HEAD RELATED IMPULSE RESPONSE INTERPOLATION FOR DYNAMIC SPATIALIZATION

Thippur V. Sreenivas
ECE Dept, Indian Institute of Science
Bangalore 560012, India

Vikas C.Raykar and Ramesh Raman
ECE Dept, Regional Engineering College
Thiruchirapalli 620015, India

Abstract

Interpolation of HRIRs is required for effective real-time synthesis of binaural signals for moving sound sources. Interpolation can be done using adjacent angle HRIRs or full azimuth HRIRs. Also, HRIRs can be interpolated in the time-domain or in a transformed domain, such as spectrum or principal-component domain. Considering several alternate interpolation techniques, it is found that linear interpolation provides the best fit to the measured data. Using principal component analysis, 8:1 reduction in HRIR representation is shown to provide satisfactory performance. Both simulation tests and blind perception tests are performed to corroborate the interpolation performance.

1 Introduction

Humans can localize the direction of a sound source based on the ITD (Interaural Time Difference) and IID (Interaural Intensity Difference) information in the binaural signals. However, for some positions that are equidistant from both ears, sound waves arrive at both ears at the same time. Yet, the human hearing system is able to percieve the direction of the source. This can be modelled as largely due to the spectral filtering by the pinnae and the head itself. If we consider the spectral differences in the binaural signals, we can define a transfer function (one for each ear) between the source and each of the ear canals. This is referred to as Head Related Transfer Function (HRTF). The impulse response corresponding to such a HRTF is referred to as Head Related Impulse Response (HRIR). It may be noted that HRIRs include the ITD and the IID information.

A pair of HRTFs or HRIRs corresponding to each ear, can be used for synthesizing spatial direction for a monophonic sound source. By using the HRIR-L and HRIR-R corresponding to a particular direction, we can filter the monophonic sound signal to obtain two signals to be heard binaurally, which would simulate the sound source at the chosen direction. This is referred to as static localization. In contrast, dynamic spatialization refers to synthesis of a moving audio source from one directional angle to another, using the monophonic signal. To a first approximation this can be done by gradually switching the HRIRs corresponding to different directional angles in static localization. This requires a large number of HRIRs for different angles which have not been measured. Thus, new HRIRs have to interpolated using measured HRIRs and also efficiently represented for real-time filtering.

The HRIRs used in the present study are obtained from MIT Media Lab [1]. The HRIR measurement consisted of a KEMAR Dummy Head placed in an anechoic chamber with microphones in the left and right ear canals. Pseudo random binary sequences were used to obtain the impulse responses at a sampling rate of 44.1 KHz. A total of 710 different angles were sampled at elevations of -40^o to 90^o and 360^o azimuth. Each HRIR is an FIR filter of length 512 samples.

2 HRIR Filtering

The real-time HRIR filtering can be done either in time-domain or frequency-domain through overlap-save technique. We have used time-domain convolution since we aimed at interpolation methods also in time-domain.

We use two TMS320C30 EVM boards, one each for processing the left and right channels. The EVM board has one TMS320C30 floating point processor, 16K X 32 bit zero wait state SRAM and one Analog Interface Controller(AIC) TLC32044 which is capable of providing a single channel audio interface upto a sampling rate of 19.2 KHz. The AIC is interfaced to the serial port-0 of the processor. The AIC provides single channel A/D conversion and D/A convertion at 14 bit resolution. The EVM also has a serial port-1 brought out to a 10 pin connector which is used for synchronization between the EVMs.

For dynamic spatialization, the HRIRs for required

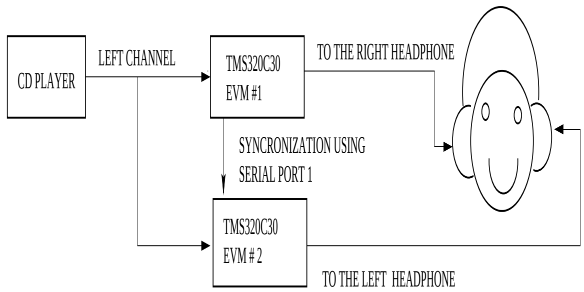

Figure 1: *HRIR filtering experimental setup*

consecutive angles are loaded in the external memory of the EVM. Here, we can control 3 parameters, viz., the initial angle, the final angle and the angular velocity of the source. Based on this, the required HRIRs are computed and saved in the memory. For a smooth movement of the source, we have chosen an angular resolution of one HRIR per 5^o azimuth. While switching HRIRs of successive angles, it is required that both the left and right ear pairs are matched. Thus, the two EVMs should switch simultaneously to HRIR-L and HRIR-R, respectively, at the same time. This is realized using a handshake mechanism between the two EVMs using serial port-1. Initially, after setting up the serial port-1 the processor getting loaded first waits for a code word from the other processor by polling the serial port-1 receive interrupt. On receiving the interrupt it reads the serial port-1 receive register, verifies the code word and then starts the filtering process. The other processor after writing the code word proceeds directly to filtering.

The EVM operates at 15MHz with the AIC having a sampling rate of 19.2 KHz. Thus, 750 cycles are available between two input signal samples. The FIR filter convolution requires 223 cycles plus some extra overhead cycles; thus, only a total of 250 cycles are used. The saved cycles could be used for online interpolation.

3 HRIR Reduction

Principal Component Analysis (PCA) is a statistical dimensionality reduction technique. The statistical characterization is through the covariance matrix of the finite impulse response vector. The aim is to seek orthogonal basis vectors, which when linearly combined would yield the required FIR vector. The basis functions are also ordered according to their contribution to the total variance of the covariance matrix. Thus, by choosing the principal contributors, one can get a lower dimensional representation of the original vectors. If p is the vector dimension, and N is the number of vectors available for PC analysis, the covariance matrix of the FIR vector $h_k(n), 1 \le n \le p$, is given by,

$$C_{ij} = \frac{1}{N}\sum_{k=1}^{N} h_k(i)h_k(j) - \bar{h}(i)\bar{h}(j), 1 \le i,j \le p, \quad (1)$$

where $\bar{\boldsymbol{h}}$ is the mean FIR vector of all angles. The pxp covariance matrix is diagonalized as $\boldsymbol{C} = \boldsymbol{BDB^t}$. The columns of the eigen-vector matrix $\boldsymbol{B}$ comprises of the basis vectors $\boldsymbol{b_i}$, $1 \le i \le p$. Among these $q < p$ vectors are retained corresponding to the largest eigenvalues. Thus, the original FIR vectors are best approximated by $\boldsymbol{h_k} = \sum_{i=1}^{q} \boldsymbol{b_i}\ \boldsymbol{w_k(i)}$, where $w_k(i)$ are the weight factors corresponding to the k_{th} specific FIR vector. These weight vectors are obtained by truncating the transformed vector $\boldsymbol{w_k} = \boldsymbol{B^t h_k}$.

Now, in the transformed sub-space the HRIRs for successive angles are assumed to be smoothly represented and hence amenable for interpolation. This will permit interpolation of a q-dimensional vector $\boldsymbol{w}$ than p-dimensional vector $\boldsymbol{h}$, reducing the complexity of online interpolation.

4 Interpolation Methods

Interpolation itself can be viewed as an HRIR reduction technique because of the possibility of using HRIRs for only coarsely sampled directional angles. However, interpolation is in general required because of the need to synthesize any angular motion for dynamic spatialization. Thus, a combination of PCA and coarse angle interpolation would result in a significant reduction of the HRIR memory requirement. To explore the influence of neighboring angles and farther angles in interpolation, we have analysed four interpolation techniques, viz., linerar, cubic, cubic spline and sinc interpolation.

Let $h(n,\theta_i)$ be the n^{th} point of the interpolated HRIR for θ_i^o azimuth. Then, for linear interpolation, we have

$$h(n,\theta_i) = \frac{(\theta_2-\theta_i)}{(\theta_2-\theta_1)}h(n,\theta_1) + \frac{(\theta_i-\theta_1)}{(\theta_2-\theta_1)}h(n,\theta_2) \quad (2)$$

where θ_i is any angle between θ_1 and θ_2.

For cubic interpolation, we have

$$\begin{aligned} h(n,\theta_i) = & \frac{(\theta_i-\theta_2)(\theta_i-\theta_3)(\theta_i-\theta_4)}{(\theta_1-\theta_2)(\theta_1-\theta_3)(\theta_1-\theta_4)}h(n,\theta_1) \\ & + \frac{(\theta_i-\theta_1)(\theta_i-\theta_3)(\theta_i-\theta_4)}{(\theta_2-\theta_1)(\theta_2-\theta_3)(\theta_2-\theta_4)}h(n,\theta_2) \\ & + \frac{(\theta_i-\theta_1)(\theta_i-\theta_2)(\theta_i-\theta_4)}{(\theta_3-\theta_1)(\theta_3-\theta_2)(\theta_3-\theta_4)}h(n,\theta_3) \\ & + \frac{(\theta_i-\theta_1)(\theta_i-\theta_2)(\theta_i-\theta_3)}{(\theta_4-\theta_1)(\theta_4-\theta_2)(\theta_4-\theta_3)}h(n,\theta_4) \quad (3) \end{aligned}$$

where θ_i is any angle between θ_1 and θ_4.

For spline interpolation, we have

$$\begin{aligned} h(n,\theta_i) &= Ah(n,\theta_j) + Bh(n,\theta_{j+1}) + Ch''(n,\theta_j) \\ &\quad + Dh''(n,\theta_{j+1}) \end{aligned} \tag{4}$$

where θ_i is any angle between θ_j and θ_{j+1} and

$$\begin{aligned} A &= \frac{(\theta_{j+1}-\theta_i)}{(\theta_{j+1}-\theta_j)} \\ B &= \frac{(\theta_i-\theta_j)}{(\theta_{j+1}-\theta_j)} \\ C &= \frac{(A^3-A)(\theta_{j+1}-\theta_j)^2}{6} \\ D &= \frac{(B^3-B)(\theta_{j+1}-\theta_j)^2}{6} \end{aligned} \tag{5}$$

$h''(n,\theta_j)$ and $h''(n,\theta_{j+1})$ should be continous across the boundary between two intervals. We get a set of $N-2$ linear equations with N unknowns $h''(n,\theta_i)$, $i = 1,2....N = 72$.

$$\begin{aligned} &\tfrac{(\theta_j-\theta_{j-1})}{6}h''(n,\theta_{j-1}) + \tfrac{(\theta_{j+1}-\theta_{j-1})}{3}h''(n,\theta_j) \\ &+\tfrac{(\theta_{j+1}-\theta_j)}{6}h''(n,\theta_{j+1}) = \tfrac{h''(n,\theta_{j+1})-(h''(n,\theta_j)}{(\theta_{j+1}-\theta_j)} \\ &\qquad + \tfrac{h''(n,\theta_j)-h''(n,\theta_{j-1})}{(\theta_j-\theta_{j-1})} \end{aligned} \tag{6}$$

Solving this tri-diagonal linear system with not-a-knot boundary conditions we get $h''(n,\theta_i)$, $i = 1,2....N$.

For sinc interpolation, $h(n,\theta_i)$, $i = 1,2,...,k < 72$ are transformed to the Fourier domain using DFT and then inverse transformed by padding with zeros upto a length of 72 points.

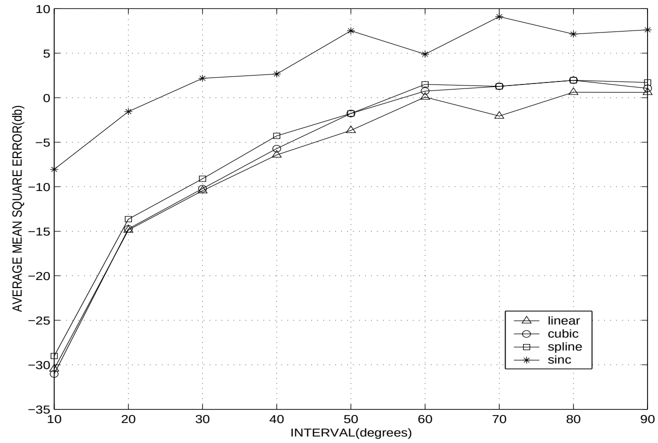

Figure 2: *Time-domain HRIR interpolation error*

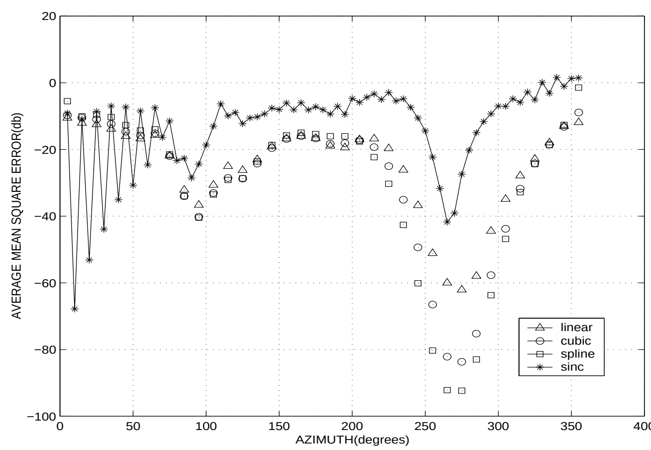

Figure 3: *Interpolation error as a function of azimuth angle for* 10^o *sub-sampling*

5 Interpolation Performance

The performance of different interpolation techniques are evaluated by angular sub-sampling of the available HRIR data and then interpolating the HRIRs at the remaining angles. These interpolated HRIRs are then compared with the available measured data using the mean-squared-error performance measure. Considering only 0^o elevation, HRIRs are available at a resolution of 5^o. With an angular sub-sampling corresponding to $(10^o, 20^o, ..., 90^o)$, all the above four interpolation methods are carried out to compute HRIRs at the remaining angles. The average mean square error γ is shown below. Let $h(n,\theta_i)$ be the measured HRIR for azimuth θ_i and let $\hat{h}(n,\theta_i)$ be the interpolated HRIR for the same direction.

$$\gamma = \frac{\sum_{i=1}^{72}\sum_{n=0}^{511}[h(n,\theta_i) - \hat{h}(n,\theta_i)]^2}{\sum_{i=1}^{72}\sum_{n=0}^{511}[h(n,\theta_i)]^2} \tag{7}$$

A value of $\gamma = -\infty$ dB indicates perfect interpolation.

Fig.2 shows that except sinc interpolation, the other three methods provide similar performance at different angular sub-sampling. Of all the four techniques, linear interpolation gave the least error. For sub-sampling > 60^o $\gamma \approx 0$dB. In Fig.3 we can see interpolation error as a function of the azimuth angle for 10^o sampled data. As expected, the errors are large at the boundary angles, but interestingly 90^o, 270^o azimuth show lower errors. This may be due to the fact that the variation of HRIR in the vicinity of 90^o and 270^o is not much. It is also noted that the error is maximum at the centre of each interval.

To combine interpolation with PCA, we again considered only 0^o elevation with 72 azimuth angles at 5^o interval. After PCA, it is found that 8 basis functions capture about 90% of the variance. The HRIR reconstruction error due to the PCA alone is found to be

$\gamma = -23$ dB. Further, the interpolation techniques are applied on the transform domain weight vectors corresponding to sub-sampled angular directions. Fig.4 shows the MSE performance; again, the same 3 interpolation methods except sinc interpolation showed better performance. From Fig.2, An error of -23 dB in time-domain interpolation corresponds to a sub-sampling interval of 15^o. This requires 24 HRIRs to be stored. So, PCA with 8 basis functions and 5^o interval HRIRs is more compact than 15^o interval time-domain HRIRs.

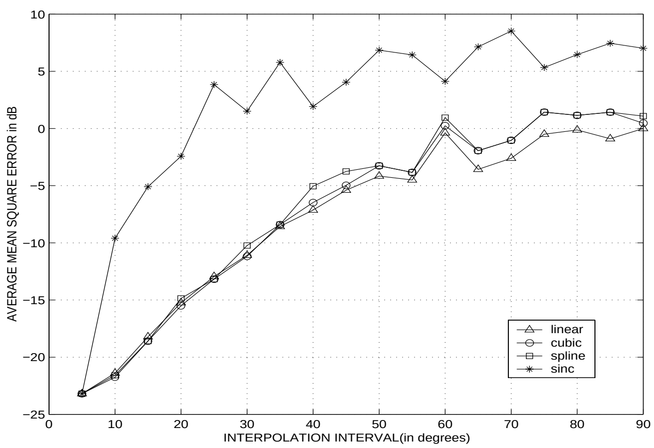

Figure 4: *PC domain interpolation error*

6 Perceptual Evaluation

The perceptual evaluation of dynamic spatialization consists of presenting to individual listeners (Fig. 1) real-time processed audio signals with a randomly chosen angular motion. The listeners are initially exposed to training samples with known angular movement, but in a random order. The angular velocity of the sound source is fixed at $22.5^o/sec$ for comfortable listening and the initial and final angular positions are randomly chosen. After reaching the final angular position, a silence period of two seconds is introduced. The source movement is then repeated cyclically. The listener can listen to the random sample indefinitely long and then judge the angular path of the source. The training samples consisted of four stimuli: $0-90^o$, $90-180^o$, $180-270^o$, $270-0^o$. During the testing phase, the listener response is recorded on stencil along with his comments on the difficulty of judgement.

For dynamic spatialization with interpolated data, the same stimuli as the previous experiment are chosen in a random order and presented to the same listeners. HRIRs from the best case interpolation method (linear) for interpolation intervals of 10^o and 90^o are chosen. Any difference in perceptual performance with respect to the previous set of experiments is noted.

We performed the tests for 5 male adult listeners with normal listening ability. For dynamic localization without interpolation all the subjects could percieve the movement of the source within a quadrant. The maximum average error (error in inital angle+error in final angle/2) is found to be 35^o. However, there were some cases of front-back confusions (i.e.movement in front of head in one direction was percieved as movement in the other direction at the back of the head.) Some subjects showed consistently higher error. This may be probably bue to the HRIR mismatch. For interpolation in time-domain, it is noted that the difference in perception is not significant for either 10 and 90 degree interpolation interval. Some subjects commented that the source appeared to move at a higher elevation. There were also some front back confusions. For all subjects the front back confusions occured only for the 90 degree interval. All subjects could percieve the source movement using the HRIR generated from the 8 basis functions. Further, for interpolation of PC weights the error in perception between 10 and 90 degrees was not noticeable.

7 Conclusion

The present study has shown that HRIRs can be reconstructed with sufficient accuracy using principal eigenvectors. Interpolation in the principal component domain is also feasible without much additional loss compared to interpolation in the time-domain. Of the various interpolation techniques, simple neighboring angle linear interpolation has shown the best reconstruction. The blind perception experiment results indicate that large errors in interpolation are tolerable for dynamic spatialization, unlike in static localization.

References

[1] MIT Media Lab Perceptual Computing - technical report#280 "A HRTF Measurement of a KEMAR Dummy-Head Microphone".

[2] D.J. Kristler and F.L. Wightman "A model of head related transfer functions based on principal components analysis and minimum phase reconstruction",J.Acoust.Soc.Am.91(3),March 1992

[3] A.V. Oppenhein and R.W. Schaffer: "Discrete-Time Signal Processing", Prentice-Hall, 1985.

[4] Texas Instruments TMS320C30 evaluation module-Technical Reference

[5] Texas Instruments TMS320C3x User's Guide