www.akustik.rwth-aachen.de
DOI: 10.18154/RWTH-2020-06760

Technical Report

A High-Resolution Individual 3D Adult Head and Torso Model for HRTF Simulation and Validation: 3D Data

Hark Braren, Janina Fels

Teaching and Research Area of Medical Acoustics,
Institute of Technical Acoustics, RWTH Aachen University, Germany
Corresponding author: janina.fels@akustik.rwth-aachen.de

Abstract: Simulation of head-related transfer functions (HRTFs) using the boundary element method (BEM) is a well established practice for calculating the direction-dependent response of the pinna, head and torso geometry to incoming sound events. This report describes the process of capturing and processing the high-resolution 3D model of a male adult, using a structured-light scanner. Special care was taken to accurately represent the pinna with up to sub-millimeter resolution in areas of high curvature. Along with the 3D data, an HRTF measurement of the same individual is available, described in the accompanying report.

Keywords: Binaural Technologies, Anthropometric Model, 3D-Scan

1 Introduction

In this report the process of capturing and processing an individual high-resolution 3D model of an adult male is described. The capturing hardware and its capabilities are presented. The processing pipeline used in the creation of the model is described and finally the accuracy of the resulting mesh is analyzed. The resulting 3D mesh is available under doi.org/10.18154/RWTH-2020-06760. An accompanying dataset and report of an HRTF measurement of the same person is also available [1].

2 3D-Scan

The 3D model of an adult male individual was obtained using a handheld structured-light scanner (Artec Space Spider, Artec Group, Luxembourg) [2]. It works by projecting a straight line pattern onto the scanned surface and capturing the deformed lines using 3 cameras with known relative positions. The known position of the cameras combined with the deformation of the projected lines on non-flat surfaces is used to construct individual frames of point cloud data. Each frame contains approximately 50,000 points from which about 100,000 faces are derived.

Figure 1 shows a single frame of a scan of the right ear. The scanner is moved along the targeted surface to capture larger areas. Individual frames are recorded at up to 8 frames per second. Redundant information within consecutive frames (overlapping areas) is used to automatically align the frames and form a large 3D point cloud. Figure 2 shows the resulting point cloud from 10 such frames. Since this process is error prone, it is also possible to realign the frames during processing, if the initial alignment went wrong.

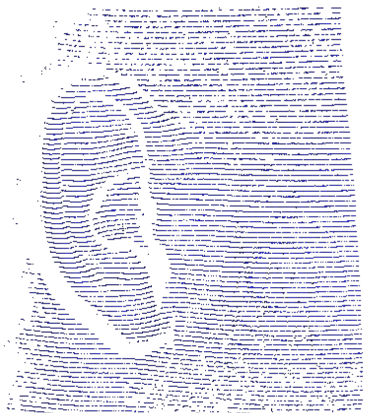

Figure 1: Point cloud data from a single frame of the right ear showing the projected line pattern.

To scan the whole head and upper torso, multiple scans, each containing a couple hundred frames, are superimposed and aligned with one another as shown by the three differently colored scans in Figure 3. This method allows to capture the head and especially the pinna geometry from different angles, which is particularly important to accurately capture the complex geometries inside of the pinna. The structured light scanner can not see through geometries like for instance a magnetic resonance imaging (MRI) scanner. It is therefore necessary to capture the ear from at least 5 different angles, including from behind, to finally capture a complete and hole-free 3D model.

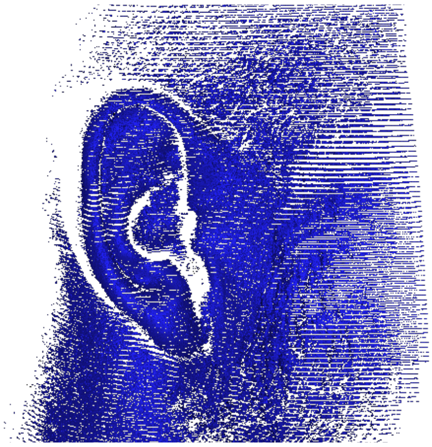

Figure 2: Point cloud data from 10 consecutive frames of the right ear.

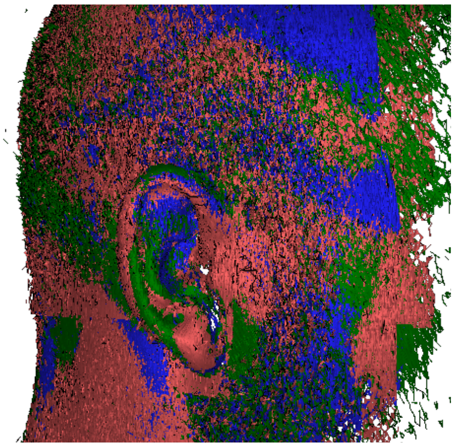

Figure 3: Overlay of three individual scans indicated by red, blue and green surface colors. Each scan is captured at a different view angle to fill in information missing from the others.

The aligned scans are then positioned more precisely, this time not only using subsequent frame information, but the data from all frames intersecting at a given location. This process is called global registration in Artec Studio [3] and constitutes a mandatory step before combining the point clouds into a surface model. For the current model, Artec Studios *Sharp Fusion* algorithm with a resolution set to 0.3 mm was used. It generates a surface model as shown in Figure 4 where small holes are automatically filled. Unwanted small details (e.g. due to a beard) are visible due to the high resolution of the scans and will be addressed in post processing.

2.1 Post Processing

Before exporting the resulting fusion-geometry as an *.stl* file from Artec Studio, the geometry is simplified using the mesh simplification algorithm. The maximum deviation from the original geometry was set to 0.02 mm. This helps keeping the number of vertices and thus the resulting file size in a manageable range whilst retaining the high resolution mesh where it is needed, namely at the ears.

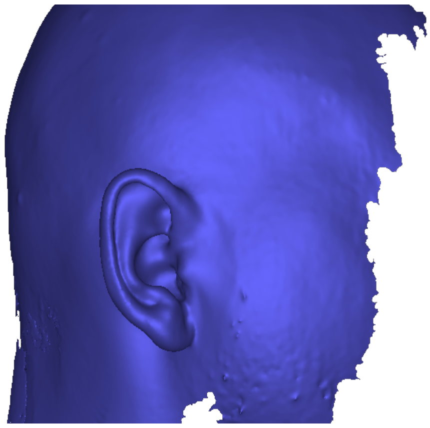

Figure 4: Resulting partial surface model from a sharp fusion of the three scans shown in Figure 3.

All further processing took place in Autodesk's Meshmixer [4] Version 3.5.474. Here the following steps were applied:

- Artifact removal,
- mesh alignment,
- creating a watertight model,
- anonymization, and
- further mesh simplification.

Artifact Removal is needed where multiple Scans were not correctly aligned before applying the fusion algorithm in Artec Studio. This generally happens when the scanned participant moves slightly during or in between scans. Thus the assumption of a rigid body no longer holds true. These artifacts manifest as double or multiple layers in close proximity as seen in Figure 5 and need to be removed manually to get a single surface water-tight model. Such effects were most prominently visible on the upper body, which was smoothed over anyhow in subsequent steps to get rid of T-shirt wrinkles, which do not effect the HRTF and only produce additional mesh nodes. So they could be taken care of by mesh replacement methods such as generous removal plus hole filling and smoothing.

Mesh Alignment refers to the process of centering the head in the global coordinate system as presented in Figure 6. HRTFs are generally defined in spherical coordinates, with the center of the coordinate system coinciding with the center of the interaural axis. The azimuth angle φ starting at $\varphi = 0°$ in the frontal direction going counterclockwise. Thus $\varphi = 90°$ is to the left. In Cartesian coordinates, this means that the x-axis points forward, the y-axis goes through the left ear and the z-axis points upwards. Since slight

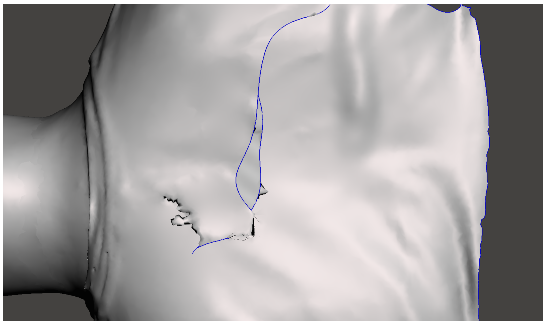

Figure 5: Mesh artifacts in the shape of multiple layers resulting from slight movements and thus surface shape deviations between overlayed scans on the back.

variations in posture can affect the frontal direction of the head and the alignment of the head to the torso quite drastically, care was taken to capture a natural posture looking straight forward and a best-effort approach was used when aligning the captured 3D model.

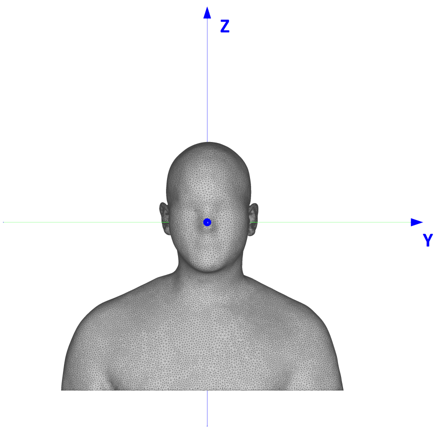

Figure 6: Alignment of the mesh with x-axis to the front, the interaural-axis as y-axis and the z-axis point upward.

A Watertight Model is a model with a closed surface, i.e. without holes or cracks in the surface. This is a mandatory feature for use in simulations. Most of the surface is free of holes after the previous post processing steps, apart from a hole in the bottom of the model, where the torso scan ends. This hole is filled by cutting off lower parts of the torso with a planar cut and flat-filling the hole. The remaining torso needs to be large enough, such that the low frequency torso reflection effects can still be resolved in simulations. At higher frequencies the torso is not as important, thus fine details in the torso geometry are not needed, and it can be smoothed over in order to reduce the size of the resulting mesh files.

Anonymization refers to the smoothing of characteristic facial features, that might aide identification. For this process, Meshmixers robust-smoothing brush was applied to the facial region, particularly the mouth, nose and eyes area, whilst keeping the general geometry of the head as close to the original as possible.

Mesh Reduction refers to the previously mentioned geometry simplifications, namely removal of unneccesary and unwanted details especially in the torso region by mesh replacement and mesh reduction methods. The latter utilizes the mesh-reduction brush.
In acoustic simulations, a trade-off between accuracy and solving time is common practice. As a general rule of thumb, a mesh accuracy of $\lambda/6$ leads to sufficiently accurate results as long as the geometry is accurately represented. Since the pinna has the greatest effect on high frequencies in HRTFs, it is kept at a high resolution, while the smooth shapes of the remaining geometry are meshed at a lower accuracy. Table 1 presents an overview of the different mesh sizes and the overall number of vertices of the final model.

Table 1: Mesh statistics.

	Edges [mm]			
	max	mean	Vertices	Faces
Pinna	5.9	0.9	6775	13442
Head	10.6	4.4	22419	44590
Torso	11.8	4.4	22769	45292
Overall	11.8	3.4	44699	89394

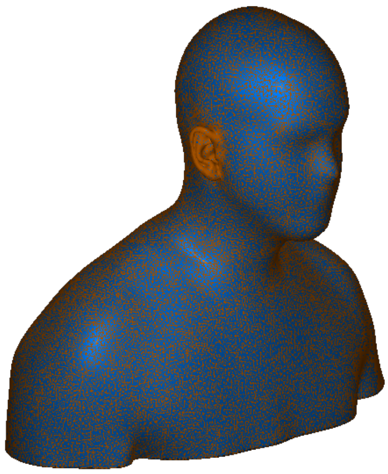

Figure 7: Final mesh with surfaces indicated in blue and mesh edges in orange showing the high resolution pinna shape while keeping a lower resolution for the remaining mesh.

3 Summary

The process of generating an individual 3D-scan of a male adult individual including the necessary post-processing of the data has been detailed. Facial features of the resulting 3D model have been removed and the resulting mesh was optimized to reduce the number of mesh nodes by reducing the resolution on the large surfaces of the torso , whilst retaining a sub-millimeter resolution in curved areas of the pinna, which are important for high-frequency HRTF calculations.

Supplementary Material and Information

The 3D-dataset can be downloaded from doi.org/10.18154/RWTH-2020-06760.

All datasets related to this report are published with consent and a review by the medical ethics committee at the RWTH Aachen University ([EK 218/18]).

The data collection is in line with the Declaration of Helsinki [5]. Participation was voluntary and subject to extensive information; personal data is treated confidentially (cf. DSGVO, [6]) and physical or mental harm to the participants is excluded.

As accompanying material, an HRTF measurement of the same participant with measurement description[1] is provided under doi.org/10.18154/RWTH-2020-06761.

An example of a BEM calculation using the provided model can be found in COMSOL's application library, Application ID: 75011 [7].

References

[1] H. Braren and J. Fels, "A High-Resolution Individual 3D Adult Head and Torso Model for HRTF Simulation and Validation: HRTF Measurement," Medical Acoustics Group, Institute of Technical Acoustics, RWTH Aachen, 2020. DOI: 10.18154/RWTH-2020-06761.

[2] Artec Group, *Artec Space Spider*. [Online]. Available: https://www.artec3d.com/portable-3d-scanners/artec-spider.

[3] ——, *Artec Studio*. [Online]. Available: https://www.artec3d.com/3d-software/artec-studio.

[4] Autodesk, *Meshmixer*. [Online]. Available: http://www.meshmixer.com/.

[5] "World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects.," eng, *Bulletin of the World Health Organization*, vol. 79, no. 4, pp. 373–374, 2001.

[6] Verordnung (EU) 2016/679, "Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung)," 2016.

[7] COMSOL Multiphysics®, "Head and Torso HRTF Computation," *Comsol Application Gallery*, Application ID: 75011, [Online]. Available: https://www.comsol.com/model/head-and-torso-hrtf-computation-75011.