A PIECEWISE INTERPOLATION METHOD BASED ON LOG-LEAST SQUARE ERROR CRITERION FOR HRTF

Jie Zhang, Zhenyang Wu

Department of Radio Engineering, Southeast University, Nanjing, 210096, P. R. China
E-mail: zjhf1978@sohu.com

ABSTRACT

This paper addresses the problem of accurately realizing the interpolation of spatially discrete head-related transfer function (HRTF) for synthesis of virtual auditory space. By analyzing the advantages and disadvantages of general bilinear interpolation method in 3D-sound, associating with human auditory system’s mechanism of band-pass filtering and consulting critical bands of psychoacoustics, the paper presents a piecewise interpolation method based on log-least square error criterion for HRTF’s magnitude to compensate the deficiency of bilinear method in intermediate frequency. As seen from the following simulations, this new method accomplishes preferable results.

1. INTRODUCTION

In the study of spatial hearing, HRTF plays an important role. It can be said that most of the study in this field has been progressing around it, including the analysis of structural or mathematical model with regard to certain subject, the measurement or computation of its data as well as their corresponding relationship’s confirmation, and so on [1, 2, 3, 4].

Nowadays, the primary approach to acquire HRTF is still actual measurement under rigorous conditions in anechoic room. Then limited by this, the measurement only represents the spatial characteristics on discrete positions. As a result, during the synthesis of virtual 3D-sound, it is indispensable to consider how to properly interpolate those unmeasured positions’ HRTFs from the finite measured ones.

At present, a straightforward method to do this is the bilinear [5], whose basic idea is to realize the target position’s interpolation with its four nearest measured HRTFs according to their ubiquity. An example of this method’s avail is shown in Figure 1. It is obvious that, in the regions of 6KHz below and 17-18KHz above, the interpolation effect is good, whereas in spacious intermediate region, the effect worsens evidently.

By investigating its reason, we think that, this method doesn’t take cognizance of the different influence for different wavelength of sound wave, and tries to uniformly deal with the entire audible spectrum from LF to HF (20Hz-20000Hz). While especially in 7-16KHz, the wavelength of sound wave is just comparative with the dimension of human’s ear, which makes for evident reflection and dispersion. Therefore, a little replacement usually brings great fluctuation, and then the bilinear’s invalidation is understandable. Nevertheless, it is reasonably true that, along with the motion of measured position up and down or right and left, there is indeed a tendency of the magnitude, just as shown in Figure 2. Eclectically, it is likely to piecewise obtain a better approach and interpolation effect. This

This work is supported by the 973 Program of China under Grant No. 2002CB312102.
paper validates the viewpoint. Furthermore, in view of the specialty of human auditory system (HAS), this paper felicitously manages to partition the acoustic spectrum and achieves some preferable results.

Another point should be mentioned. The above discussion only talks about the magnitude of HRTF, and ignores the phase. The reason is that, during the measurement, phase is easy to be contaminated by equipments and environments. Then the phase we measured is not accurate in the beginning. Here we will only take into account how to interpolate the magnitude of HRTF. And the information of phase can be recovered by means of the minimum phase characteristic corresponding to magnitude, which cannot cause special perceptive reduction in synthesis [1, 5, 6].

In addition, aiming at the amelioration of the bilinear, Inter-Positional Transfer Function (IPTF) is introduced in [5], which gives expression to the transform relationship between the target position and its most neighboring one, then indirectly works out the desired HRTF. Compared with the bilinear, this approach improves the interpolation effect to a certain extent. However, unfortunately, all the same as bilinear, in intermediate frequency region, the result is still unsatisfactory. The detailed comparison sees also [5].

Synthetically, the paper is organized as follows. In Section 2, the data of HRTFs used for simulation are simply described. And then our procedure with the elementary simulation results is detailedly presented in Section 3. After that, an elaborate partition of the intermediate spectrum is shown, which has a significative result. Section 5 is the conclusion and future work.

2. DESCRIPTION OF THE DATA FOR SIMULATION

Here, the data of KEMAR’s HRTFs, which offered by MIT Media Lab [7], are applied to the simulation in this work. The measurements of the data were made with speaker on the discrete positions every 10° in elevation, and 5°-30° unequally in azimuth. Table 1 lists the detailed sample points. And it is apparent that the spatial sampling is quite discrete. Therefore, interpolation is absolutely requisite for practical application.

Table 1. The distribution of KEMAR’s measurement

<table>
<thead>
<tr>
<th>Elevation (°)</th>
<th>Number</th>
<th>Step (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-20-20</td>
<td>72</td>
<td>5.00</td>
</tr>
<tr>
<td>-30-30</td>
<td>60</td>
<td>6.00</td>
</tr>
<tr>
<td>-40-40</td>
<td>56</td>
<td>6.43</td>
</tr>
<tr>
<td>50</td>
<td>45</td>
<td>8.00</td>
</tr>
<tr>
<td>60</td>
<td>36</td>
<td>10.00</td>
</tr>
<tr>
<td>70</td>
<td>24</td>
<td>15.00</td>
</tr>
<tr>
<td>80</td>
<td>12</td>
<td>30.00</td>
</tr>
<tr>
<td>90</td>
<td>1</td>
<td>—</td>
</tr>
</tbody>
</table>

As stated above, the measured data may be contaminated by deficient factors. Thus it is necessary to get rid of the infection before further processing. As far as the actual operation is concerned, a Hamming window with width of 128 points is applied to the original data [3]. Some other meticulous and important processing refers to [3, 7].

In the following simulation work, we only use the data of right ear, just as in the above Figure 1 and Figure 2.

3. PRESENTATION OF THE INTERPOLATION PROCEDURE

By analyzing the extensive literatures involved with spatial hearing and the above-mentioned contents, a few assumptions are applied in this method.

Firstly, human auditory characteristics can be felicitously approximated under the circumstances of logarithmic magnitude spectrum [9]. Here we are also conscious of the great comparability and pertinence among closer HRTFs. And on second thoughts, this paper assumes that the logarithmic magnitude of approximate frequencies and neighboring positions can be approached by linear equations, which will be used in our interpolation procedure. As for the positions at a distance, comparability of HRTFs is markedly different from one another. So it is reasonable to ignore their effects. Furthermore, just as in the bilinear, we will only make use of the four neighboring positions to model the relationship.

Secondly, when the difference of magnitude responses is below 2 dB, human cannot differentiate them [9]. Thus the error below 2 dB will be overlooked in simulation programming.

In the following contents, the new method and its simulation results will be presented detailedly.

In the first place, for intermediate and high frequency between 7KHz and 16KHz, according to the aforesaid assumption, we take it for granted that the target position’s HRTF can be interpolated with neighboring positions’ HRTFs by means of linear equations in a narrow frequency span. Whereas when the span become wide, it is necessary to alter the interpolation coefficients to make after the variational trend.

In the next place, for the other two ranges, here are the lower frequency range below 7K and higher range above 16KHz, the influence of outer ear’s reflection and dispersion on HRTFs’ interpolation is relatively little between neighboring positions. Whereas considering the preferable result of bilinear interpolation, it is correctly the electee for these two ranges. The following are the detail process.

Let’s suppose that the position to be interpolated is denoted by \((\text{elev}_0, \text{azim}_0)\), where \(\text{elev}\) is the abbreviation of elevation, and \(\text{azim}\) is azimuth. Moreover, the four adjacent positions of the target are as follows: \((\text{elev}_1, \text{azim}_1)\), \((\text{elev}_2, \text{azim}_2)\), \((\text{elev}_3, \text{azim}_3)\) and \((\text{elev}_4, \text{azim}_4)\). According as the linear assumption of the relationship between the neighboring positions, the following equation can be educed

\[
\text{HRTF}(\text{elev}_0, \text{azim}_0, f) = x_1\text{elev}_0 + x_2\text{azim}_0 + x_3 f. \tag{1}
\]

Here, \(x_j\) \((j \in \{1, 2, 3\})\) are the interpolation coefficients and can be regard as the latent influence factors of the three variables, i.e. \(\text{elev}, \text{azim}\) and \(f\), in a certain frequency span, which can be worked out based on the minimization algorithm of the undermentioned formula,

\[
\min_x \sum_{j=1}^4 \left[ \ln |H_j| - p_j x_j \right]^2 \tag{2}
\]
\[ H_i = HRTF(elev_i, azim_i, f(n)) \]
\[ p_i = [elev_i, azim_i, f(n)] \]

Where, \( H_i \) (\( i = 1, 2, 3, 4 \)) are the four measured HRTFs around the desired position, and \( p_i \) are the corresponding position’s parameters. The vector \( x \), just as discussed above, is the best interpolation coefficient used to describe their variational trend in the log-least square sense of formula (2). In addition, \( f(n) \) is the central frequency of divided frequency regions, and \( n \) is the number index of certain frequency span, which varies according to the different spectrum and corresponding partition.

To sum up, in different frequency regions, we first calculate different \( x \) from formula (2), and then use them in formula (1) to get the desired HRTFs.

Fig. 3. A simple simulation result of piecewise method

Fig. 4. Difference between the measured and the piecewise interpolated HRTFs on the horizontal plane

A simple simulation result of this method is shown in Figure 3. It is apparent that the piecewise interpolated HRTF extraordinarily approaches the measured one in intermediate frequency region, which effectively indicates the rationality of the piecewise interpolation.

In Figure 4, a general example is given. Here, the abscissa denotes frequency, whose range is from 7KHz to 15KHz. And the frequency span is 86.133Hz, that is, the interpolation coefficients are needed to be calculated every 86.133Hz. The ordinate is azimuth angle, \( 10^\circ \) to \( 170^\circ \), which represents the spatial area at \( elev = 0^\circ \) for \( 10^\circ \leq azim \leq 170^\circ \) (for the convenience of programming). On the contour line, the number is the absolute difference of measured and interpolated HRTFs’ magnitude in dB.

It can be seen that most of the difference is around 0dB, which indicates the generality of this method. However, in the diagonal area of about 9.5KHz < \( f < 10.6KHz \) and \( 50^\circ < azim < 120^\circ \), there is relatively obvious interpolation error. The reasons are just that, the wavelength of sound wave in this area (3.1cm-3.6cm) is comparable to the physical dimensions of pinnae, which has intricate shape and delicate configuration; moreover these positions are just exposed to the sound source. Under these conditions, sound wave obviously reflects or disperses on outer ear. Therefore, small variations of space or wavelength usually can produce large changes in the HRTF. However, we just choose a set of interpolation coefficients of certain frequency to approximate the trend in its frequency span, and then the error occur in the nature of things. On the other hand, it just shows that, there is profuse spatial information in this audio spectrum [2].

As for this problem, the solution is nothing but more elaborate segmentation of this part spectrum. In our other simulations, it has been validated.

4. SIMPLIFICATION OF COMPUTATION IN CRITICAL BANDS

From the above simulation, we can see that, this piecewise method acquires better results in the intermediate frequency, where the bilinear becomes invalid. However, it needs a lot of computations in each section of the spectrum to solve Equation (2). Therefore, based on the consideration of simplifying processing, and inspired by the perspective of psychoacoustics, we assume that, in term of the particular human auditory band-pass filtering characteristic, it is possible to make it by appropriate division of the spectrum.

<table>
<thead>
<tr>
<th>Sequence number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower limit frequency (Hz)</td>
<td>6029.3</td>
<td>7149</td>
<td>9043.9</td>
<td>11197</td>
</tr>
<tr>
<td>Central frequency (Hz)</td>
<td>6632.2</td>
<td>8527.1</td>
<td>9905.3</td>
<td>12834</td>
</tr>
<tr>
<td>Upper limit frequency (Hz)</td>
<td>7062.9</td>
<td>8957.8</td>
<td>11111</td>
<td>15073</td>
</tr>
<tr>
<td>Bandwidth (Hz)</td>
<td>1033.6</td>
<td>1808.8</td>
<td>2067.1</td>
<td>3876</td>
</tr>
<tr>
<td>Corresponding CB (Bark)</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
</tr>
</tbody>
</table>

As an efficient physiological measure, critical bands approximate the band-pass filtering characteristics of human auditory system and quantify the minimal bandwidth in which sound signal masking is especially notable. Here, in accordance with the psychoacoustical concept, we design a filter-bank model, whose bandwidth varies from 1KHz to 4KHz along with the increase of acoustic signal frequency. Table 2 is just obtained in this way. Furthermore, we choose the central frequency of intersected band as the representative to calculate
the interpolation coefficients in each band. As a result, the simulation has some special significance.

Fig. 5. A simple simulation result using piecewise method in critical bands

![Fig. 5. A simple simulation result using piecewise method in critical bands](image)

According to the above partition, a simple simulation result is shown in Figure 5(a). It can be seen that, the interpolated HRTF basically accords with the measured one. However, we only calculate four sets of interpolation coefficients in four critical bands we choose in intermediate frequency. Compared with the before times 93 ((15073-7062.9)/86.133=93) in Section 3, the computation decrease greatly. Of course, there is still a connection problem on the turning point between adjacent critical bands. Therefore, it is required to smooth the discontinuity.

Here, we just want to smooth the interpolated magnitudes in numerical value, which is just like the usual “low-pass filtering”. In our procedure, a simple smoothing technique, which is similar to that in [10], is used. Detailedly, DCT (discrete cosine transform) is applied to the above-interpolated data; then the first ten coefficients, which correspond to the gradual slope, are preserved for IDCT (inverse discrete cosine transform) to get the smoothed HRTF shown in Figure 5(b). It is clear that the result just approaches the measured one.

A more general simulation result on the horizontal plane is shown in Figure 6. Just as in Figure 4, the most obvious error also exist in the diagonal area of about 9.5 KHz < f < 10.6 KHz and 50° < azim < 120°, but the whole result is basically equivalent to that in Figure 4.

Fig. 6. Simulation result using piecewise method in critical bands on the horizontal plane

![Fig. 6. Simulation result using piecewise method in critical bands on the horizontal plane](image)

5. CONCLUSION AND FUTURE WORK

In this paper, a piecewise method based on the log-least square error criterion is applied to compensate the deficiency of bilinear interpolation method for HRTFs’ magnitude in intermediate frequency. As seen from the results, this new method achieves the aim on the whole. Furthermore, the psychoacoustical measure of critical band is also used for reference in the partition of intermediate frequency spectrum. As a result of this, the computational complexity greatly decreases. At the same time, its result is still comparative to the original method. All these prove that, in the whole acoustical spectrum, the HRTF’s variation has certain characteristics in different frequency range, which basically accords with human auditory system.

As for future work, we think that, the results of psychoacoustics should be attached more importance to the study in this field, including the analysis of certain HRTF and the design of its corresponding filter.

6. REFERENCES