COMPARATIVE STUDY ON HRTFS OF SIMPLIFIED DUMMY HEADS AND SUBJECT LISTENING EVALUATION

LIN Yang, FU Zhong-Hua, TANG Ling

School of Computer Science, Northwestern Polytechnical University, Xian, China 710129

The HRTFs of a standard Chinese dummy head and its different simplified versions are calculated and compared. Numerical comparisons are performed to analyze the influences of different head simplifications. Seven different simplified artificial heads are modeled based on the standard Chinese dummy head, where the head is simplified as ellipsoids of different sizes, including head height, head breadth, and head length as in GB/T 23461-2009[S], and the pinnae remain unchanged. All HRTFs are calculated using the boundary element method. The relation between numerical errors and listening distortions are highlighted.

Key words: HRTF; boundary element method; simplified artificial head; subject evaluation;

1. Introduction

HRTF (Head related transfer function) describe the transmission from sound source to two ears and reflect the overall effects caused by the diffraction (filtering) of head, torso, and pinnae [1]. They also depend on anthropometric structures and parameters and consequently individual.

Many researches have shown that HRTFs have intense relationship with human physiological structures and their sizes. In [1], it have been proved that the details of the shape of the pinna and head come into effect when frequency increases based on the measurements from 52 Chinese subjects. And in [2], it have been shown that the head size has effect on the HRTF frequency spectrum based on simulation and practical HRTF measurements. However, the actual influence of different heads on the HRTFs are not clear. The effect of different types of heads bring to HRTF is well worth discussing.

This paper compares the HRTFs of seven dummy heads and a standard dummy head called BHead210, which is built according to Chinese adult male [3]. The seven dummy heads are different simplified versions of the BHead210, where the heads are simplified as ellipsoids of different shapes by computer modeling and the pinnae remain unchanged. With these simulations, we can focus on the effects of the head on the HRTFs. The BHead210 is designed by Communication Acoustic Laboratory of Communication University of China. It has been scanned into computer using structural light [4] and the boundary element method (BEM) is used for calculation of its HRTFs [5].

We use the BEM to calculate the HRTFs of the dummy heads and analyze their differences from numerical and perceptual aspects. The relation between numerical errors and listening distortions are highlighted.
2. HRTF calculation

2.1 HRTF calculation based on the boundary element method

The assumption of numerical calculation of HRTFs is that only the surface characteristics of head are pertinent and the propagation through head structure is ignored [6]. The numerical calculation method involved in our work is the BEM. It is one of the commonly used methods to approximate the solution of acoustic radiation and scattering problem. It takes boundary integral equation as constrained equations and divides the acoustic boundary into tiny units. Then by integrating all the constrained integral equations on the units and solving the algebraic equations, it can obtain the acoustic pressure of any given observation point.

There are two main approaches in BEM which are commonly used: the Direct method (DBEM with a collocation formulation) and the Indirect method (IBEM) with a variational formulation. In general, the latter is more efficient for solving large problems since the matrices are symmetric. The boundary conditions can be applied on each node and can include the known pressure (Dirichlet boundary condition), velocity (Neumarm boundary condition), or the impedance/admittance boundary condition. Although it is a very powerful tool, BEM is very computationally expensive [9].

In the calculation of HRTF, acoustic source can be in any position in space. The boundaries are composed of heads and other biological structures which are simplified as ellipsoid head models. It causes a lot of time calculating acoustic resources’ HRTFs in different distances and azimuths. So reciprocity principle is used to reduce computational effort. The principle means that the acoustic system is a reciprocal system from the launch point to the receive point in linear acoustic range. The acoustic analysis module in Virtual.Lab Acoustic of LMS company (abb.VL) is used to calculate the ellipsoid heads’ HRTFs. We put the acoustic source in the geometrical center of the eardrum and set the field point mesh for the calculation.

2.2 The mesh models of the dummy heads

Since the purpose of this paper is to study the relation between different head shapes and the corresponding HRTFs, we need to preserve all physiological structures but the head shape so as to make the result more convincing. Therefore, based on the scanned model of the BHead210, we morph the head shape and keep the pinnae unchanged. Then with the carefully modified mesh models, the HRTFs are obtained using the BEM method mentioned above.

2.2.1 Design of the dummy heads

For the flexibility of computer morphing, we simplified the head of the BHead210 as a ellipsoid. It has been shown that the structure of human head is very similar to ellipsoid and their frequency characteristics are extremely close [10].

To make the research work more traceable, seven different ellipsoids are simulated according to GB/T 23461-2009[S], which specifies seven typical Chinese head shapes. It has been reported that the height, width and depth of human head have a profound influence on the HRTFs [11]. Therefore, for each simulated ellipsoid, we set the two equatorial radius of the ellipsoid as halves of the width and depth of the corresponding head, respectively, and set the polar radius of the ellipsoid as half of the head height [12]. The ellipsoids are built in Computer Aided Three-dimensional Interactive Application (CATIA). It is a multi-platform CAD/CAM/CAE commercial software suite developed by the French company Dassault Systems. In CATIA, a tool called “sweep” is used to draw the ellipsoids. The simulated seven ellipsoids are named respectively as round-high head (RH), round-right head (RR), middle-round head (MR), middle-high head (MH), round-super-high head (RSH), super-round-high head (SRH), and super-round-super-high head (SRSH). The detailed parameters of these head sizes is referred in GB/T 23461-2009[S].
To obtain complete simplified dummy heads, we need to “stick” pinnae model to the simulated ellipsoids. The original pinna model is from the BHead210, which is an “average ear” model (or standard auricle) designed and manufactured based on the most representative auricle shapes of Chinese adults [3]. The left pinna is scanned into computer using structure light [4]. Then in the computer, the pinna mesh model is cut from its cylinder base, which is for fixed embedding into the ear canal of the BHead210.

One important problem to “stick” the pinna mesh model to the ellipsoids is to correctly locate the position of each pinna. The corresponding tragion positions of the seven typical Chinese heads are clearly defined in GB/T 23461-2009[S]. We define five reference points to correctly locate the pinna as shown in figure 1. The point A refers to the tragion point. Suppose X-axis points to the front, Y-axis points to the right, and Z-axis points to the upside. Across A, we set two planes that are paralleled to X-Y plane and Y-Z plane, respectively. The two planes intersect with the contour of the base of the pinna mesh model to get four points, which are labeled as B, C, D, and E, respectively. With the four reference points, the pinna mesh model is correctly embedded into each ellipsoid. The fusion of one ellipsoid with the pinna model is figure 1.

![Figure 1. Fusion of the ellipsoid models and the pinna models](image)

Another thing needs to be noted is that the HRTFs considered in this paper is measured with closed ear canal. Since the pinna model contains part of the ear canal, we cut it before doing calculations. The final pinna model is fused with each ellipsoid using DAVID3 software.

### 2.2.2 Mesh modification

Mesh modification are necessary to make the size of each mesh suitable for the HRTF calculation. The proper length of each triangle mesh should be no longer than 1/4-1/6 of the wavelength of max frequency acoustic wave [13]. The important information of the HRTFs exists within frequencies between 0-12kHz, hence the max length of the mesh should be less than 7.08mm. However, both the ellipsoid models and the pinna mesh model are not qualified. Each ellipsoid has approximately 25,000 triangular elements with maximum length of approximately 19.50mm, so it needs subdividing. The scanned pinna model has approximately 14,000,000 elements with maximum edge of approximately 0.19mm, so it needs decimating.

1. **Mesh reprocessing.** The mesh of each ellipsoid is subdivided by splitting every triangular element into three smaller triangles. And the mesh of the pinna model is decimated by merging adjacent small element. As a result, the numbers of the triangular elements of both the ellipsoid mesh and the pinna mesh are no more than 60,000 and their edges are all less than 7.08mm. Thus, the mesh grids are appropriate for the BEM calculations.

2. **Mesh refinement.** When the ellipsoid and the pinna models are fused in DAVID3, the joint parts will appear sharp edges and some dispersed nodes. What’s more, the cutting and fusing make the triangular elements unequally distributed. So we firstly clear those dispersed nodes, secondly smooth the joint parts, and thirdly, redistrict the mesh grid by distribute the elements equally. All the procedures are accomplished in RapidForm.

3. **Mirror the left head to the right head.** For simplicity, after the mesh refinement, the left head
with pinna is mirrored to the right head, so as to reduce the deviations between the left and right sides. The final mesh modification are shown in figure 2.

Figure 2. Partial meshes of the ellipsoid (Left), the scanned pinna (Middle), and the final mesh (Right).

3. Comparisons of the calculation results

3.1 HRTF comparisons of the BHead210 and the simplified heads

The spatial directions in all the following experiments are denoted in form of (elevation, azimuth). For example, the direct front is labeled as (0°, 0°), and the right hand direction is (0°, 90°). Due to the page limitation, we only compare the HRTFs of left ear on directions of (0°, 0°), (0°, 90°), (0°, 180°), and (0°, 270°). The distance of the sound source to the center of the heads is set to 1m.

We firstly compare the HRTFs of the BHead210 model [4], the smoothed model of the BHead210, and an average simplified model. The smoothed model is obtained by smoothing off the facial structures of the BHead210 using RapidForm. The average simplified model contains a ellipsoid head with average size of the seven typical heads and the same pinnae as the BHead210. The comparisons are shown in figure 3. BHead210 and the smoothed model are shown in figure 4.

Figure 3. Left-ear HRTF comparisons of the Bhead210, the smoothed BHead210, and the simplified head

Figure 4. BHead210 (Left) and the smoothed model (Right)
It’s clear that the HRTFs of these heads are basically matched especially in the low and medium frequencies. In 0° azimuth, the variations are less than 3dB in 0-5kHz but the error become larger in 9kHz. In 90° azimuth, the mean error is large in 6kHz and there are more differences in high frequency. In 180° and 270° azimuths, the curves are matched very well. The biggest variation appears in 90° azimuth, where the reflection and diffraction of the acoustic wave are quite complicated. Above all, the simplified head shows the same characteristics as the other heads.

3.2 HRTF comparison of different simplified head

Now the HRTFs of different simplified heads are compared, both the far-field and the near-field cases are considered, where the distances from the sound source to the head center are set to 1m and 0.2m respectively. The results are shown in figure 5 for the far-field case, figure 6 for the near-field case and figure 7 for the azimuth 90° case, respectively.

![Figure 5](image1)

**Figure 5.** HRTFs of 1m in horizontal plane with azimuth angle of 0°, 90°, 180°, 270°

![Figure 6](image2)

**Figure 6.** HRTFs of 0.2m in horizontal plane with azimuth angle of 0°, 90°, 180°, 270°

From figure 5 we can see:

Differences among HRTFs of seven dummy heads are small. They have basically the same form especially in the low frequency. The differences are mainly in the magnitude which result from every detail of the structures. In (0°, 90°), the deviation are bigger than 5dB in 5kHz and higher frequency. The absolute error grow with the increasing frequency. In (0°, 0°), (0°, 180°), (0°, 270°), little deviation is observed which remain less than 1dB in the whole frequency domain. In azimuth of 0°, 180° and 270°, elevation of 30° and 60° the regulars are similar.

From figure 6 we can see:

In (0°, 0°), (0°, 180°), (0°, 270°), distortion among HRTFs of different heads remain small with the deviation less than 1dB. The seven dummy heads follow basically the same trend. In azimuth of
Figure 7. HRTFs of 0.2m and 1m with azimuth angle of 90°, elevation of 30°, 60°, 0°, 180°, 270°, elevation of 30° and 60°, regulars are similar. In (0°, 90°), the deviation is bigger than 5dB in 8kHz and higher frequency. The pinnae valleys remain almost the same position with different magnitude in all frequency domain.

From figure7 we can see:

In azimuth of 90°, pinnae valley position have greater distortion due to different sizes of head. In (30°, 90°) of 1m, the most severe distortion happen in about 9kHz. But in (60°, 90°) of 1m, distortion is less than 2dB mainly in the whole frequency domain. In (30°, 90°), (60°, 90°) of 0.2m, distortion is less than 8dB in about 7kHz.

From the above analysis, we can come to a conclusion that different sizes of typical heads have influences on the azimuth of 90° but not much for the other three azimuth namely 0°, 180°, 270°. Distortion happen in azimuth of 90° because of the reflection and diffraction of the acoustic wave. When the sound source is observed from the opposite side of ear, the wave path of sound is more complex as a consequence of head’s shadow. What’s more, due to the influences of the head’s shadow, the noise-signal ratio is lower which may bring some deviation.

4. Subject evaluation

From the above numerical comparisons, the HRTF differences between the dummy heads are shown. It is worthy of examining the subject perception with the corresponding HRTFs. So we conduct the following preliminary subject evaluation.

4.1 Experiment setup

The far-field HRTFs of the seven simplified heads and the BHead210 are involved. We simulate a white gaussian noise circling clockwise around the listener’s head from 0° to 350° with a step of 10°. Therefore there are totally 36 azimuths on the horizontal plane. Since the HRTFs in this paper are within 0-12kHz, the sampling rate of the white noise is set to 24kHz. The length of the original white noise is 0.5s, then binaural signals are obtained by convoluting the original noise with each pair of the HRIRs of each azimuth. Between each azimuth, a short silence of 0.5s is insert into the binaural signal. With the eight heads (7 simplified + BHead210), we get eight binaural segments.

The listening experiments are conducted in a quite listening room. Sennheiser HD600 headphone is used for sound reproduction. There are 24 students (19 males and 5 females) involved in the evaluation. We choose more male listeners because the dummy head sizes are from Chinese male adults. Before the formal test, each student is required to listen to a real-recorded binaural audio clip so as to be familiar with the listening condition. Then the eight audio segments are randomly delivered to the headphone. Note that the content of the test segments are told to the listeners but
the segment order. Each listener is required to select the best segment and score each of the eight segments with a 5-score rule, where from 5 to 1, each means very-good, good, general, poor, and very-poor, respectively.

4.2 Experimental results and analysis

<table>
<thead>
<tr>
<th>Head Style</th>
<th>RH</th>
<th>RR</th>
<th>MR</th>
<th>MH</th>
<th>RSH</th>
<th>SRH</th>
<th>SRSH</th>
<th>BHead210</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best Option</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>Average Score</td>
<td>3.11</td>
<td>3.33</td>
<td>3.69</td>
<td>3.72</td>
<td>3.92</td>
<td>3.64</td>
<td>3.66</td>
<td>4.06</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.08</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

From table 1 we can see the average scores of eight segments remain almost unanimous with a variance of 0.08. Although most listeners choose BHead210 as their priority followed by other dummy heads, BHead210 only has a slightly higher score.

Due to the small differences between eight head models, we come to a conclusion that the ellipsoid head model can be a simplification of BHead210. And different sizes of the ellipsoid have not much influences on reproducing the sound field.

5. Conclusion

It can be concluded from numerical results that ellipsoid head model is acceptable as a simplification of head model with facial features. The simplified head shows the same characteristics as the other heads. Different sizes of typical heads have influences on the azimuth of 90° but not much for the other three azimuth namely 0°, 180°, 270°.

The subject evaluation reveals that there is not much differences among the eight models. B-H210 is chosen to be the best one. It may be a better model to reproduce the sound. But the simplified head model is still a good version for simplification. And the sizes of different types of heads have not much influences on the subject evaluation. What’s more, the relation between listeners’ head types and their subject evaluation results need further discussion.

6. Acknowledgments

The authors would like to thank teacher Na Qi from Institute of communication acoustics of Communication University of China for the experimental data of China standard head model-the BHead210 she provides.

REFERENCES


