z., and Reale, R. A. (**1994**). ''Application of least-squares FIR filter to virtual acoustic space,'' Hearing Res. **80**, 153–166.

Chen, J., Van Veen, B. D., and Hecox, K. E. (**1995**). ''A spatial feature extraction and regularization model for the head-related transfer function,'' J. Acoust. Soc. Am. **97**, 439–452.

Gardner, W. G., and Martin, K. D. (**1995**). ''HRTF measurements of a KEMAR,'' J. Acoust. Soc. Am. **97**, 3907–3908.

Genuit, K. (**1986**). ''A description of the human outer ear transfer function by elements of communication theory,'' Proceedings of 12th International Congress on Acoust. (Toronto, Canada), ADSTR. B6-8.
Hakin, S. (**1986**). *Adaptive Filter Theory* (Prentice-Hall, Englewood Cliffs, NJ).
Hiranaka, Y., and Yamasaki, H. (**1983**). ''Envelope representation of pinna impulse responses relating to three-dimensional localization of sound sources,'' J. Acoust. Soc. Am. **73**, 291–296.
Kistler, D. J., and Wightman, F. L. (**1992**). ''A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction,'' J. Acoust. Soc. Am. **91**, 1637–1647.
Musicant, A. D., Chan, J. C., and Hind, J. E. (**1990**). ''Direction-dependent spectral properties of cat external ear: New data and cross-species comparisons,'' J. Acoust. Soc. Am. **87**, 757–781.
Rhode, W. S. (**1976**). ''A digital system for auditory neurophysiology research,'' in *Current Computer Technology in Neurobiology*, edited by P. Brown (Hemisphere, Washington, DC), pp. 543–567.
Shaw, E. A. G. (**1974**). ''Transformation of sound pressure level from the free field to the eardrum in the horizontal plane,'' J. Acoust. Soc. Am. **56**, 1848–1861.
Wenzel, E. M. (**1992**). ''Issues in the development of virtual acoustic environments,'' J. Acoust. Soc. Am. **92**, 2332(A).
Wightman, F. L., and Kistler, D. J. (**1989**). ''Headphone simulation of free-field listening: I. Stimulus synthesis,'' J. Acoust. Soc. Am. **85**, 858–867.
Wu, Z., Chen, J., and Reale, R. A. (**1996**). ''Measurement and estimation of free field to eardrum transfer function for virtual acoustic space research,'' Acta Acust. (China) **21**, 84–92.
Zhou, B., Green, D. M., and Middlebrooks, J. C. (**1992**). ''Characterization of external ear responses using golay codes,'' J. Acoust. Soc. Am. **92**, 1169–1171.