PERCEPTUALLY MOTIVATED ANALYSIS OF NUMERICALLY SIMULATED HEAD-RELATED TRANSFER FUNCTIONS GENERATED BY VARIOUS 3D SURFACE SCANNING SYSTEMS

Manoj Dinakaran[1,2], *Fabian Brinkmann*[2], *Stine Harder*[3], *Robert Pelzer*[2], *Peter Grosche*[1], *Rasmus R. Paulsen*[3] and *Stefan Weinzierl*[2]

Huawei Technologies, German research center, Munich, Germany[1], Technical university of Berlin, Audio Communication Group, Germany[2] and Technical university of Denmark, Lyngby, Denmark[3]

Abstract

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 this paper, 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 with 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

- In recent years, several approaches have been proposed with a focus on increasing the accuracy of the simulated HRTF 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].
- Each of them provides a different resolution and its accuracy directly affects the quality of the numerically simulated HRTFs which are subject to research.
- 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.
- 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 correspondent in an earlier 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-I *(GOM-Ref):*** Stationary, structured light scanner (0.01 mm point resolution).
b) **Artec Spider *(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 scanning at a working distance 1 m (1 mm point resolution).
d) **Microsoft Kinect *(KIN):*** Low cost IR scanner with a working distance between 0.5 m to 1 meter.
e) **Autodesk 123D catch *(123D):*** Mobile application which allows the user to get a 3D model from at least 5 to 6 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.

(- iPhone 6 mobile was used to take photos for 123D and PPT method)

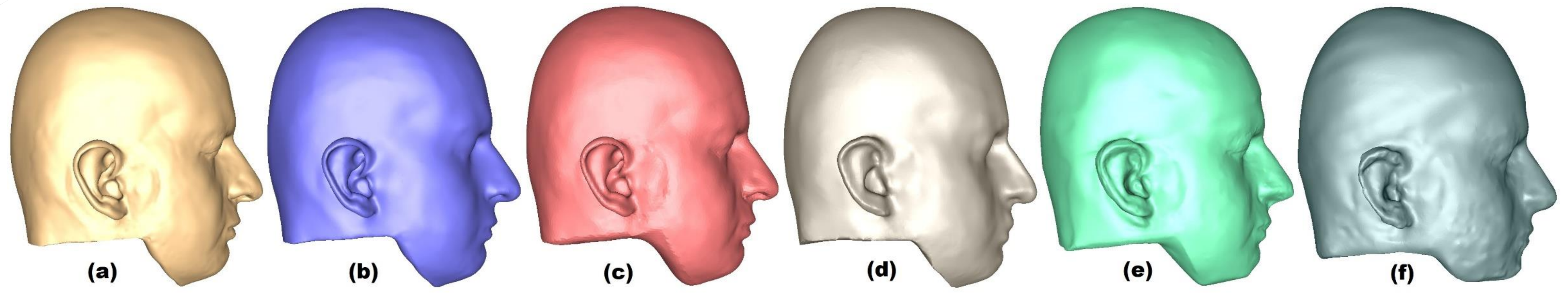

Fig 1: FABIAN 3D Surface Scans using (a) GOM ATOS-I Scanner ***(GOM-Ref)***, (b) Artec Space Spider Scanner ***(SPY)***, (c) Canfield Vectra M3 scanner ***(CAN)***, (d) Kinect scanner ***(KIN)***, (e) Autodesk 123D ***(123D)***, and (f) PPT ***(PPT)***.

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 iterative closest point (ICP) algorithm from the surface manipulation and transformation toolkit (SUMATRA) [3].
3) A priori mesh grading algorithm (resulting in non-uniform meshing) was deployed according to Ziegelwanger et al. which result in increases in size of the mesh element with respect to the distance from the ear [4].
4) Two different models were generated for each scanning method: One for the left pinna (with small mesh elements at the left ear, and 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.
6) For numerical HRTF simulation, the Mesh2HRTF implementation of the 3-dimensional Burton-Miller collocation BEM was used [5].

Difference in terms of Geometrics and Localization performance

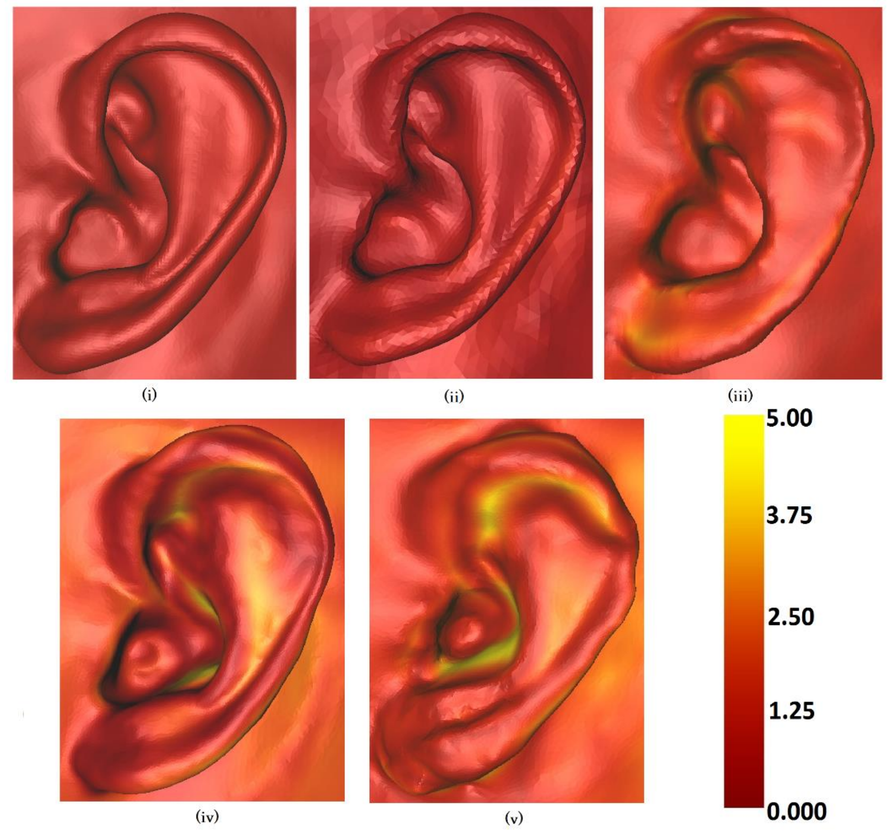

Fig 2: Geometric difference of (i) ***SPY***, (ii) ***CAN***, (iii) ***KIN***, (iv) ***123D*** and (v) ***PPT*** with respect to ***GOM-Ref*** (in mm)

Scans	$X_1(\mu,\sigma)$	$X_2(\mu,\sigma)$	X_3 *(max)*	X_4 *(max)*	X_5 *(max)*	X_6 *(max)*
SPY	0.14 (0.24)	0.17 (0.29)	0.50	0.75	0.75	0.75
CAN	0.66 (0.80)	0.66 (0.54)	0.66	0.90	0.90	0.80
KIN	1.53 (0.28)	1.50 (1.08)	1.50	2.50	1.25	1.25
123D	1.98 (1.42)	2.10 (1.41)	5.00	3.75	2.50	3.75
PPT	1.77 (1.72)	1.68 (1.56)	5.00	5.00	3.75	5.00

Table 1: Geometric difference in mm ($\boldsymbol{\mu}$ → mean, $\boldsymbol{\sigma}$ → standard deviation, ***max*** → maximum difference). $\boldsymbol{X_1}$→ head, $\boldsymbol{X_2}$→ head without pinna, $\boldsymbol{X_3}$→ Concha, $\boldsymbol{X_4}$→ Antihelical fold, $\boldsymbol{X_5}$→ Antihelix and $\boldsymbol{X_6}$→ Fossa.

Measures/ Scans	*SPY*	*CAN*	*KIN*	*123D*	*PPT*
PE	$< 0.7^{\circ}$	$< 0.7^{\circ}$	6°	11°	12°
QE	< 0.4	< 0.4	4	6	6
$\mathbf{S_d}$	< 0.5	< 0.5	< 1	1-2	1-2

Table 2: **PE, QE**→ increase in polar error (in degree), quadrant error (in %) and $\boldsymbol{S_d}$→ Average spectral difference (in dB)

Difference in terms of HRTF magnitude, ILD and ITD

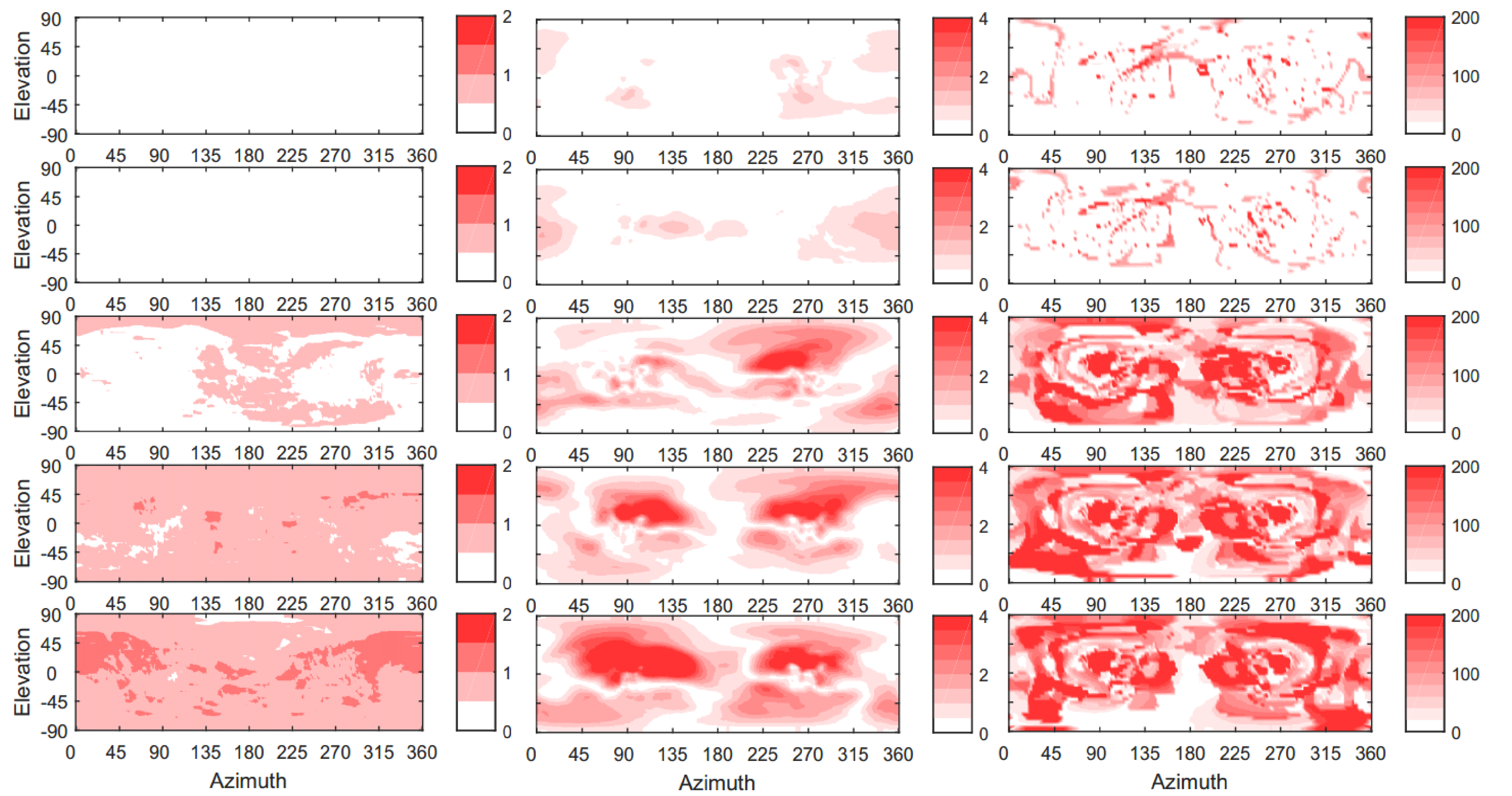

Fig 3: Absolute differences in the HRTF magnitude spectra averaged across the entire frequency range for both ears (left column) (in dB), absolute differences in **ILD** (middle) (in dB) and absolute differences in **ITD** (right) (in μs) with respect to ***GOM-Ref*** : ***SPY*** (1st row), ***CAN*** (2nd row), ***KIN*** (3rd row), ***123D*** (4th row) and ***PPT*** (5th row).

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

Reference

[1]. A. Reichinger, P. Majdak, R. Sablatnig and S. Maierhofer, "Evaluation of Methods for Optical 3-D Scanning of Human Pinna's," International Conference on 3D vision, France, pp. 390-397, 2013.
[2]. F. Brinkmann, A. Lindau, S. Weinzierl, S. van de Par, M. Muller-Trapet, R. Opdam, and M. Vorlander, "A high resolution and full-Spherical head-related transfer function database for different head-above-torso orientations," J. Audio Eng. Society, pp. 841-848, 2017.
[3]. R. R. Paulsen, "MRF-Surface," URL: http://www2.imm.dtu.dk/image/MRFSurface
[4]. H. Ziegelwanger, W. Kreuzer and P. Majdak, "A priori mesh grading for the numerical calculation of the head related transfer functions," Applied Acoustics, vol. 114, pp. 99-110, 2016.
[5]. H. Ziegelwanger, W. Kreuzer and P. Majdak, "Mesh2HRTF: An open-source software package for the numerical calculation of head-related transfer functions," 22nd International Congress on Sound and Vibration, Florence, Italy, 2015.