CREATING THE SYDNEY YORK MORPHOLOGICAL AND ACOUSTIC RECORDINGS OF EARS DATABASE

P. Guillon, R. Zolfaghari, N. Epain, A. van Schaik and C.T. Jin

C. Hetherington, J. Thorpe and A. Tew

THE UNIVERSITY OF SYDNEY

THE UNIVERSITY of York

Presented by Craig Jin
OUTLINE

› Introduce a “Holy Grail” in 3D Audio
› Describe the SYMARE database
› Genesis of the database
› Preliminary validation of the data.
The Problem: Customize 3D audio for listeners based on 2D photos.
Head-Related Impulse Responses (HRIRs): are acoustic filters for earphones that produce 3D Audio.
Acoustically measuring HRIRs can be difficult and time-consuming. It would be better if we could predict HRIRs from morphology.
Listener’s ears differ and so do their HRIRs

HRTF is the HRIR Spectrum
The SYMARE database is a joint effort between the universities of Sydney and York.

The database contains morphological and acoustic data for 60 subjects. For each subject, the database provides:

- The subject’s measured HRIRs for approximately 400 directions.

- High-resolution surface meshes of the head and combined head and torso at varying resolution suitable for acoustic simulation. In the meshes with the highest resolution, the size of the elements is on the order of 1 mm.

- Two sets of HRIRs simulated by applying the Fast-Multipole Boundary Element Method (FM-BEM) to the head-only and head-and-torso meshes.
Acquiring Morphological Data

MRI scanning → DICOM images → Surface mesh
Generation of Surface Meshes

› For each subject, 2 MRI scans were performed: a high resolution scan of the head and a lower resolution scan of the head and torso.

› The surface meshes were first obtained by extracting iso-surfaces from both sets of MRI images.

Some artifacts are present inside and outside the mesh. These artifacts were removed manually.
Generation of Surface Meshes

› The high resolution head-only meshes were aligned and merged with the low-resolution torso mesh.

› The torsos’ surfaces were smoothed.

Before smoothing the torso

After smoothing the torso
Generation of Surface Meshes

› The meshes were re-meshed so that the triangular faces are as equilateral as possible.
› The heads (except the ear pinnae) were then smoothed.

Triangle surface elements satisfy:
1. maximal to minimal edge length ratio < 5
2. Smallest angle > 15°
3. Largest angle < 160°
The heads were separated from the torso to create head-only meshes.

The Head-only and Head-and-Torso meshes were then re-meshed at various resolutions.

High-resolution Head-and-Torso mesh (~150k elts., $f_{\text{max}}=16\text{kHz}$)

Detail of a High-resolution Head-only mesh (~130k elts., $f_{\text{max}}=20\text{kHz}$)
Terracotta Army in Xi’an China
The subjects HRTFs were simulated using Coustyx, a Fast-Multipole Boundary Element Method (FM-BEM) software.

The size of the mesh elements determine the maximum frequency at which the BEM simulation can be done:

\[
f_{\text{max}} = \frac{c}{6 \ l_{\text{max}}}
\]

where \( c \) is the speed of sound (~ 340 ms\(^{-1}\)) and \( l_{\text{max}} \) is the maximum edge length.

For each subject, two BEM simulations were performed:
- Using a low-resolution Head-and-Torso mesh, up to 5.6 kHz.
- Using a high-resolution Head-only mesh, up to 17 kHz.
Use Reciprocity: $P(E,S_i) = P(S_i,E)$ to derive all HRTFs in one go.

Source placed in the ear; apply equivalent velocity boundary condition.
We calculated the Interaural Time Delays (ITDs) for both the measured and simulated sets of data (Head-and-Torso meshes). We then applied Kuhn’s model to estimate the corresponding effective head diameters.

The difference in the effective diameters is on the order of a few percent.
We applied Middlebrooks’ frequency scaling technique to estimate the scaling factor that minimises the mismatches between the measured and simulated HRTF data.

Frequency scaling factors for Subjects 1 to 10

<table>
<thead>
<tr>
<th>Subject</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Head only</strong></td>
<td>1.053</td>
<td>1.031</td>
<td>1.099</td>
<td>1.042</td>
<td>1.064</td>
<td>1.020</td>
<td>1.020</td>
<td>1.075</td>
<td>1.020</td>
<td>0.990</td>
</tr>
<tr>
<td><strong>Head &amp; T.</strong></td>
<td>1.021</td>
<td>1.010</td>
<td>1.031</td>
<td>1.010</td>
<td>1.010</td>
<td>1.010</td>
<td>1.020</td>
<td>1.031</td>
<td>1.020</td>
<td>1.053</td>
</tr>
</tbody>
</table>
A Spatial Frequency Response Surface (SFRS) is defined as the magnitude of the HRTFs for every direction at a given frequency. It characterises the directivity of the ear at this frequency.
We compared the data obtained via BEM simulation using the Head-only meshes with that obtained using the Head-and-Torso meshes. The figure below compares Subject 1’s SFRS’s at 2kHz.

Important HRIR features are present only in the data corresponding to the head-and-torso mesh.
Comparison of SFRS’s corresponding to the measured and simulated data sets (the Head-and-Torso data is used below 5 kHz):

Many of the features of the SFRS corresponding to the measured data set are present in the BEM-simulated data set.
Spatial correlation between the SFRS’s corresponding to the measured data and that corresponding to the simulated data, as a function of the frequency:

The simulated HRIR data captures some of the individual characteristics of the outer ear acoustics.
Thanks for your attention.