Evaluation of Methods for Optical 3-D Scanning of Human Pinnas

Andreas Reichinger
VRVis Forschungs GmbH
Vienna, Austria
andi@vrvis.at

Piotr Majdak
Acoustics Research Institute
Austrian Academy of Sciences
piotr@majdak.com

Robert Sablatnig
Computer Vision Lab
Vienna University of Technology
sab@caa.tuwien.ac.at

Stefan Maierhofer
VRVis Forschungs GmbH, Vienna, Austria

Abstract

In the context of computational acoustics, a detailed evaluation of various 3-D scanning methods for the purpose of capturing the geometry of the human pinna (the visible part of the ear) is presented. Over 80 full 3-D scans of the head and ears of three subjects were performed with six different 3-D scanning methods, including photogrammetry with nine different camera and texturing options. We provide a numerical comparison of the scanning performance in terms of accuracy and completeness, and show the effect of ambient occlusion on the accuracy. The numerical and practical issues for scanning the ears for computational acoustics are discussed.

1. Introduction

Our particular interest in scanning the pinna is rooted in computational acoustics, where knowledge of the so-called head-related transfer functions (HRTFs) is required [12]. HRTFs describe the directional filtering of the incoming sound by the torso, head, and pinna, recorded in the ear canal of a listener [12]. While HRTFs in the frequency range up to 16 kHz are required for accurate sound reproduction, they depend on the specific features of the pinna geometry especially in frequencies above 4 kHz [12]. Therefore, for binaural audio reproduction systems, listener-specific HRTFs are of interest.

HRTFs are usually acquired in acoustic measurements [11] and require the listener sitting motionless on a rotating chair surrounded by loudspeakers in an anechoic chamber with microphones placed in the ear canals. The measurements last for tens of minutes, are tedious and uncomfortable. Thus, much effort is put into the development of non-contact measurement methods for capturing HRTFs like numerical calculation of HRTFs from optically scanned geometry (e.g., [9]). A mobile and non-contact scan of the ear geometry promises a convenient method to simulate listener-specific HRTFs.

While the exact requirements on the geometry for HRTF simulation are still subject to research, the available accuracy and practical implications of current state-of-the art scanning methods are not fully explored yet, in particular with respect to scanning ears. Thus, this work investigates mobile non-contact 3-D surface scanning methods that are applicable at living ears and heads with a special focus on photogrammetry [16, 17] because of its perspective at consumer-grade scanning for applications offering personalized HRTF-enabled audio equipment. Note that available comparisons of the performance of long-range 3-D laser scanners, either mutually [1, 7], or by comparing a single laser scanner with photogrammetry [10, 8] do not apply to our requirements because the accuracy of long-range scanners is usually too low for our application. Also, volumetric methods like computer tomography (CT) or magnetic resonance imaging have been excluded from our study because compared to the acoustic HRTF measurements, they do not provide any advantages in terms of scan time, size or cost, and in the case of CT, at the price of exposure to x-ray radiation.

Although this work originates from questions in computational acoustics, it can also be valuable for applications in fields like ear biometrics [6], forensics [13], or scans of the inner parts for the design of hearing aid devices [14]. The results of our study can be also seen as a general evaluation of non-contact, near-field, full 3-D surface scanning methods, especially for complicated geometries with many undercuts.

In the following we describe the evaluation of near-field full 3-D surface scanning methods for over 80 full 3-D scans. For the data analysis, we adapted various mesh comparison techniques for an automated comparison and evaluate the ambient occlusion of mesh parts as an indicator for the scan performance.

2. Scanning ears in general and related work

Scanning ears with a *single* method has been investigated, e.g., in acoustics [9] and biometrics [6], and it seems to be challenging in various ways: Pinna geometry has many undercuts and deep cavities, yielding occlusions which are well-known to pose difficulties for surface scanning methods. Shoulders and hair may block views and make scanning from all directions difficult. Ears are not rigid and may deform between scans or even during the same scan. Furthermore, body hair adds a further layer of deformable complex geometry, which seems to be acoustically irrelevant but may interfere with the visual scans. Also, human skin performs complex light interaction (reflections, subsurface scattering) which might produce scanning artefacts.

A comparison of *various* methods seems to be a gap in the literature. Probably closest to ours is the comparison of five 3-D scanning methods for scanning faces [2]. For the ground-truth reference geometries, they used ten synthetic face masks made of a white solid plastic polymer. While this method allowed them to precisely compare between the scanning methods, the polymer "scans well, and does not pose the challenges to a scanner that skin does with its translucency" [2]. In our work, we were interested in *in vivo* scan performance on human ears.

3. Methods

3.1. Subjects and procedure

Ears of three human subjects were scanned. The subjects' ears covered a range of potential difficulties in the scans. S1 had small ears and medium length hair, S2 had larger ears with narrow cavities and long hair, and S3 had large ears with more open cavities and short hair.[1]

For each subject, both ears were scanned once, using the scanning methods described below. During the scans subjects wore a custom cap with holes for the ears in order to cover sculp hair, which otherwise might cloak the ears and interfere with the scanning process. The scans lasted up to one hour per subject.

Further, a plaster mold of the left ear of S2 was scanned. Compared to the skin, the plaster is more rigid and its surface is nearly diffuse. Thus, scans performed on the plaster mold aimed at providing a useful information on the effect of rigidity and problematic surface properties of living skin.

All scans were performed over the course of 6 months.

3.2. Scanning methods

In this section, we will briefly introduce the selected scanning methods and outline practical implications during the scans. Each method is labeled with a 3 letter code throughout this work.

3.2.1 MET: Metris ModelMaker MMD200

MET[2] is a hand-held device (now offered by Nikon) that triangulates line profiles by projecting a sheet of laser light, detected by a rigidly connected camera. The laser power was set using the auto-leveling function while pointing on skin. Scanning was performed by slowly sweeping the laser line in single continuous stripe movements from different angles, making post-alignment possible and typically increasing the scan quality. The scanner is attached to a MCA3600M7 7-axis coordinate measurement arm. All scans are directly registered in a common coordinate system, as long as the subject does not move. Although the subjects had their head resting on a post, some motion was unavoidable. Therefore, we scanned in individual dense stripes that were realigned as a post-process. Best-fit alignment and mesh merging (PolyWorks V10[3]) were used to create two datasets, one with best-fit alignment (BFA) until convergence and the other (MET_2) with BFA performed only twice. Scanning lasted for a long time, mainly caused by the stripe scanning pattern, which introduced redundant samples while trying to capture stripes from different views and noticeable system setup times before the system is ready to scan the next stripe. Also, the restraining measurement arm made the scan inconvenient for the operator.

3.2.2 ZSC: ZCorp. ZScanner 700CX

The scanning technology of the hand-held device ZSC[4] is similar to MET but it projects two laser sheets at a 90° angle in order to get a denser coverage in a single sweep. The biggest difference is the use of optical tracking that operates on retro-reflective markers randomly placed on the subject. Therefore, slight variations in the subject's pose did not introduce large errors and the operator was not restrained in movements. The accompanying software allowed arbitrary scan motions, and computed a mesh representation that is apparently exported from an intermediate voxel-representation. These features made scanning convenient and resulted in a short actual scanning time, even though a pre-selection of the scanning volume had to be performed.

3.2.3 MIN: Minolta VIVID-9i

MIN[5] performs light-sectioning using a laser plane, which is swept via a galvanic mirror over the scanned volume and

[1] Note that Subjects S1–3 correspond to Subjects NH131, NH130 and NH5 respectively in the HRTF Database of the Acoustic Research Institute (available at http://www.kfs.oeaw.ac.at/hrtf/).

[2] http://www.nikonmetrology.com

[3] http://www.polyworks.com

[4] http://www.zcorp.com

[5] http://sensing.konicaminolta.asia

captured by a static camera generating a depth image. In our setup, one scan was composed of three laser sweeps with different laser powers lasting for about 4 seconds during which the subject was not allowed to move. For a full 3-D model, depth images from multiple views were combined. In our scans we combined per subject on average 12 views for the head and 19 for each ear. A view change took approximately 30 s, since the scanner is heavy and autofocus was slow. Since a direct alignment method was not supported by the scanner, the caps were augmented with wooden spheres which can be detected and used for automatic alignment (Geomagic[6]). However, detection and alignment failed in approximately 30% of depth images, and had to be corrected manually. The scans were merged into a polygon model similar to MET (Polyworks), but the spheres were manually removed in all scans to yield a robust BFA. The plaster ear was scanned on a turntable which provided reliable alignment data. To study the influence of creating dedicated ear scans, we included datasets created from head-scans only (postfixed _ho).

3.2.4 BRE: Breuckmann Smart Scan 3D HE

BRE[7] is based on white-light stripe projection, using time sequential structured light scanning with 7 gray code and 4 phase slides. A two-camera wide setup with 15 and 30 degrees between projector and cameras (equipped with M lenses) was used for the scans, and handling was similar to MIN. A view change took approximately 30 s and each scan took about 2 to 4 s (depending on noise reduction settings) during which the subject was not allowed to move. Scan alignment was performed manually after each scan by specifying corresponding points. We do not have any information concerning the post-processing effort since this was performed by a Breuckmann operator who provided a "raw" and a "processed" version for each scan (postfixed _raw and _pro).

3.2.5 ART: Artec S and Artec MHT

ART[8] are hand-held scanners based on structured light using a white light flash. In contrast to time sequential structured light techniques, a full depth image is generated with each flash with a frame rate of up to 15 Hz. Tracking is performed in real-time based on the geometry captured. When tracking is lost, it can be regained by getting back to the last tracked position, or a new scan can be started which has to be manually aligned in the software by indicating corresponding points. Similar to ZSC the subject was allowed to move moderately. The small-scale S-model was used for the ears, and the medium-scale MHT-model for the head. Scanning was performed by an experienced representative of Artec. Each sub-scan was performed in approximately 30 s long scans, totaling to about 2 to 3 minutes per ear or head. After the scanning process, the meshes were stitched and finished using Artec Studio 8.0.

3.2.6 Pxy: Photogrammetry

Multi-view reconstruction was used to generate 3-D models from photographs. Since these methods require sufficiently textured surfaces, we placed stickers with pseudo-random patterns on the caps to aid automatic alignment.

In order to study the effect of texture on the reconstruction quality, we photographed each subject under three conditions: U) with *u*ndecorated ears, M) with approx. 50 *m*arker-stickers with pseudo-random patterns applied all over each ear, and S) with black water color *s*prayed all over each ear to produce a dense random pattern.

In order to study the effect of the camera we also captured each scene with three different cameras: N) professional camera equipment (*N*ikon[9]) operated by a professional photographer, F) a pocket camera (*F*ujifilm FinePix F30) without flash, and P) a smart-*p*hone camera (Samsung Galaxy S2 i9100) without flash.

Each subject was captured with most of those nine conditions. Each condition is labeled by a three letter code P + decoration + camera.

On average, we took 58 head and 114 photos per ear with the professional equipment (N), and 36 head and 47 ear photos with the other two cameras. For processing we used Agisoft PhotoScan 0.8.5 Build 1423 (64-bit)[10]. Blurry photos were removed manually, and photo sets of the head and the two ears were aligned independently using settings "high, no pair selection". These chunks were aligned using method "point based, high" and a dense reconstruction performed on the whole dataset with geometry type "smooth" and quality "medium". With these settings we obtained watertight meshes suitable for acoustics, but the "smooth" setting possibly degraded quality by *blending* geometry reconstructed from slightly misaligned sets of photos. In order to study better performance by sacrificing fully automatic processing, we also processed the PSN datasets with geometry type "sharp" (postfixed _sha). This setting does not generate a watertight mesh but outputs all potential surfaces which are consistent in subsets of photos. The most reliable surface has been semi-manually extracted (Geomagic).

In contrast to the other methods, photogrammetry reconstructs the geometry only up to a scale factor. Therefore, we manually scaled all meshes based on the average of several manually measured distances. In order to exclude

[6] Geomagic Studio 2012, http://www.geomagic.com.
[7] http://www.breuckmann.com
[8] http://www.artec3d.com
[9] Nikon D700 camera with AF-S 105 mm f/2.8 macro lens, aperture f/32 to maximize depth of field and Rayflash ring flash for fast shutter speed.
[10] http://www.agisoft.ru

this source of error from the quantitative evaluation, we included the scale factor as a further optimization parameter during the final alignment phase.

3.3. Data analysis

3.3.1 Ground-truth reference geometry

For the analysis of the scan performance, the exact shape of the ears as ground-truth (GT) meshes is required. Approximations of the GT meshes were obtained for both ears of each subject by creating negative impressions of the corresponding ear, and scanning the impressions with a high-energy industrial computer-tomography (CT) scanner [19]. Note that CT scanning is a volumetric method neither sensitive to the complexity of the surface nor prone to errors arising from occlusion.

The impressions were made of silicone. In the impressions, skin is replaced by a homogeneous high-contrast impression material, yielding a large contrast in the CT scans. Further, during the creation of impressions, body hair was either pressed away or formed narrow canals which did not affect the CT scanning. The effective accuracy of the CT scans was "in the range of two to four voxels, i.e., 0.1 to 0.2 mm" [19], thus, approximately an order of magnitude higher than those in the visual surface scans. For more details see [19].

3.3.2 Accuracy and Completeness

Our analysis of the scan performance is based on *accuracy* of a scan and the *completeness* [16]. The accuracy represents the deviation of the scan relative to a GT mesh. The completeness represents the percentage of the GT mesh modeled by the scan.

For the accuracy, we calculated a signed distance map, where for each vertex of the scan, the normal distance to the nearest point of the GT mesh was calculated. Then, a normalized cumulative distance histogram of the absolute error was calculated and the distance at the cumulative frequency of 95% was defined as accuracy. In other words, accuracy shows the maximum distance (in mm) between the scan and the GT mesh for 95% of the vertices of the scan. Note that 95% corresponds to two standard deviations of the mean for a normal distribution, and seems to be an appropriate cumulative frequency for our investigations.

For the completeness, we calculated the signed distance map, where for each vertex of the GT mesh, the normal distance to the nearest point of the scan was calculated. Then, a cumulative distance histogram was calculated and the percentage of points within an inlier threshold of 1 mm was defined as completeness. In other words, completeness shows the percentage of points in the scan that are within 1 mm of the GT mesh.

Note that the signed distance maps are later used for the visualization of the comparison.

3.3.3 Ambient occlusion

The accuracy of triangulation-based scanners is dependent on the triangulation-angle, i.e., the angle between the lines of sight of sensor and projector (or other sensor for photogrammetry) to a scanned point on the object. The accuracy has a geometrical optimum at 90 degrees and decreases with a decreasing triangulation-angle. Most of the scanning methods have a cut-off angle (sometimes also expressed as maximum depth-to-width ratio), below which the triangulation is not possible. The cut-off angle naturally limits the scanning performance for objects with narrow passages, which permit only small triangulation angles. In such cases, scanners may fail and give no or erroneous data.

In an attempt to estimate the scan difficulty and to quantify the potential narrowness of a point on the reference geometry, we use *ambient occlusion* [15], a measure coined by the computer graphics community. In principle, ambient occlusion assigns every point a value between 0 and 1, indicating the fraction of directions over its hemisphere from which the point is directly visible. An application of the ambient occlusion is shown in Fig. 1a where the GT mesh is shown with the brightness corresponding to the ambient occlusion. In our study, ambient occlusion was calculated for each GT mesh (per-vertex ambient occlusion filter, MeshLab[11]) and is used in further analysis of the errors.

3.3.4 Mesh preparation

The (raw) meshes obtained from the individual scanning methods showed often obvious artifacts. Further, the resolution of the meshes varied between scanning methods. In computational acoustics, a watertight mesh with a moderate polygon count is required. Thus, the raw meshes were post-processed by the following procedure for all scans.

Meshes were cleaned by applying automatic corrections (Mesh Doctor, Geomagic) repeatedly until the automatic corrections did not find any issues any more. Small mesh parts, not connected with the main mesh, and usually effect of noise in the measurements, were removed by deleting everything but the largest connected component. Further, the main mesh was manually trimmed to contain only the head, removing, e.g., parts of the body and head-stabilizing posts. In the main mesh, holes were filled with a curvature-based method (Fill All, Geomagic) and the automatic corrections (Mesh Doctor) were re-applied in order to fix issues potentially introduced by the filling. Finally, meshes were downsampled to 1 mm average edge length (Remesh, Geomagic) and the automatic corrections were again applied in order

[11] http://meshlab.sourceforge.net/

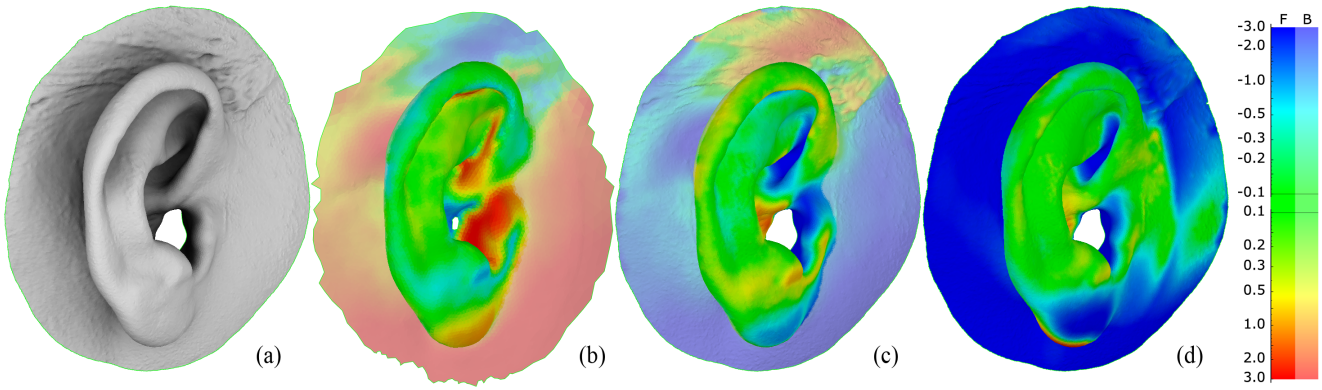

Figure 1. Right ear of S2: (a) GT mesh, brightness = ambient occlusion; (b) mesh scanned with MET, color = signed error (scan→GT); (c) GT mesh analyzed for MET, color = signed error (GT→scan); (d) average signed error (GT→scan) over all scanning methods visualized on the GT mesh.

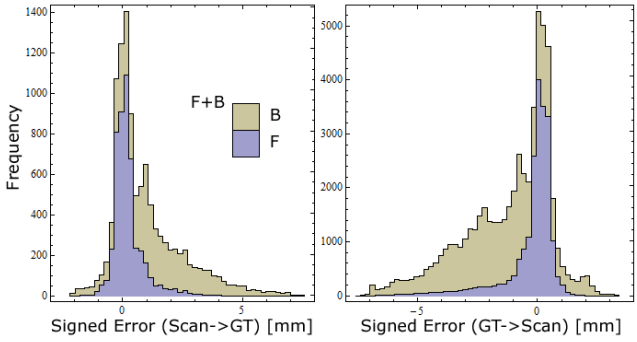

Figure 2. Right ear of S2: Stacked histograms of signed errors.

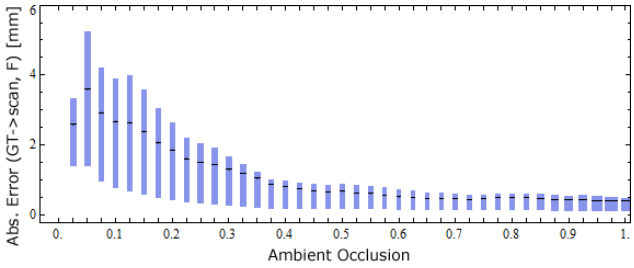

Figure 3. Absolute error GT→scan (25% and 75% quartiles and average) as a function of occlusion averaged over all scans. Note the increasing error with decreasing visibility.

to fix issues potentially created by the down-sampling, like highly creased edges or spikes.

The post-processing procedure for the GT meshes was a different one. In these meshes, air bubbles have been removed and filled (see [19] for more details), and the results were then down-sampled to an average edge length of 0.35 mm (Fig. 1a).

For the quantitative analysis, all meshes were divided into two parts by assigning each vertex in a mesh either to front (F) or to back (B). This was done to particularly investigate the performance of scanning methods for the front part of the pinna and was motivated by two aspects. First, the front part of the pinna seems to be acoustically the most relevant part [18] and thus we were interested most in the scan performance for that part. Second, the back part of the pinna, with its narrow passage between ear and head, might yield problems in triangulation-based scanners, resulting in a substantially worse scanning performance in that region of the pinna. Note that also the cap might be a source of additional problems especially behind the ear. In an analysis considering both parts of the pinna, the performance of the scanning methods for the front part of the pinna might have been overshadowed by the large problems in the back part of the pinna.

Note that in the quantitative analysis, we focus on the results for the front part of the pinna.

3.3.5 Alignment

In order to compare two meshes, they must be in the same coordinate system. This was achieved by aligning the scanned meshes to the coordinate system of the GT mesh.

The aligning was done by human-guided pre-alignment with a subsequent iterative closest point (ICP, e.g., [3]) optimization. For the pre-alignment, manual registration was performed by manually selecting up to nine corresponding points (Manual Registration, Geomagic) and then automatically fit based on geometry (Best Fit Alignment, Geomagic) with the options “fine adjustments only” and “automatic deviator elimination”. For the fine-alignment, a customized version of the ICP algorithm was used based on the optimization by Chen and Medioni [3]. Each vertex of the scan was matched to the nearest vertex on the GT mesh within a radius of 3 mm. Matches of backsides of thin structures were avoided by removing all boundary vertices and ver-

Meth.	Scan Time	MM	Comfort	Post Time	Manual Input	Potential
MET	60 min	no	very low	60 min	when alignment fails	low
MIN	45 min	4 s	medium	60-120min	yes, alignment	low
BRE*	20-30min	2-4 s	medium	N/A	yes, aid alignment	low
ZSC	40 min	yes	med-high	10 min	only define volume	medium
ART*	8 min	yes	high	10-20 min	yes, aid alignment	high
P*xy*	10 min	yes	high	40-120min	no	high

Table 1. Potential of methods based on practical factors. Methods marked with * were scanned by professionals. MM=May Move.

tices with differences in the normals larger than $45°$. The matches were weighted by two factors. First, the difference in the local curvatures, calculated for both meshes with the algebraic point-set surfaces method [5] and weighted by a Gaussian. This weight reduces the impact of dissimilar local structures potentially introduced by scanning errors. Second, the square root of the ambient occlusion term [15] of the GT mesh. This weight reduces the impact of occluded regions, which were assumed to be more error prone and thus to affect the matching quality. Note that only the front part of the GT mesh and only vertices of the scan mesh that were not created by hole filling were used for the alignment.

All aligned meshes were trimmed to the approximate size as the GT mesh.

4. Results and discussion

By scanning the same subjects with different scanning methods under comparable conditions, valuable practical experience has been gained, revealing the relevant factors for our future application. Table 1 summarizes the most important factors.

The average scan time was in many cases larger than the time required in acoustic HRTF measurements. Note, however, that all scans were made ad-hoc with only few preliminary tests and would probably last shorter for experienced operators. In Table 1, the comfort for the scanned subject is based on a combination of the required scan time and whether the subject were allowed to move or not. For example, MET showed the largest scanning time of all tested methods, and the subjects were not allowed to move at all during the full scan, rendering this method as impracticable for our future application.

The non-hand-held scanners (BRE and MIN) did not allow the subjects to move during each single sub-scan, but the subjects were allowed to move in between. However, the long scan time and the required manual input for the post processing show a low potential for these methods.

Although the *total* scan time of ZSC is large, the time was dominated by computation time between the scans. The *actual* scan time where the subject had to remain seated was approximately 6 minutes each for head and ears, rendering the process more convenient and yielding thus a higher overall potential.

ART and Pxy have the highest potential since the scan time was among the lowest and the subjects could move to a certain extent. Nevertheless, we ranked ART slightly lower, since it required some manual post processing.

Note that post-processing time has only a low relevance for our future application, as long as it can be performed mostly automatically.

4.1. Vertex-specific evaluation

The error analysis is visualized by showing the signed distance map on the meshes encoded by the color (inspired by [4]). The visualization for an exemplary ear (right ear of S2) is shown in Fig. 1b and c. Negative distances (going into the object) smaller than -3 mm, exact matches (within ± 0.1 mm), and positive distances (above the surface) larger than 3 mm are encoded by the colors blue, green, and red, respectively. Values between those ranges are encoded by a logarithmic gradient in the hue. The back parts (B) of the pinna are shown in reduced saturation. Note that two meshes are shown. Fig. 1b shows the signed distances calculated for each vertex of the scan and represents the accuracy. Fig. 1c shows the signed distances calculated for each vertex of the GT mesh on which the completeness measure is based. Fig. 2 shows the same data in histogram form.

The error analysis revealed that the accuracy in the region of the ear lobe was large for most scanning methods. In order to evaluate this issue more specifically, a signed distance map was calculated as the average over all scanning methods and is shown in Fig. 1d. Indeed, the earlobe seems to show a large deviation from the GT mesh even when averaged over all scanning methods. Since the scans were performed in different sessions and on different days, it is rather unlikely that every scanning method resulted in such a way. We rather assume that the impression made for the GT mesh had a bias caused by the pose of the subject required to create the impression (lying on the side). Also, the errors seem to be large at the parts with narrow cavities and the backside of the ear, indicating that all scanning methods had difficulties scanning these parts.

Figure 3 shows the accuracy error as a function of occlusion averaged over all scans. The accuracy error clearly increases for decreasing visibility, showing that indeed all scanning methods had problems scanning more occluded parts of the geometry like narrow cavities. It seems like ambient occlusion is an appropriate metric to estimate the scanning performance when a GT mesh is not available.

4.2. Scan-specific evaluation

Table 2 shows the accuracy and completeness for all ears and scans. Further, the performance is provided for each ear averaged over the scanning methods (bottom row) and for each scanning method averaged over the subjects (second-right column). Finally, the performance is provided for the

F	S1 L		S1 R		S2 L		S2 R		S3 L		S3 R		avg S1-3		Plaster	
95% 1mm	acc	c[%]	acc	c[%]	acc	c[%]	acc	c[%]	acc	c[%]	acc	c[%]	acc	c[%]	acc	c[%]
ART	2.16	60.8	1.74	80.0	2.18	74.0	2.61	81.2	2.22	77.2	1.56	80.3	2.08	75.6	0.21	98.9
BRE_raw	2.17	58.0	2.05	68.5	1.43	85.9	1.64	82.8	1.61	81.4	1.41	82.8	1.72	76.6	0.08	99.5
BRE_pro	1.99	64.7	2.19	61.3	1.46	86.5	1.64	82.7	2.27	78.9	1.61	79.2	1.86	75.6	0.07	99.5
MET	1.45	80.8	1.58	76.8	1.95	67.4	1.40	85.2	1.72	85.5	1.27	86.8	1.56	80.4	0.50	95.0
MET_2	1.34	83.3	1.68	75.4	1.93	70.8	1.42	84.9	1.46	90.5	1.25	87.6	1.52	82.1		
MIN	2.18	76.3	2.05	75.4	2.09	80.1	1.86	81.1	1.86	79.4	1.67	74.1	1.95	77.7	0.22	99.8
MIN_ho	1.94	60.5	2.40	62.4	2.03	70.2	1.79	74.1	2.50	66.5	1.84	70.9	2.08	67.4		
ZSC	1.76	83.1	1.28	89.2	1.23	90.1	1.45	88.2	1.46	82.8	0.99	86.3	1.36	86.6	0.34	99.1
PSN_sha			1.45	84.2	1.76	84.3	2.29	63.3	1.82	85.9	1.17	86.9	1.70	80.9	0.36	97.3
PSN	1.46	75.0	1.75	76.6	1.98	78.7	2.41	65.5	1.81	83.7	1.31	84.8	1.78	77.4		
PSF	1.48	75.0	1.90	69.6	1.92	78.9	2.62	54.1	2.02	79.0	1.64	78.7	1.93	72.5		
PSP	1.82	68.4	2.19	61.6	1.79	71.2	2.28	66.3	2.13	77.0	1.29	81.4	1.92	71.0		
PMN					1.78	69.9	1.44	80.8	1.93	73.9	2.08	76.7	1.80	75.3		
PMF					1.86	64.9	1.96	70.2	1.93	73.0	1.58	75.3	1.83	70.8		
PMP					2.74	59.0	2.17	66.5					2.46	62.7		
PUN					1.96	63.5	1.97	70.7	2.07	75.9	2.28	69.0	2.07	69.8	8.78	76.9
PUF					2.80	54.0	2.27	60.9	3.23	55.3	3.13	52.5	2.86	55.7	2.38	79.8
PUP					2.94	52.0	2.64	49.8	3.19	54.7	3.17	52.8	2.98	52.3	1.47	88.7
average	1.80	71.4	1.86	73.4	1.99	72.3	1.99	72.7	2.07	76.5	1.72	76.8	1.90	73.9	1.44	93.5

Table 2. Accuracy (acc in mm) and completeness (c in %) for the front part of the ears. Column *avg S1–3*: accuracy and completeness averaged over all scans of the living ears. Plaster: accuracy and completeness for the plaster ears. Color indicates the rank in each column (red: worst, green: best, grey: dataset not available).

plaster ear (rightmost column).

Note that the individual histograms used for calculation of the accuracy and completeness are provided on our project website[12], where in browser-based graphics and plots further aspects of our data analysis can be explored. The charts on the website further allow to modify the parameters used to calculate the accuracy and completeness, and provide an interactive ranking of the scanning methods evaluated in this study.

In the conditions depicted in Table 2, ZSC shows the smallest average accuracy error and the largest average completeness. It is followed by MET which is surprising, since one would assume registration problems because of the fixed coordinate system and inevitable head movements in that method. Obviously, these problems did not substantially affect the results. BRE yielded also a high performance on average. These averages are, however, clearly biased by the poor results for S1, indicating that while BRE might be generally capable of yielding better performance, some problem occurred in our scans on a subject-individual basis. Interestingly, the less processed raw version is on average more accurate and complete than the processed version. ART yielded a lower performance on average, which might be a result of tracking problems during the scans. This method uses only the scanned geometry for position tracking and an additional attachment of unique geometric features might improve the results. Note that the current version of the software is capable of additional texture tracking which could also improve the results.

The best average accuracy was 1.36 mm (ZSC) and the best average completeness was 86.6% (also ZSC). This is in contrast to the plaster model, for which the accuracy was 0.34 mm and completeness was 99.1% for the corresponding scanning method. Also, the performance averaged over all scanning methods showed a substantially better accuracy and larger completeness for the plaster ear (1.44 mm; 93.5%) than for the living ears (1.90 mm; 73.9%). Such a striking difference might be caused by the non-rigidity of the living ear, by the optical differences in material between skin and plaster, as well as by the possibility to conveniently scan the plaster ear from all directions. This difference also clearly demonstrates the challenge when *in vivo* scanning the ears.

As an interesting side-effect, we found that BRE, MIN, and ART performed in particular well on the plaster ear, which might indicate that these scanners—especially when combined with turntables—have been optimized for scanning rigid objects.

In several of the conditions, photogrammetry performed worse than the other scanners, especially on the undecorated ear (PUy conditions). This might indicate, that the ear has too few features on its own for reliable stereo matching. A further source of error might have been alignment difficulties between groups of photos and the subsequent blending of these misalignments (the "smooth geometry" option). Indeed, additional tests with the "sharp" option—which does not perform the blending step but requires manual interventions—yielded much better results, suggesting that the automatic alignment was not the optimal choice. With manual intervention and the "sharp" option, the accuracy and completeness was in the range of those for other 3-D scanning methods. Note, however, that the photogrammetry results might be biased, since we allowed to optimize the scale factor during the alignment phase. Without this optimization the results were significantly worse, indicating that a good metric calibration is mandatory.

The professional camera equipment (PxN conditions) yielded better results than the conditions using the pocket and cell-phone cameras. This is probably because of better lighting conditions since this was the only method using a flash. With an appropriate flash, the smaller cameras might have yielded better results because of their small aperture and therefore potentially larger depth of field. In addition, the better performance may also be due to the larger amount of photos taken with the professional camera.

The reliability of a scanning method can be investigated by comparing the performance for the left and right ears of a subject. While it would be tempting to conclude that most of our scanning methods yielded reliable results, the comparison is based on only three subjects, with an unknown number of outliers (scans may occasionally fail). What is clear is that the scanning results depend on many factors, and the scanning skills of the operator have a large impact on the results, especially for hand-held scanners.

[12] http://locaphoto.vrvis.at/eval/

5. Conclusions

This study investigated the quality of pinna meshes acquired with various scanning methods. The quality was expressed in terms of accuracy and completeness with respect to ground-truth meshes of the corresponding ears. Our results show that scanning living ears is indeed difficult and the results are generally poorer than scanning plaster impressions of the same ears. The vertex-specific analysis showed that narrow cavities and the backside of the ear have been captured with large errors, probably caused by the simple curvature-based interpolation of holes in raw meshes. Advanced interpolation schemes are necessary to reduce the errors in these regions.

Also, our evaluation shows that the best scanners in terms of accuracy are unacceptable in terms of practicability. Methods PSN_sha or ZSC seem to be promising and yielding already now acceptable results. Note that this does not necessarily mean that the other scanners are not suitable. All scans were performed in an ad-hoc setup leaving room for improvements in the future. In addition, scanners evolve and might perform better now.

The main problems are a practical combination of reliable alignment and short scanning time. While the necessary accuracy for acoustic simulations has yet to be clarified, the photogrammetry using professional equipment, and reliable alignment seems to provide potential for an applicable method for HRTF simulations in the future.

Acknowledgments

Special thanks to Michael Mihocic, Michael Schwärzler, Christa Rothner, Florian Rist (Institute of Art and Design, TU Vienna), Daniel Habe (ÖGI, Leoben), Jaroslav Pěček (ABBAS, Brno), Jörg Fauland (Breuckmann), Stephan Mantler, Matthias Kucera (Ludwig Boltzmann Institute, Vienna) and Fabian Hollaus (Computer Vision Lab, TU Vienna). This work was supported by the Austrian Science Fund (FWF): P24352-N23.

References

[1] W. Boehler, M. Bordas Vicent, and A. Marbs. Investigating laser scanner accuracy. *Int Arch Photogram Rem Sens Spatial Inform Sci*, 34(Part 5):696–701, 2003. 1

[2] C. Boehnen and P. Flynn. Accuracy of 3d scanning technologies in a face scanning scenario. In *3D Digital Imaging and Modeling, 2005. 3DIM 2005. Fifth Int. Conf. on*, pages 310–317. IEEE, 2005. 2

[3] Y. Chen and G. Medioni. Object modelling by registration of multiple range images. *Image and Vision Computing*, 10(3):145 – 155, 1992. 5

[4] P. Cignoni, C. Rocchini, and R. Scopigno. Metro: Measuring error on simplified surfaces. *Comput. Graph. Forum*, 17(2):167–174, 1998. 6

[5] G. Guennebaud and M. Gross. Algebraic point set surfaces. *ACM Trans. Graph.*, 26(3):Article No. 23, July 2007. 6

[6] S. M. S. Islam, M. Bennamoun, R. A. Owens, and R. Davies. A review of recent advances in 3d ear- and expression-invariant face biometrics. *ACM Comput. Surv.*, 44(3):14:1–14:34, June 2012. 1, 2

[7] M. Johansson. Explorations into the behaviour of three different high-resolution ground-based laser scanners in the built environment. In *Proc. Int. Society for Photogrammetry and Remote Sensing*, pages 33–38, 2002. 1

[8] R. Kadobayashi, N. Kochi, H. Otani, and R. Furukawa. Comparison and evaluation of laser scanning and photogrammetry and their combined use for digital recording of cultural heritage. *Int Arch Photogram Rem Sens Spatial Inform Sci*, 35(5):401–406, 2004. 1

[9] W. Kreuzer, P. Majdak, and Z. Chen. Fast multipole boundary element method to calculate head-related transfer functions for a wide frequency range. *J Acoust Soc Am*, 126(3):128090, Sep 2009. 1, 2

[10] D. Lichti, S. Gordon, M. Stewart, J. Franke, and M. Tsakiri. Comparison of digital photogrammetry and laser scanning. In *Proc. Int. Society for Photogrammetry and Remote Sensing*, pages 39–44, 2002. 1

[11] P. Majdak, P. Balazs, and B. Laback. Multiple exponential sweep method for fast measurement of head-related transfer functions. *J. Audio Eng. Soc*, 55:623–637, 2007. 1

[12] H. Møller. Fundamentals of binaural technology. *Applied Acoustics*, 36(34):171218, 1992. 1

[13] M. S. Nixon, I. Bouchrika, B. Arbab-Zavar, and J. N. Carter. On use of biometrics in forensics: gait and ear. In *European Signal Processing Conference*, Aalborg, Denmark, Aug 2010. 1

[14] R. R. Paulsen, C. Nielsen, S. Laugesen, and R. Larsen. Using a shape model in the design of hearing aids. In *Proc. SPIE 5370, Medical Imaging 2004: Image Processing*, pages 1304–1311, May 2004. 1

[15] M. Pharr and S. Green. *GPU Gems*, chapter Ambient occlusion, pages 279–292. Addison-Wesley, 2004. 4, 6

[16] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Proc. IEEE Conf. on Comp. Vis. and Pat. Recog.*, CVPR '06, pages 519–528, Washington, DC, USA, 2006. IEEE. 1, 4

[17] C. Strecha, W. von Hansen, L. Van Gool, P. Fua, and U. Thoennessen. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *Proc. IEEE Conf. on Comp. Vis. and Pat. Recog.*, CVPR '08, pages 1–8, Los Alamitos, CA, USA, June 2008. IEEE. 1

[18] H. Takemoto, P. Mokhtari, H. Kato, R. Nishimura, and K. Iida. Mechanism for generating peaks and notches of head-related transfer functions in the median plane. *J Acoust Soc Am*, 132(6):3832–3841, 2012. 5

[19] H. Ziegelwanger, A. Reichinger, and P. Majdak. Calculation of listener-specific head-related transfer functions: Effect of mesh quality. In *Proc. of Meetings on Acoustics (POMA)*, volume 19, page 050017, Montréal, June 2013. Acoustical Society of America. 4, 5