Perceptually motivated analysis of numerically simulated head-related transfer functions generated by various 3D surface scanning systems

Dinakaran, Manoj; Brinkmann, Fabian; Harder, Stine; Pelzer, Robert; Grosche, Peter; Paulsen, Rasmus Reinhold; Weinzierl, Stefan

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Numerical simulations offer a feasible alternative to the direct acoustic measurement of individual head-related transfer functions (HRTFs). For the acquisition of high quality 3D surface scans, as required for these simulations, several approaches exist in the market. We systematically analyze the variations between different approaches and evaluate the influence of the accuracy of 3D scans on the resulting simulated HRTFs. To assess this effect, HRTFs were numerically simulated based on 3D scans of the head and pinna of the FABIAN dummy head generated by 6 different methods. These HRTFs were analyzed in terms of interaural time difference, interaural level difference, energetic error in auditory filters and by their modeled localization performance.

From the results, it is found that a geometric precision of about 1 mm is needed to maintain accurate localization cues, while a precision of about 4 mm is sufficient to maintain the overall spectral shape.

**Motivation**

1. In recent years, several approaches have been proposed with a focus on increasing the accuracy of the simulated HRTFs by acquiring high quality 3D scans of head and pinna. Different techniques for the acquisition of 3D surface scans exist such as MRI scanners, structured light scanners, laser scanners, infrared scanners, stationary scanners, hand held scanners, or by using mobile camera pictures [1].

2. Each of them provides a different resolution and its accuracy directly affects the quality of the numerically simulated HRTFs which are subject to research.

3. Here, we systematically analyze the accuracy of 3D surface scans obtained by different approaches and study their influence on the resulting HRTFs by means of interaural time difference (ITD), interaural level difference (ILD), energetic error in equivalent rectangular bandwidth (ERB) auditory filters, and their simulated localization performance.

4. To isolate the influence of the scanning method on the HRTF, the different scanning methods were evaluated against a high resolution structured light scan (ground truth) which showed a very good agreement to its acoustically measured counterpart in an auditory study [2].

**Acquisition of Meshes using different scanning systems**

We acquired 3D surface scans of the head and pinna of the FABIAN dummy head by using 6 different methods (cf. Fig. 1).

- a) GOM ATOS® (GOM-Ref): Stationary, structured light scanner (0.01 mm point resolution)
- b) Artic Space Scanner (SPY): Hand-held structured-light scanner, scanning at a working distance of 0.2 m to 0.3 m (0.05 mm point resolution)
- c) Canfield Vectra M3 (CAN): Stationary, stereo photogrammetry technology scanner (500 points per second at 1 m resolution)
- d) Microsoft Kinect (KIN): Low cost IR scanner with a working distance between 0.7 m and 3.5 m
- e) Autodesk 123D catch (123D): Mobile application which allows the user to scan a 3D model from at least 5 to 8 overlapping photos
- f) The Python Photogrammetry Toolbox (PPT): An open source tool which has a pipeline to construct a 3D model from a set of photos.

**Alignment, Re-meshing & HRTF Simulations**

1. In the first step, interaural axis and interaural center of the GOM-Ref mesh were aligned to the origin of coordinates.
2. Then, the remaining FABIAN surface scans were then aligned with respect to GOM-Ref using the least square closest matching algorithm (SPY, CAN, 123D, PPT).
3. A priori mesh grading algorithm (resulting in non-uniform meshing) was deployed according to Ziegelwanger et al. which result in scans where the mesh element size was reduced by the factor of 2 (KIN).
4. Two different models were generated for each scanning method: One for the left pinna (with small mesh elements at the ear surface) and the right (with large elements at the right); and one for the right pinna.
5. The target lengths used were 1 mm to 10 mm, which resulted in around 20,000 elements per mesh.

**Conclusion**

- A high precision of about 1 mm is needed when capturing the pinnae geometry to assure accurate localization cues. This criterion was met only by the SPY and CAN scanning methods.
- However, the overall coloration showed to be below 1 dB, even for geometric errors of up to 4 mm, which occurred for the KIN method.
- The remaining methods (123D & PPT) showed geometric deviation of up to 5 mm and slightly larger coloration of up to 1.5 dB.

**Table 1: Geometric difference in mm (Ref 1), change in polar error (in degree), quadrant change (in %) and Average spectral difference (in dB)**

<table>
<thead>
<tr>
<th>Method</th>
<th>Scan</th>
<th>X-Err (°)</th>
<th>Y-Err (°)</th>
<th>Z-Err (°)</th>
<th>X-Err (°)</th>
<th>Y-Err (°)</th>
<th>Z-Err (°)</th>
<th>X-Err (°)</th>
<th>Y-Err (°)</th>
<th>Z-Err (°)</th>
<th>X-Err (°)</th>
<th>Y-Err (°)</th>
<th>Z-Err (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPY</td>
<td>0.14 (0.24)</td>
<td>0.17 (0.20)</td>
<td>0.50</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAN</td>
<td>0.66 (0.90)</td>
<td>0.66 (0.54)</td>
<td>0.50</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>123D</td>
<td>1.53 (0.75)</td>
<td>1.25 (0.18)</td>
<td>1.50</td>
<td>2.50</td>
<td>2.50</td>
<td>2.50</td>
<td>2.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPT</td>
<td>3.71 (1.27)</td>
<td>3.68 (1.15)</td>
<td>1.50</td>
<td>3.50</td>
<td>3.50</td>
<td>3.50</td>
<td>3.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Fig 1:** ABSOLUTE DIFFERENCES IN THE HRTF MAGNITUDE SPECTRA AVERAGED ACROSS THE ENTIRE FREQUENCY RANGE FOR BOTH EARS (LEFT COLUMN) (in dB), ABSOLUTE DIFFERENCES IN ILD (MIDDLE) (in dB) AND ABSOLUTE DIFFERENCES IN ITD (RIGHT) (in μs) WITH RESPECT TO GOM-Ref (SPY 1st row, CAN 2nd row, KIN 3rd row, 123D 4th row and PPT 5th row).