



Analysis of laser scanning and photogrammetric scanning accuracy on the numerical determination of Head-Related Transfer Functions of a dummy head

Fabio Di Giusto^{1,2,*} , Sjoerd van Ophem^{1,2}, Wim Desmet^{1,2}, and Elke Deckers^{3,2}

¹Department of Mechanical Engineering, KU Leuven, Celestijnenlaan 300 B, B-3001 Heverlee, Belgium

²Flanders Make@KU Leuven, B-3001 Heverlee, Belgium

³Department of Mechanical Engineering, KU Leuven, Campus Diepenbeek, Wetenschapspark 27, 3590 Diepenbeek, Belgium

Received 17 November 2022, Accepted 15 September 2023

Abstract – Individual Head-Related Transfer Functions (HRTFs) are necessary for the accurate rendering of virtual scenes. However, their acquisition is challenging given the complex pinna shape. Numerical methods can be leveraged to compute HRTFs on meshes originating from precise scans of a subject. Although photogrammetry can be used for the scanning, its inaccuracy might affect the spatial cues of simulated HRTFs. This paper aims to assess the significance of the photogrammetric error affecting a Neumann KU100 dummy head scan. The geometrical differences between the photogrammetric scan and a laser scan are mainly located at the pinna cavities. The computed photogrammetric HRTFs, compared to measured and simulated data using objective and perceptually inspired metrics, show deviation in high frequency spectral features, stemming from the photogrammetric scanning error. This spectral deviation hinders the modelled elevation perception with photogrammetric HRTFs to levels comparable to renderings with nonindividual data. Extracting the photogrammetric geometry at individual ear cavities and merging it to the laser mesh, an assessment of the influence of the inaccuracy at different pinna structures is conducted. Correlation analysis between acoustic and geometrical metrics computed on the results is used to identify the most relevant geometrical metrics in relation to the HRTFs.

Keywords: Head-Related Transfer Functions, Photogrammetry, Boundary Element Method, Binaural rendering, Spatial sound localisation

1 Introduction

The auditory information carried by the sound signals arriving at the eardrums of a subject contains the attributes enabling spatial hearing. These are generated by the torso, head and outer ears, operating as a linear filter whose transfer function, the Head-Related Transfer Function (HRTF), mainly depends on sound source position and individual shape of the ears [1, 2]. Several phenomena affect the sound impinging on the pinna, coding the sound-field characteristics into spatial cues perceivable with one or both ears, defined monaural and binaural cues, respectively. The first are significant for elevation perception and front-back discrimination [3], while the latter are important for localisation in the lateral dimension and relate to Interaural Time Differences (ITDs) and Interaural Level Differences (ILDs) of the sound signals at the two ears [4]. Additional cues are also employed in spatial hearing, i.e. dynamic cues related to movement and nonacoustic factors such as visual cues [5]. By virtually replicating the auditory cues in signals

presented to a subject, a reliable spatial sound experience can be rendered. This is commonly done through HRTF filtering and headphone presentation [4], for which generic and individual HRTFs can be used; however, the first relate to increased front-back confusion and localisation error in static renderings without visual information [6]. Although in Virtual Reality (VR) dynamic and visual cues are available, individual filters are necessary to obtain the most realistic spatial rendering [7].

Individual HRTFs are usually acquired through acoustic measurements in anechoic conditions [8]; however, this tends to be a tedious procedure, affected by errors generating variations shown to cause differences in localisation performance between repeated measurements [9]. Alternatively, numerical methods, e.g. the Boundary Element Method (BEM), can be leveraged to calculate HRTFs on individual ears, head and torso geometries [10]. In this approach a precise 3D scan is required, which is challenging to acquire due to the complex pinna shape and the high resolution needed to obtain perceptually valid HRTFs [11]. Several techniques have been tested to scan the outer ear, e.g. Computed Tomography (CT), Magnetic Resonance

*Corresponding author: fabio.digiusto@kuleuven.be

Imaging (MRI), laser scanning and structured-light scanning [12–15]. The first two, being volumetric scanning techniques, are not sensitive to the complexity of the pinna surface generating self-occlusion; however, they entail exposure to radiation or magnetic fields and expensive equipment. The last two are surface scanning techniques based on line-of-sight propagation; hence, inherently hindered by the occluded pinna geometry. Nonetheless, previous studies have shown that laser scanning is capable of creating suitable geometries for HRTF computation [16, 17]. The most precise scanners typically require specialised equipment and long scanning time, making their use impractical outside of a scientific framework [14]. An alternative scanning technique is photogrammetry, based on obtaining information on the surface of an asset using overlapping pictures of it [18]. Recent commercial applications of photogrammetry have been released, aiming at rapidly synthesising individual HRTFs from images of a subject [8]. The advantage of this method lies in the affordability of the scanning equipment and the short scanning time, though the drawback is a generally lower precision. The pinna shape, plotted in Figure 1, presenting several cavities, i.e. the cavum conchae (concha), cymba conchae (cymba), fossa triangularis (fossa) and scapha, creates a lack of visibility which particularly hinders the photogrammetric reconstruction algorithm. Given the low accuracy at these ear locations, relating to noise and incompleteness in the scans, manual corrections are required, making the application of photogrammetry challenging. However, the widespread availability of camera sensors suited for this scanning technique, e.g. smartphone cameras, makes this method one of the most appealing for a large-scale application [14, 16].

The photogrammetric scanning of the outer ears of 3 subjects has shown an average accuracy of 1.7 mm and completeness of 81%, defined as the maximum distance between CT and photogrammetric scans for 95% of the acquired vertices, and the percentage of points in the scan within 1 mm from the reference CT scan, respectively. Therefore, this scanning technique is indicated to be potentially applicable for numerical HRTF computation [14]. Comparisons of scans of a dummy head, obtained through various techniques including photogrammetry, to a reference structured-light scan have been carried out, showing maximum differences of 5 mm at the ear between the photogrammetric and reference geometry [15]. Objective and perceptually inspired metrics evaluated on HRTFs computed on the photogrammetric and reference meshes show differences in interaural features and modelled localisation performance, which are thought to relate to perceptual discrepancies between the HRTFs [15]. However, the reference HRTFs are not validated against measured data on the same dummy head. Another study has reported a typical deviation of 1 mm between photogrammetric and reference scans, and higher values in locations difficult to scan but acoustically non-significant, e.g. the back of the ear. The spectral features of horizontal plane HRTFs simulated on the scans show a good match below 10 kHz; however, no comparison is given at higher frequencies [19]. Furthermore,

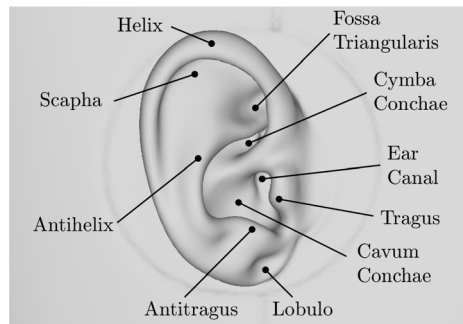


Figure 1. Anatomical pinna structures.

no perceptual assessment is conducted on the photogrammetric HRTFs.

The various anatomical pinna structures are thought to have a different influence on the HRTFs. Several studies have tried to define the effect of ear morphology using databases of matching HRTFs and anthropometric measurements, e.g. the CIPIC database [20]. Correlation analyses carried out on this data between selected ear measures and horizontal plane HRTFs have found 4 pinna dimensions having strong correlation, relating to concha and overall pinna dimensions [21]. Analyses of numerically obtained pressure distributions on 4 MRI scanned human heads over the full audible range have revealed that the outer ear cavities are responsible for creating the HRTFs spectral peaks and notches [22]. The latter tend to shift in frequency depending on sound source elevation; therefore, they are indicated as perceptual cues in the median plane. They are generated by resonances in various pinna cavities, e.g. the cymba and fossa, changing in location and frequency depending on source position and interfering with the impinging sound, creating a node at the entrance of the ear canal. However, no clarification on the relationship between pinna geometry and spectral features is provided, given the complexity of the pinna shape; thus, the use of a simplified pinna model is suggested. Analyses of numerical HRTFs computed on a simplified ear, head and torso model up to 8 kHz have been conducted [23], in which 6 head and 6 pinna parameters are separately modified within ranges derived from anthropometric measures of adults and children, and spectral deviation, ITD and ILD variations are evaluated. The results show that concha dimensions and pinna rotations greatly affect the HRTFs. A similar study, using a schematic head and torso geometry and a realistic parametric pinna model, has analysed the effect of modifying 8 selected CIPIC parameters within 2 standard deviations from the mean evaluated on measured data [24]. Various spectral distance metrics are used up to 16 kHz, and the outcome has emphasised the importance of the concha and fossa dimensions, while showing a lower effect of other parameters such as the pinna width, which was conversely indicated as influential in [21]. It is suggested that this disagreement could stem from the different ear shapes under consideration, which can possibly reduce the effect of some parameters, making it challenging to obtain coherent results. However, no in-depth analysis is provided

to assess this hypothesis. Another study has used a parametric pinna model modified through control points related to anthropometric ear measures, varying them separately in ranges extracted from a database, while trying to maintain all others at their spatial average [25]. Objective and perceptual metrics are employed to assess the results up to 12 kHz, showing that parameters changing the concha and fossa dimensions have the strongest influence, along with those controlling the depth of some pinna structures. In general, apart for concha dimensions, disagreement is observed between the reported significant parameters, indicating a non-trivial link between HRTFs and anthropometry.

This paper aims at assessing the effect of the scanning error inherent to photogrammetry on HRTFs acquired on a Neumann KU100 dummy head. Although standard dummy head HRTFs tend to lead to a lower localisation accuracy in comparison to selected human subject HRTFs [26], the KU100 is used given the availability of several measurements [27]. Furthermore, the possibility of directly acquiring its shape with accurate scanners makes it easier to obtain a reference geometry for comparison with the photogrammetric results. MRI scans of this dummy head have been carried out [28]; however, some problems are reported in the scan, mainly relating to the resolution of the scanner being comparable to the pinna thickness, and requiring manual post-processing and mirroring of the left part of the geometry to the right, on the assumption of symmetry of the dummy head. HRTFs simulated on this geometry show similar patterns to measured results, but significant shifts in spectral features above 5 kHz. Laser scanning is used in the current study to approximate the reference geometry of the dummy head. The investigation focuses on evaluating the photogrammetric scan by comparing it to the laser scan using several geometrical metrics, and assessing the validity of the numerical HRTFs computed on it. Given that by modifying the depth of the pinna cavities by a small amount significant shifts in the centre frequency of the HRTFs peaks and notches are observed [29], the level of deviation of photogrammetric scans is expected to alter the high frequency spectral characteristics of simulated HRTFs. This is assumed to have a detrimental impact on their perceptual attributes, especially on the median plane where the spectral features are considered to be important localisation cues. Based on results of previous studies, it is expected that scanning errors in the proximity of the concha and fossa have a strong effect on the HRTFs. However, these studies have focused on deformations of pinna morphology within ranges extracted from inter-subject variability analyses. The photogrammetric error is assumed to generate different discrepancies, potentially having a different impact on the HRTFs. These errors might stem from the scanning noise inherent to photogrammetry, which tends to mostly appear “outside” the scanned surface [30], e.g. creating shallower cavities, and holes due to occlusion, which could be tackled by interpolation from the surrounding points [18], potentially generating high deviation. Limited in-depth studies have been carried out regarding the influence of the photogrammetric error on HRTFs numerically computed on the acquired geometries,

and the perceptual similarity of these HRTFs to reference data. To assess and quantify this effect, objective and perceptually inspired metrics are employed for comparing numerical HRTFs to reference measured and simulated data. Given the variability observed in repeated measurements, attention should be given in selecting the reference HRTFs for comparison. For this purpose, a method combining previously defined metrics and comparison procedures is proposed to evaluate the perceptual similarity between measured and simulated data, and identify reference HRTFs which can be used to assess the results obtained on the photogrammetric scan. Furthermore, the influence of the scanning error at different pinna structures is evaluated by merging the photogrammetric geometry at different anatomical regions to the laser mesh and assessing the numerical HRTFs. The goal of this analysis is to identify at which locations the photogrammetric scanning error produces the most critical effect on the HRTFs, and further discuss the applicability of photogrammetry for scanning related to individual HRTFs computation. Correlation analysis is carried out between the geometrical and acoustic metrics evaluated on the scanned results and the related HRTFs, to identify the most relevant geometrical metrics evaluated between a target and a reference scan in terms of similarity of the HRTFs computed on them. Although the geometrical metrics are assumed to show high correlation between them, given that they are generally calculated on the distance between target and reference geometry, it is yet unclear which of these better relate to the objective and perceptual HRTF comparison metrics.

This paper is structured as follows. In [Section 2](#) the scanning procedure is described, the geometrical evaluation metrics are introduced, and the achieved results are discussed. [Section 3](#) addresses the numerical HRTFs acquisition, presents the metrics used for their analysis and the outcome of their assessment. [Section 4](#) studies the influence of the inaccuracy at different scanned ear parts, explaining its evaluation and reporting the obtained results. A general discussion is carried out in [Section 5](#), followed by the conclusion in [Section 6](#).

2 Geometry acquisition and assessment

2.1 Scanning and processing methods

The KU100 dummy head is scanned using a Nikon LC60Dx line laser scanner mounted on a Coord3 MC16 coordinate measuring machine. The head is scanned entirely, while a separate scan is carried out on the dismounted ear elements to achieve a higher precision. A similar procedure using laser scanning to obtain accurate pinna geometries has been employed in [17]. Given the sub mm accuracy of this scanner¹, the obtained geometry is used as an approximation of the reference dummy head shape. The ear canals are closed at their entrances by the measurement microphones; therefore, the scan presents blocked

¹ Specifications available from https://www.nikon.com/products/industrial-metrology/support/download/brochures/pdf/lc60dx_en.pdf [Accessed on March 2, 2021].

ear canals. Laser scanning is chosen for the scan given the unavailability of volumetric scanning techniques, and the fact that it has already been used to acquire accurate meshes for numerical HRTF computation [17]. Nonetheless, HRTFs computed on this scan are compared and validated against experimental data.

The photogrammetric scan is carried out on images of the dummy head extracted from videos taken with the 12.2 MP camera sensor of a smartphone (Google Pixel 4a 5g) using the maximum resolution (Full HD 1080 P). To achieve a high reconstruction quality, non-reflective white spray is applied on the dummy head to create a complex surface pattern, given that the uniform dark coloured texture of the dummy head hinders the photogrammetric reconstruction algorithm. Although this tends to increase the scanning accuracy compared to an untreated scenario, application of a complex pattern prior to scanning has been previously used on dummy heads and human subjects, e.g. employing matte makeup powder, black water paint or markers [14, 19]. Nevertheless, results of ear scans without optical treatment are presented for comparison. Videos of the head and ear elements are captured separately in a room mainly lit by artificial light; camera parameters are maintained constant during the acquisition. The video shooting light is turned on for the full duration of the scan. Although its usage is not recommended, given the nonconsistent lighting between images captured from different angles [31], this is activated since it relates to a more complete reconstruction in some parts of the ear which are challenging to scan, e.g. the cymba. The video of the entire head is acquired by manually holding the smartphone camera at a distance of roughly 0.5 m, pointing at the dummy head and ensuring that the full object fits in the camera frame, while moving the camera parallel to the horizontal plane at different elevations from bottom to top; hence, capturing the object from several angles. Videos of each ear are similarly taken, but at a closer distance of roughly 0.2 m and on a hemisphere frontal to the ear. Each of the acquired videos has a duration of 120 s. One frame per second is extracted from each video to obtain a uniform sampling of the object among all the recorded directions; images in which the dummy head is out of focus or not centred in the camera frame are discarded. Thus, between 90 and 100 images are extracted from each video and used for the reconstruction through a structure-from-motion photogrammetry algorithm [32, 33]. The choice of the number of images is based on a study of outer ear photogrammetric scans deriving from a set of 30, 60 and 90 pictures taken with a smartphone camera [18]. The results have been analysed in terms of completeness and accuracy, evaluated as the percentage of scanned points within 2 mm from a reference, i.e. a structured-light scan, and the Root Mean Square (RMS) value of the distance between points in the reference and photogrammetric scan, respectively. Reconstructions with the highest number of images relate to slightly better results than with fewer pictures; however, the differences between them are reported not to be statistically significant.

The procedure from laser and photogrammetric scanned point clouds to meshes is summarised in Figure 2. Prior to

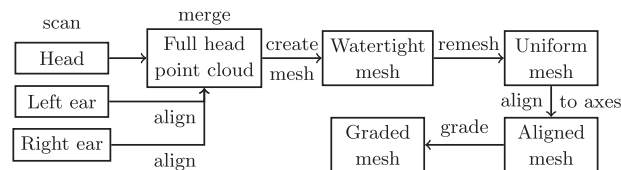


Figure 2. Block-diagram of the steps used to obtain a graded mesh from the laser and photogrammetric scan results. MeshLab and OpenFlipper are used to carry out the processing steps [34, 36].

merging the head and ear point clouds, manual interventions are required to remove unnecessary scanned points, e.g. supports needed for the scan and background objects. The alignment between head and ear scans is done in MeshLab with the “Point Based Gluing” tool, in which at least 4 homologous points are manually selected in the scans to perform an initial alignment, which is then optimised running the Iterative Closest Point (ICP) algorithm [34]. The method selected to create a watertight triangle mesh from the merged point cloud is Screened Poisson Surface Reconstruction (resolution depth: 9, screening weight: 2, samples-per-node: 10) [35]. While the laser mesh is directly uniformly remeshed to an average edge length of 0.6 mm, an additional step is needed on the photogrammetric mesh prior to the remeshing, to remove inaccuracies due to the higher scanning error of this technique, e.g. foldovers in the mesh. This is carried out employing the OpenFlipper “Mesh Repair” tool [36], using an automated procedure to detect foldovers, by selecting edges having an angle difference bigger than 30° , deleting them and algorithmically filling the holes [37]. The photogrammetric geometry is reconstructed up to an arbitrary scale; thus, a scaling factor estimated on measurements taken on the real and scanned geometry is applied. The uniform meshes are aligned to the coordinate system using a procedure in which 4 points in the mesh are manually selected, corresponding to the left and right closed ear canal centres and the inferior margin of the left and right orbits. These points are used to translate and rotate the mesh such that the origin corresponds to the interaural centre, the y -axis passes through the ear canal centres and the x -axis is parallel to the Frankfurt plane [11]. Thus, the meshes are aligned to the measurement positions on which the HRTFs are computed. These are defined by the distance (r), the azimuth angle (θ) and the elevation angle (ϕ), given in the interaural-polar coordinate system having $\theta \in [-90^\circ, 90^\circ]$ and $\phi \in [-90^\circ, 270^\circ]$, with positive θ and $\phi \in (0^\circ, 180^\circ)$ relating to left and above horizontal plane positions, respectively. Given that this alignment procedure requires manual input, its outcome can vary, and this could have an impact on the computed HRTFs. To assess this effect an analysis is carried out by repeating the alignment of the laser mesh with the axes, using the ICP algorithm to minimise the distance between the meshes related to the repeated alignment attempts and a reference one, for which the distance from its left-right mirrored version is minimised. The maximum displacement and rotation errors of the repeated alignments

are estimated and used to analyse the effect of misalignment on the numerical HRTFs. Furthermore, the ICP algorithm is employed to minimise the distance between the reference aligned laser mesh and the photogrammetric mesh by optimising the alignment and scaling of the latter, to focus the analyses on the effect of photogrammetric scanning errors rather than misalignment and incorrect scaling [14]. Nonetheless, HRTFs computed on the photogrammetric mesh with the original manual alignment are compared to the results after the ICP algorithm application, to further evaluate the effect of alignment and scaling errors. The scanned meshes are further processed to make them suitable for numerical computations. A grading algorithm is applied to obtain meshes with an edge length (l_ϵ) ranging between a minimum (l_{\min}) and maximum (l_{\max}) value, set to 1 mm and 15 mm, respectively. The graded edge length is calculated as $l_\epsilon = l_{\min} + (l_{\max} - l_{\min})\mu(d_\epsilon)$, with the grading function $\mu(d_\epsilon) = 1 - \cos^2(\pi d_\epsilon/2)$, and d_ϵ being the relative distance from the mid-point of a selected element at the centre of an ear canal and an edge, divided by the maximum distance between this midpoint and the farthest edge in the mesh [38]. This procedure decreases the number of elements from more than 1 000 000 to around 18 000, achieving a great reduction in computational cost while maintaining adequately accurate results. HRTFs calculated on the audible frequency range on similarly graded meshes results in terms of numerical accuracy and predicted sound localisation performance [38]. Thus, one graded mesh per ear is generated, on which the HRTFs are computed.

2.2 Geometrical error analysis

Several metrics have been defined to assess the geometrical error affecting scanned point clouds and meshes. These metrics are generally evaluated on the Euclidean distance ($d(x, y)$) between each point (x) of a reference geometry (X) and the points (y) of a target geometry (Y). If X is orientable a signed distance can be defined, where a positive sign indicates that y is outside the surface of X [39]. Accuracy (**Acc**) and completeness (**Cmp**) are two metrics defined in [14] to evaluate different techniques for scanning the outer ear, and are used in the following analyses since they allow a direct comparison of the results. **Acc** is defined as the 95th percentile of the absolute distance from each y to the closest x , while **Cmp** is defined as the percentage of $x \in X$ at a distance smaller than 1 mm from the closest y . Hence, while the first evaluates the maximum distance from the reference for 95% of the scanned points, the latter focuses on incomplete parts in the scans, e.g. holes. Mean (**AvG**) and maximum (**Max**) absolute distance between y and the closest x are also commonly employed metrics [15]. Furthermore, another metric is introduced, namely the Chamfer Distance (**CD**), which is typically used for measuring similarity between point sets [40]. The **CD** is defined as:

$$\mathbf{CD}(X, Y) = \frac{\sum_{x \in X} \min_{y \in Y} d(x, y)^2}{|X|} + \frac{\sum_{y \in Y} \min_{x \in X} d(y, x)^2}{|Y|}, \quad (1)$$

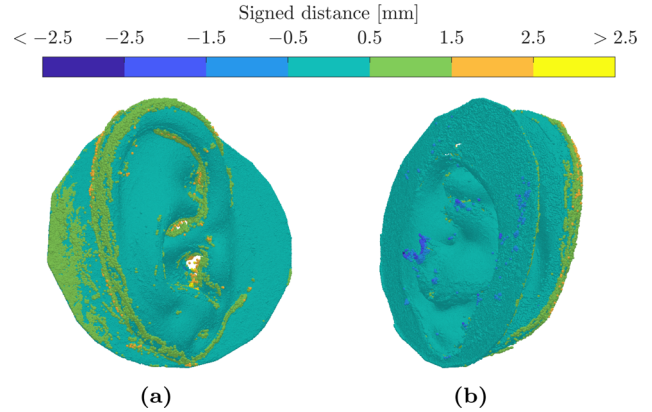


Figure 3. KU100 right ear point cloud obtained through photogrammetry scanning of the optically treated dummy head. The colours show the signed distance from the laser mesh, cropped at ± 2.5 mm. (a) Front side and (b) back side.

with $|X|$ and $|Y|$ denoting the cardinality of X and Y , respectively. This metric is considered to be relevant since it takes into account both the distance from y to x and from x to y ; thus, it is sensitive to both inaccuracy and incompleteness of the scanned geometry.

2.3 Results

The right ear photogrammetric point cloud obtained on the sprayed pinna is presented in Figure 3, alongside the signed distance of each point from the laser mesh. Noise and incompleteness can be noticed at the concave ear structures, e.g. the concha and cymba, owing to a low visibility. The most convex parts, e.g. the helix, tend to be surrounded by erroneous points with a generally positive distance, probably arising from artefacts in the images due to the non-optimal lighting conditions.

An attempt has been made to scan the left ear in an untreated scenario, i.e. without using the scanning spray. The comparison between these results and those obtained on the optically treated scans are presented in Table 1, using the laser mesh as the reference for evaluating the geometrical metrics. **Acc** and **Cmp** evaluated on the optically treated and untreated scans are better than the outcome reported in [14] for the average scan of the full ear of real subjects, i.e. 1.70 mm and 80.9%, respectively, for a scan in which black water colour has been sprayed on the ears, and 2.07 mm 69.8% in the untreated case. This might be related to the use of a dummy head, since the translucency of the skin and involuntary movements of human subjects are considered as factors capable of hindering the scanning procedure. However, photogrammetric scans of a plaster ear using a professional camera setup show **Acc** and **Cmp** of 0.36 mm and 97.3%, respectively [14]. The poorer results obtained on this KU100 scan might be related to the use of a smartphone camera with lower resolution and the dark and uniform texture of the dummy head, which can hinder the photogrammetric reconstruction algorithm. The scanning outcome related to the two sprayed ears shows comparable results. The results obtained on the ear

Table 1. Geometrical error metrics of optically treated and untreated photogrammetric KU100 ear point clouds. The KU100 laser ear meshes are used as reference.

Point cloud ID	Acc [mm]	Cmp [%]	Avg [mm]	Max [mm]	CD [mm ²]
Untreated left ear	1.13	78.5	0.31	6.01	1.26
Sprayed left ear	0.71	95.8	0.18	3.21	0.26
Sprayed right ear	0.53	91.6	0.14	3.70	0.43

without optical treatment are worse in all the analysed metrics, with **Avg** and **Max** values almost twice as high, while the **CD** is the most impacted metric, likely relating to the low completeness of the untreated scan.

Figure 4 presents the uniform right ear meshes, prior to applying the grading algorithm, obtained through laser and photogrammetry scanning on the optically treated dummy head. Figure 4a displays the laser mesh on which the signed distance from the photogrammetric mesh is plotted, while in Figure 4b the photogrammetric mesh is shown alongside the signed distance from the laser mesh. It is noticeable that the largest differences between the two meshes are found in the concave ear structures, where the photogrammetric point clouds show noise and incompleteness. The sign of the distance, negative on the laser mesh and positive on the photogrammetric mesh, implies that the photogrammetric geometry tends to lie outside the surface of the laser mesh; hence, reducing the depth of the concave pinna structures. This is attributed to the lack of points and high noise observed in the scanned data and the subsequent application of the meshing and hole filling algorithms, which interpolate the shape based on the surrounding points [35, 37], and are expected to only partially decrease the difference between the meshes at these locations.

The geometrical metrics evaluated on the left and right ear meshes deriving from the photogrammetric scans with spray, and the left ear mesh from the untreated scenario, are reported in Table 2. This analysis is limited to the front of the ear since the back is considered to be acoustically non-significant [19]. These metrics differ from the point cloud results since the meshing procedure interpolates the scanned points; moreover, the analysis is limited to the frontal part of the pinnae. The meshes deriving from the sprayed scans show a better outcome in all the considered metrics. Low differences are seen between the sprayed left and right ear results. Comparison to previous studies can be done on the maximum deviation reported in [15] for the photogrammetric scan of another dummy head, in which a difference of 5 mm from a reference scan is reported at the concha. An overall mean deviation of 1.98 mm is indicated on the full head scan, which compares to a considerably smaller **Avg** of 0.21 mm in the current head scan with scanning spray. This can be attributed to differences in the scanning procedure, since in [15] no optical treatment is employed and a low number of images is used for the reconstruction, i.e. 16 on the full head plus 10 on each ear. Given the overall worse results displayed in the untreated left ear mesh, and to focus on the minimal deviation achievable with photogrammetry, the following

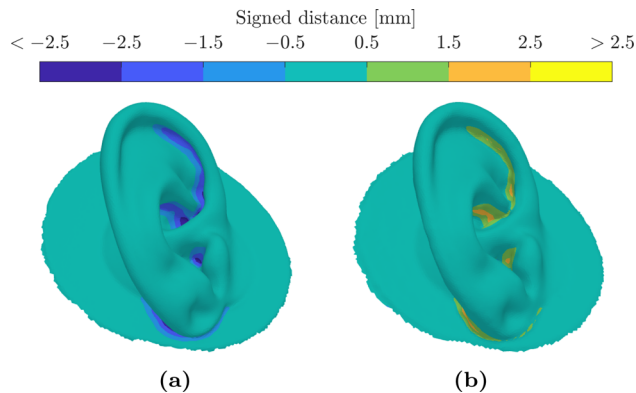


Figure 4. KU100 right ear uniform mesh obtained through laser (a) and photogrammetry (b) scanning. The colours show the signed distance from the laser to the photogrammetric mesh (a) and vice-versa (b), cropped at ± 2.5 mm. (a) Laser mesh and (b) photogrammetric mesh.

analyses are conducted only on the photogrammetric scans with scanning spray.

The alignment of the laser mesh with the axes is repeated three times using the algorithm described in Section 2.1; the deviation from these meshes to the reference aligned mesh is calculated in terms of translation and rotation error. A maximum translation error of 2.7 mm is found on the x -axis, while for the y and z -axis this is below 1.5 mm. Maximum values of rotation error of 2.4° are observed for the y -axis, while this value is below 1° for x and z -axis rotations. For the photogrammetric case two meshes are assessed, the first manually aligned to the axes with the described procedure and the second after using the ICP algorithm to minimise the alignment and scaling error of this mesh to the reference aligned laser mesh. The maximum translation and rotation error in this case reach maximum values of 3 mm for the x -axis and 2.8° for the y -axis, respectively. The estimated optimal scaling factor between the two meshes is 1.01, indicating that the manually scaled photogrammetric mesh is slightly smaller than the laser geometry.

The right ear graded head mesh deriving from the laser scan is presented in Figure 5, in which the dimension of each element, calculated as the average between its edges, is highlighted. It can be seen that elements at the right ear have dimensions below 3 mm and their size increases gradually in function of the distance from the right ear canal. On the left side the overall edge size is above 7 mm and it is possible to notice that the left ear geometry is lost, which relates to a low element count in the graded mesh.

Table 2. Geometrical error metrics of optically treated and untreated photogrammetric KU100 ear meshes. The KU100 laser ear meshes are used as reference.

Mesh ID	Acc [mm]	Cmp [%]	Avg [mm]	Max [mm]	CD [mm ²]
Untreated left ear	1.46	83.6	0.40	3.39	1.16
Sprayed left ear	1.04	90.4	0.23	3.18	0.65
Sprayed right ear	0.87	90.5	0.19	2.28	0.50

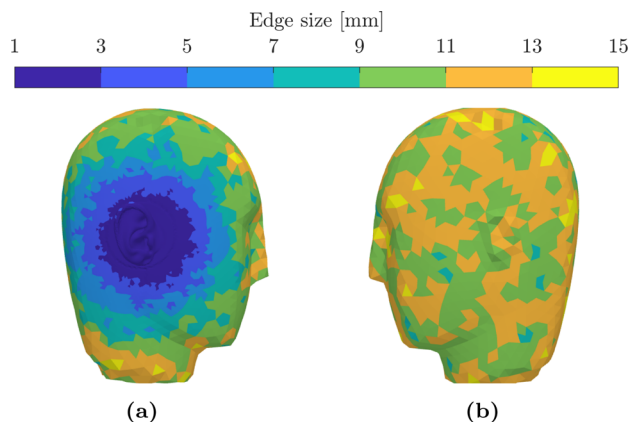


Figure 5. KU100 mesh obtained through laser scanning after application of the grading algorithm for the right ear. The colours show the average edge size of each element. (a) Right side and (b) left side.

3 HRTF acquisition and assessment

3.1 Numerical HRTF computation

The BEM is applied for the numerical computation of the laser and photogrammetric HRTFs using the reciprocity principle; hence, applying an excitation at the ear and computing the results in a surrounding grid [41]. The excitation is defined as a constant normal velocity of one element with edge length of around 1 mm at the centre of the closed ear canal. This choice is shown to result in the computation of HRTFs comparable to acoustically measured ones [11]. Sound hard boundary conditions are imposed, since they relate to the best match between measured and simulated KU100 HRTFs below 7 kHz, in comparison to cases in which the measured impedance of the dummy head material is assigned to the full model or the pinnae elements alone [42]. The medium used in the simulations is air at a temperature of 20 °C, given that no temperature information is available for the measured data. This corresponds to a density (ρ_0) of 1.2041 kg m⁻³ and a speed of sound (c) of 343.21 m s⁻¹. Numerical simulations are carried out for frequencies ranging from 0.1 kHz to 22 kHz in steps of 0.1 kHz; the results are evaluated on a full spherical grid of 5042 points with centre corresponding to the interaural centre, $r = 1.2$ m, θ and ϕ resolution of 2.5° and 5°, respectively. The results of the simulation are post-processed to obtain HRTFs [4]. The inverse Fourier transform is applied to the complex spectra, converting them in Head-Related Impulse Responses (HRIRs) of

441 samples at a sampling rate of 44.1 kHz. The results are further converted into Directional Transfer Functions (DTFs), dividing each of the HRTFs spectra by its minimum phase Common Transfer Function (CTF), calculated as the spatially weighted average magnitude at each frequency bin over all directions [43]. Hence, the DTFs isolate the directional dependent HRTFs components, while the CTF contains the common spectral features among all measurement positions [44].

To further reduce the deviation between simulated and measured DTFs, potentially arising from a mismatch in ρ_0 and c , a frequency scaling factor (α) is estimated to optimise the match between the data, and applied to the numerical results by resampling the HRIRs according to the computed α [13]. Furthermore, additional simulations are performed on the laser mesh with displaced measurement grids to assess the effect of misalignment. This is done by computing the laser mesh HRTFs on spherical grids to which a rotation of $\pm 2.5^\circ$ on each axis is separately applied, resulting in six additional HRTFs computed at displaced locations. These displaced positions are then redefined as the ones from the original unmodified grid; hence, resulting in a rotation of the head rather than the measurement grid. The value of 2.5° is chosen since it is slightly higher than the maximum rotation error observed in the repeated alignments of the laser mesh. The translation error is not taken into account since its effect is assumed to be lower than that of rotation error, given that a rotation of 2.5° at $r = 1.2$ m corresponds to a displacement of the measurement grid of more than 50 mm; hence, considerably higher than the maximum observed translation error, below 3 mm. The HRTFs are also computed on the initially manually aligned photogrammetric mesh, prior to using the ICP algorithm to optimise its alignment to the reference aligned laser mesh.

3.2 HRTF dataset creation

The numerical results are assessed by comparing them to acoustically measured HRTFs taken from a database of repeated KU100 measurements [45]. A subset of 8 HRTFs is extracted, maintaining only data with a common set of $\phi \in [-30^\circ, 70^\circ] \cup [110^\circ, 210^\circ]$ and a resolution of 10°, to avoid potential interpolation errors in the following analyses. An overview of the dataset (\mathcal{D}) of KU100 HRTFs used in this study is presented in Table 3 (associations of database HRTFs to the current study are reported), including the numerical data obtained on the scans [46]. To match the processing carried out on the acoustic measurements, the numerical results are low-pass filtered at 18 kHz.

Table 3. Dataset of analysed KU100 HRTFs.

ID	Database ID	ϕ range [°]
<i>aco1</i>	ClubFritz1	−45:5:225
<i>aco2</i>	ClubFritz2	−35:5:215
<i>aco3</i>	ClubFritz3	−40:10:70 ∪ 110:10:220
<i>aco4</i>	ClubFritz4	−40:10:70 ∪ 110:10:220
<i>aco5</i>	ClubFritz8	−40:10:220
<i>aco6</i>	ClubFritz9	−50:5:230
<i>aco7</i>	ClubFritz10	−80:10:260
<i>aco8</i>	ClubFritz12	−30:10:80 ∪ 100:10:210
<i>las</i> ^a	hrir_ku100_las_sph	−90:5:270
<i>pho</i> ^b	hrir_ku100_pho_sph	−90:5:270

^a HRTFs computed on the laser scanned mesh.

^b HRTFs computed on the photogrammetry scanned mesh.

Measured HRTFs are also converted to DTFs and CTFs. The spatial grid for calculating the CTFs is selected to be in common to all the data in \mathcal{D} ; hence, entailing the common set of ϕ and $\theta \in [-90^\circ, 90^\circ]$ with a resolution of 10° . The weights for the CTFs evaluation are calculated as the area of spherical rectangular segments with dimensions of 10° in θ and ϕ , centred around each measurement point [43].

3.3 Interaural features analysis

Horizontal plane interaural features are computed on the HRTFs in \mathcal{D} and compared. The ITDs are evaluated on the HRIRs by leading-edge detection:

$$\text{ITD}(\theta) = \tau_{R,\eta}(\theta, \phi = 0^\circ) - \tau_{L,\eta}(\theta, \phi = 0^\circ), \quad (2)$$

where τ_R and τ_L represent the instants at which the 10 times upsampled right and left ear impulse responses first reach a certain percentage (η) of their maximum peak amplitude [4]. A threshold value of -10 dB is selected in equation (2), i.e. $\eta \approx 32\%$. No additional low-pass filtering or post-processing is applied to the HRIRs. The ILDs are evaluated at each frequency bin as:

$$\text{ILD}(\theta, f) = 20 \log_{10} \left| \frac{H_R(\theta, \phi = 0^\circ, f)}{H_L(\theta, \phi = 0^\circ, f)} \right|, \quad (3)$$

where H_R and H_L represent right and left ear HRTFs, respectively, and f indicates frequency [4]. Broadband ILDs are obtained taking the mean of equation (3) over the frequency range from 0.5 kHz to 18 kHz, in linear steps of 0.1 kHz. This limited frequency range is also used in the following comparison metrics. The higher limit matches the filtering applied to the experimental data, while the lower limit is used to avoid low frequencies where the measured HRTFs shows a steep increase in deviation, as further discussed in Section 3.7. Although using a linear frequency axis for computing mean ILD values might emphasise high frequency details, this approach is found to be in line with metrics used in previous studies, calculating the ILDs as broadband RMS level differences between left and right ear HRIRs [15].

3.4 Symmetry analysis

Symmetric properties of the HRTFs are assessed using a right-left symmetry coefficient (ρ) [27], defined as:

$$\rho(\theta, \phi, f_k) = \frac{\sum_f \tilde{H}_R^k(\theta, \phi, f) \tilde{H}_L^k(\theta, \phi, f)}{\sqrt{\sum_f \tilde{H}_R^k(\theta, \phi, f)^2 \sum_f \tilde{H}_L^k(\theta, \phi, f)^2}}, \quad (4)$$

where \tilde{H} indicates the single-sided amplitude spectrum of gammatone filtered HRTFs using a spacing of the k th auditory filter-band, with centre frequency f_k , corresponding to 1 Equivalent Rectangular Bandwidth (ERB). Equation (4) is used to evaluate the symmetry coefficient on the median plane (ρ_M) and the horizontal plane (ρ_H). For the latter, the left ear HRTFs at each θ are mirrored to the right. The mean value is taken across the auditory filters.

3.5 Sagittal localisation analysis

Perceptual attributes, not measurable with objective metrics alone, have a strong influence on spatial hearing. These are estimated using a sagittal plane localisation model [3], implemented in the Auditory Modelling Toolbox version 0.10.0 [47]. The elevation perception in sagittal planes is predicted by a comparison process between an incoming sound filtered with an arbitrary set of “target” filters, and an internal “template” set representing a subject’s own auditory filters. The outcome of this comparison is an estimation of the localisation error in terms of Quadrant Error (QE) and Polar Error (PE). The first quantifies the percentage of predicted responses falling in a different quadrant than the true position of the target sound source, the second relates to the angular accuracy of the localisation within $\pm 90^\circ$ from the target position. A listener specific sensitivity parameter (S) is employed to take into account non-acoustic factors necessary to model high inter-individual differences observed in real sound localisation experiments [48]. It should be mentioned that this model estimates results of static localisation experiments by approximating the perceptual factors related to spatial hearing, and its results seem to transfer well to sound perception in VR scenes [7]. Nevertheless, the outcome should not be taken as a precise perceptual assessment of the spatial cues in the data, rather as a perceptually inspired analysis providing additional insight.

Given the variability reported in the experimental data, making it challenging to identify a valid reference, an analysis is carried out by computing and comparing the sagittal localisation performance obtained with each pair of DTFs in \mathcal{D} , to assess their perceptual similarity. Median plane DTFs, resampled at 48 kHz, are extracted for the largest set of common ϕ , i.e. $\phi \in [-30^\circ, 70^\circ] \cup [110^\circ, 210^\circ]$ in steps of 10° ; QE and PE are evaluated on this range using each of the DTFs as template and target. A value of $S = 0.21$ is selected, corresponding to the lowest S used in [3] to match the model predictions to the outcome of real localisation experiments, and representing a “good localiser”. Gammatone filtering with 1 ERB is applied in the model. Similarly to [9], the

results are stored in two all-against-all matrices, i.e. the QE and PE Matrices (QEM, PEM), defined as:

$$\text{QEM}_{i,j} = \text{QE}(H_i^r, H_j^t, S) \quad \forall i, j \in \mathcal{D}, \quad (5)$$

with i and j corresponding to indices of the DTFs in \mathcal{D} used as template (r) and target (t). The PEM is similarly obtained. Therefore, the rows and columns of these matrices have a common set of DTFs used as template and target, respectively. The diagonal elements represent the localisation errors obtained with the same DTFs as both template and target, i.e. the baseline condition, corresponding to a listener localising sound processed with individual auditory filters [48]. Small deviations between the spectral characteristics of the various DTFs create differences in their baseline performance, which can bias the comparison. Therefore, the results of equation (5) are normalised by subtracting the baseline error of each template from the errors obtained with the same template but different target DTFs, obtaining the ΔQEM defined as:

$$\Delta\text{QEM}_{i,j} = \text{QEM}_{i,j} - \text{QEM}_{i,i}. \quad (6)$$

Applying equation (6) to the PEM, the ΔPEM is obtained. The entries of these matrices can be seen as a perceptual distance between the tested DTFs; therefore, employing hierarchical clustering with average linkage function on the rows of ΔQEM and ΔPEM , dendrograms are created and used to identify clusters of template DTFs characterised by a similar localisation accuracy. Application of hierarchical clustering to assess the similarity between HRTFs has been proposed in [49]; however, in that study the employed distance metric is the correlation coefficient of HRTFs ratings between subjects, and the rating is given by a perceptual evaluation of the binaural rendering of sound trajectories on the horizontal and median plane, using individual and non-individual HRTFs. The choice of using the outcome of the sagittal plane localisation model in the current study is driven by the fact that this can be used as a perceptually inspired distance metric between HRTFs pairs, even for the case in which the subject of the study is a dummy head. To define a threshold for cluster formation, measured DTFs of 4 different human subjects are included in the analysis, corresponding to subjects *nh2*, *nh5*, *nh8*, *nh10* of the Acoustic Research Institute (ARI) database [50], and the lowest QE and PE distance between these DTFs is used to cut the dendrograms. These levels are chosen to approximately represent the distance between non-individual data. It is expected that the majority of the KU100 DTFs will form a cluster since they originate from the same dummy head; however, noise and errors could influence their localisation performance and decrease their similarity. The photogrammetric scanning error observed in Section 2.3 is expected to affect the spectral features of the related DTFs, and decrease their modelled perceptual similarity to the other KU100 DTFs.

To better assess the perceptual accuracy of the numerical photogrammetric results, their localisation error is compared to that of the measured and laser DTFs. Based on the outcome of the previous analysis, one of the DTFs is

selected as the reference and used as template; the other KU100 and ARI DTFs are grouped depending on their similarity to this reference. The expected QE and PE using the various groups of DTFs as the target are compared to the localisation error of the photogrammetric DTFs.

3.6 Spectral difference analysis

The results of the employed sagittal localisation model are generally strongly correlated with objective metrics such as the Mean Spectral Distortion (MSD) [25]. The latter is also highly correlated with the Inter-Subject Spectral Difference (ISSD) [24]. Therefore, it is assumed that DTFs pairs showing a similar localisation performance share similar spectral characteristics. This is tested by computing the MSD and the ISSD on the DTFs in \mathcal{D} and assessing their correlation with QE and PE. The MSD between DTFs i and j is calculated as:

$$\text{MSD}_{i,j} = \frac{\sum_{\Psi} \sqrt{\frac{1}{N_f} \sum_f \left(20 \log_{10} \left(\frac{|H_i(\theta, \phi, f)|}{|H_j(\theta, \phi, f)|} \right) \right)^2}}{N_{\Psi}}, \quad (7)$$

taking the mean between the N_f frequency bins, the two ears and the number N_{Ψ} of incidence angles Ψ corresponding to the limited ϕ range selected in Section 3.2. The ISSD is defined as:

$$\text{ISSD}_{i,j} = \frac{\sum_{\Psi} \text{Var}_{f_k} (\hat{H}_i(\theta, \phi, f_k) - \hat{H}_j(\theta, \phi, f_k))}{N_{\Psi}}, \quad (8)$$

with Var_{f_k} indicating variance averaged over f_k of \hat{H} representing the dB amplitude of gammatone filtered DTFs. The results are also averaged between the two ears and the considered incidence angles.

3.7 Results

Prior to all the analyses, frequency scaling is applied to the numerical HRTFs with the average α computed between the numerical laser and all the measured DTFs in \mathcal{D} ; the chosen optimisation metric is the ISSD. The resulting average α has a value of 0.99, corresponding to a mean decrease in ISSD of 0.08 dB². The reported α relates to a small shift of the numerical DTFs spectral features to higher frequencies, to obtain a better match with the measured data. An α close to 1 is also reported in [13]; similarly to that study, this is attributed to the value of c in the simulations being mismatched and, in this case, lower than the average during the measurements. This mismatch probably arises from the temperature value used in the simulations, likely lower than the actual value during the measurements.

The acoustically measured, referred to as acoustic, and numerically simulated, referred to as numerical, KU100 DTFs and CTFs in \mathcal{D} are initially visually compared. Figure 6 shows the acoustic and numerical right ear DTFs magnitude for a representative frontal direction, and CTFs. The acoustic data is plotted as an envelope representing the 25th and 75th percentile of the DTFs and CTFs magnitude

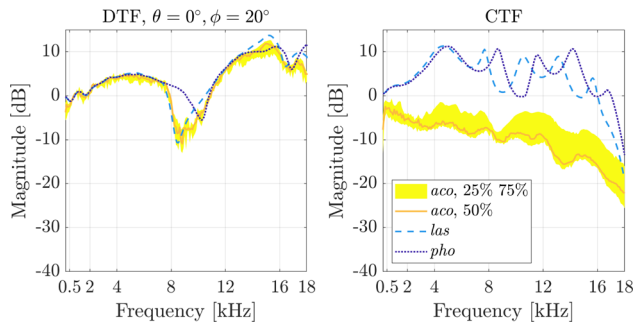


Figure 6. Magnitude of acoustic (*aco*), numerical laser (*las*) and photogrammetric (*pho*) KU100 right ear DTFs for frontal source, and CTFs. The *aco* results are shown as an envelope representing the 25th and 75th percentile of the measured DTFs and CTFs magnitude, the 50th percentile is also displayed.

distributions, the 50th percentile is also highlighted, while the numerical results are displayed individually. Large discrepancies are observed between the numerical and acoustic CTFs, making a direct comparison challenging. This is attributed to the fact that the experimental HRTFs do not present a common level between the various measurements, and no gain normalisation is applied in the current analysis. Furthermore, some of the measured data still contains the responses of the KU100 mics [27]. High variance is also noticed within the acoustic results, which can stem from several factors related to the measurement procedure, e.g. a different alignment of the dummy head or non-linearity of the acquisition setup [27]; their in-depth analysis is out of the scope of the current manuscript. A good match in all DTFs is observed for frequencies below 7 kHz; at higher frequencies the laser results show similarities to the acoustic data, while the photogrammetric curve presents large discrepancies. Although the photogrammetric DTFs seem to fit the acoustic data between 11 kHz and 15 kHz in the displayed direction, this match is not consistent for all the tested angles. In the acoustics DTFs, an increase in deviation is observed for frequencies below 0.5 kHz, with a median value tending towards a level lower than 0 dB, which is attributed to errors affecting some of the experimental measurements, showing a steep decrease in magnitude towards low frequencies, also noticeable in [27]; therefore, this lower limit is chosen for the following analyses. The match between the curves from 0.5 kHz to 8 kHz relates to the weak effect of the pinna due to its small dimension in comparison to the wavelength. Differences between laser and acoustic DTFs could be explained by mismatch between measurement and simulations, e.g. inaccuracies in the scan or the choice of rigid boundary conditions which differs from the real case; however, the latter is a common choice in numerical HRTFs computation and it is indicated to relate to only small perceptual difference in localisation performance from experimentally measured data [16]. Previous results comparing simulated HRTFs on a KU100 MRI scan and a reference measurement show a relatively good match at low frequencies but a dilation above 5 kHz, potentially attributed to the

mismatch between real and virtual microphone positions [28]; this is not noticed in the current results. The discrepancy observed in the photogrammetric data, showing a significant upward shift of more than 1 kHz in the centre frequency of the first pinna notch (N1) [22], appearing at around 8 kHz in the laser and median of the acoustic DTFs for the displayed direction, cannot be attributed to differences in microphone position, given that the deviation at the closed ear canal between laser and photogrammetric meshes is below 1 mm. Comparing the centre frequency of N1 in the full median plane for the laser and photogrammetric HRTFs, this appears to be shifted upwards in the latter by an average of 1.4 kHz on all the tested elevation angles. The difference between photogrammetric and laser meshes reported in Section 2.3, showing maximum values exceeding 2 mm, is assumed to be capable of creating this mismatch, based on results indicating that modifying the pinnae by moving all data points inward or outward by 1 mm caused a shift of at least 500 Hz in the HRTFs spectral peaks and notches above 7 kHz [12]. The laser and photogrammetric CTFs show agreement at the first peak around 4.5 kHz, while at higher frequencies the photogrammetric peaks tend to be shifted upward by as much as 1 kHz. A similar behaviour is observed in [25] when decreasing the depth of some control points in the parametric pinna model. The first peak in the HRTFs (P1) is due to an overall resonance developing in the pinna, while the second and third peaks (P2, P3) relate to resonance modes developing separately in the concha, cymba and fossa. The N1 is reported to be generated by resonances mainly appearing at the cymba and fossa, changing in amplitude and frequency depending on the elevation angle. No additional information is given for higher frequency peaks and notches [22]. Given that the main deviation between photogrammetric and laser DTFs relates to N1, P2 and P3, associated to resonances in the cymba and fossa, this is attributed to the photogrammetric scanning error, highly affecting these pinna structures.

It should be mentioned that the frequency scaling approach has also been tested on the photogrammetric DTFs, by finding an optimal factor reducing the deviation between these and the measured DTFs, and the results indicate an average α of 1.08. This value, relating to a high compression of the spectral features, is not considered to be caused by a mismatch in c , but rather by the algorithm trying to minimise the deviation between photogrammetric and measured DTFs.

The comparison of ITDs and ILDs of the data in \mathcal{D} is plotted in Figure 7. The acoustic results are presented as an envelope representing the 25th and 75th percentile of the estimated ITDs and ILDs distribution, the 50th percentile is also highlighted, and compared to the interaural features of the numerical results. A slightly different coordinate system is used in this and the following plots taking into account the horizontal plane, having $\theta \in (-180^\circ, 180^\circ]$ for ease of visualisation. Although broadband ILDs tend to vary with distance from the sound source [51], their analysis is nonetheless presented given that the measured data is acquired at comparable distances between 0.9 m

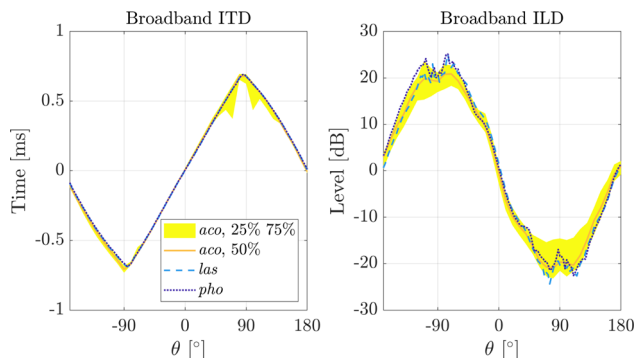


Figure 7. Horizontal plane ITDs and ILDs of acoustic (*aco*), numerical laser (*las*) and photogrammetric (*pho*) KU100 HRTFs. The *aco* results are shown as an envelope representing the 25th and 75th percentile of the measured data ITDs and ILDs, the 50th percentile is also displayed.

and 2.0 m [27], while the numerical HRTFs are computed at a common distance of 1.2 m. Higher deviation in the acoustic ITDs and ILDs tends to appear for lateral directions, i.e. $\theta = \pm 90^\circ$. An increase in ITD variations proportionally to the distance from the median plane is also observed in [27], and can be attributed to positional misalignment. The maximum ITD and ILD deviation between the acoustic data shows values of 500 μs and 8.4 dB, respectively. The high deviation in ITDs for some of the considered θ close to lateral directions can stem from measurement error, which seems to affect the low frequency components of some experimentally acquired HRTFs. Although the same algorithm for ITD calculation is used, horizontal plane ITD analysis on the full database from which the measured HRTFs are extracted reports a maximum ITD deviation of 235 μs [27]. The differences between these and the current results are difficult to determine and might relate to different post-processing of the data. Application of alternative methods for estimating the ITDs, e.g. low-pass filtering the HRIRs at 3 kHz prior to applying threshold detection, which is identified as the most perceptually relevant procedure for ITD estimation [52], results in increased deviation in the ITDs. Nonetheless, the estimated median acoustic ITDs do not seem to be impacted by the deviation observed in the data distribution; therefore, the median is used to assess the difference between acoustic and numerical ITDs. Differences between the computed interaural features exceed the anechoic Just Noticeable Difference (JND) values for ITDs and ILDs of around 20 μs and 1.0 dB, respectively [53]. Interaural features of the numerical results lie within 44 μs and 3.3 dB from the acoustic median ITDs and ILDs, respectively. In comparison, differences between the ITDs measured and simulated on an MRI scan of the same dummy head reach a maximum value of 100 μs , attributed to deviation between the real and virtual microphone position [28]. The numerical ITDs and ILDs tend to generally lie within the plotted distributions of the acoustic interaural features, although for some lateral directions the numerical ILDs appear to be outside this interval. This could be due to a slightly different

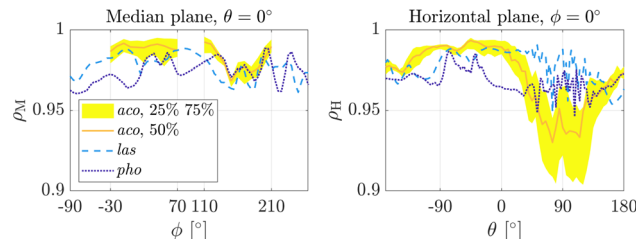


Figure 8. Median and horizontal plane right-left symmetry coefficients of the acoustic (*aco*), numerical laser (*las*) and photogrammetric (*pho*) KU100 HRTFs. The *aco* results are shown as an envelope representing the 25th and 75th percentile of the measured data symmetry coefficients, the 50th percentile is also displayed.

alignment of the dummy head between measurement and simulations or the choice of using rigid boundary conditions in the latter, deviating from the measurement case. Comparing the interaural features of the laser and photogrammetric results, deviation up to 6 μs and 3.7 dB are observed. ITD and ILD of a reference and a photogrammetric scan of a different dummy head show maximum deviations of 300 μs and 4 dB for lateral directions [15]. The worse results reported in that study could be related to the low number of images used for the photogrammetric reconstruction, as mentioned in Section 2.3.

The outcome of the symmetry analysis on the median and horizontal plane is shown in Figure 8, where the acoustic results are presented as an envelope representing the 25th and 75th percentile of the estimated coefficients distributions, the 50th percentile is also highlighted. Furthermore, on the median plane only the symmetry coefficients of the numerical data span the full range of ϕ , while the acoustic results are plotted on the common set of ϕ across the measurements. Median plane results show that ρ_M tends to be generally high for frontal directions, i.e. $\phi \in [-30^\circ, 70^\circ]$, and to decrease towards the rear of the head, i.e. $\phi > 110^\circ$. Similar results are reported in [27]. The laser data shows a lower symmetry than the acoustic median ρ_M for frontal directions, close to the lower limit of the acoustic distribution interval, while the photogrammetric results show higher asymmetry, probably related to the scanning error differently affecting the two ears. On the back of the head this pattern seems to be inverted, with the photogrammetric results showing a higher symmetry than the laser data, close to the median value of the acoustic results. The reasons for this behaviour are difficult to determine and might be related to the scanning and alignment procedure. The ρ_H shows similar trends to the frontal median plane results on the ipsilateral side, i.e. $\theta < 0^\circ$, while on the opposite side the acoustic results show a lower value, reaching a minimum around 90° ; similarly to [27], this is attributed to the lower signal-to-noise ratio of contralateral HRTF measurements due to head shadowing. The numerical results, not affected by this problem, show a higher symmetry at these locations. Although not displayed, analysis of ρ_M and ρ_H for each f_k are carried out, to directly compare the current results to the same

analyses conducted in [27] on the full measured HRTFs database. In that study, the symmetry coefficient is presented as $1 - \rho$, and maximum values of 0.4 are reported for two analysed HRTFs. These maximal values appear on the horizontal plane for contralateral positions at frequencies around 9 kHz and 15 kHz, and on the media plane for rear positions and frequencies above 7 kHz. Similar patterns are observed in the two numerical HRTFs symmetry coefficients, indicating high symmetry below 7 kHz and decreasing for higher frequencies. Comparable maximal coefficients of 0.44 are observed for $1 - \rho_M$ of the numerical data at frequencies of 15 kHz and elevations below the horizontal plane. For $1 - \rho_H$, values reaching 0.49 are seen on the laser results at the contralateral side and at a frequency of 15 kHz, and 0.54 on the photogrammetric data at similar angles and frequencies. As suggested in [27], causes of the observed asymmetry could relate to inherent asymmetric properties of the dummy head or repositioning variations of its removable pinna elements. Furthermore, the increased asymmetry observed in the numerical results could stem from random errors during the scan which could differently impact the left and right ear acquisition.

The dendrograms obtained on the ΔQEM and ΔPEM matrices are shown in Figure 9. Different clusters are observed, indicating that the localisation performance tends to be different between the data in \mathcal{D} . The results of the QE and PE metrics show similar trends; therefore, the following observations apply to both results. Cutting the dendrograms at the chosen thresholds two clusters appear, the first containing the numerical laser and several acoustic KU100 DTFs, and the second containing only acoustic data. The fact that the laser results are included in the first group, relating to a low distance to various acoustics data, indicates similarity of these numerical and acoustic DTFs. The second cluster, showing a generally higher distance between the data, is thought to contain DTFs affected by a higher level of measurement noise, capable of hindering their elevation perception. The link between these two clusters, just above the defined threshold, indicates that these DTFs are more similar between them than to the ARI DTFs, which is expected given that the latter belong to different subjects. Although the photogrammetric results show a link to the other tested KU100 DTFs, the distance between them indicates that the geometrical error in this scan has a detrimental effect on the HRTFs localisation cues. On the assumption that a high distance in QE and PE between the DTFs relates to discrepancies in their perceptual attributes, the first cluster is defined to contain DTFs creating an accurate perceptual experience, while the second is thought to relate to inaccuracies in spatial sound perception. This assumption should however be tested through real perceptual experiments.

The low distance of the laser results to the acoustic DTFs in the first cluster seems to validate this numerical acquisition procedure. Therefore, the laser KU100 DTFs are chosen as the reference and used as the template to compute the expected QE and PE with different groups of DTFs. The data in these groups is chosen to represent individual accurate DTFs, i.e. the acoustics DTFs contained in

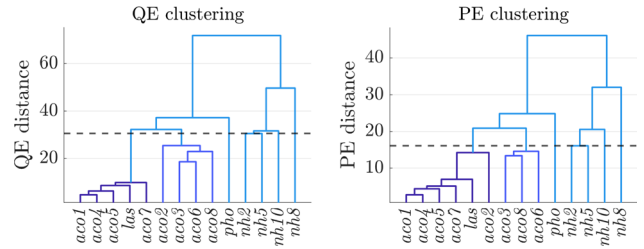
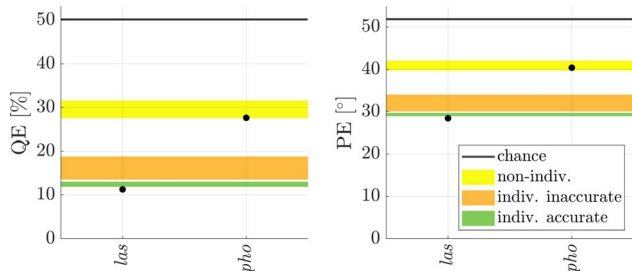


Figure 9. Dendrograms of the hierarchical clustering of the sagittal plane localisation error with different acoustics (*aco*), numerical laser (*las*) and photogrammetric (*pho*) KU100 DTFs, and ARI human subjects DTFs (*nh*).

the same cluster as the laser data for both QE and PE, individual inaccurate DTFs, i.e. the remaining KU100 measurements, non-individual DTFs, i.e. the DTFs of the ARI subjects included in the analysis, and the photogrammetric DTFs. A summary of the data contained in these groups is reported in Table 4. The outcome of the analysis is displayed in Figure 10. QE and PE of the laser DTFs in baseline condition have a low value in line with experimental results [3]. The expected errors using the DTFs in the various groups as targets are presented as envelopes representing the 25th and 75th percentile of the error distribution in each group. The individual accurate DTFs show small variations and median close to baseline conditions. The maximum deviation in QE and PE between baseline and errors obtained in group 1 is 1.8% and 1.8°, respectively. This is found to be lower than QE and PE differences from individual data used to select non-individual minimal deviant DTFs in [7]. The perceptual VR experiment carried out in that study shows that DTFs characterised by this level of deviation have low differences in terms of localisability and realism compared to individual filters. Group 2 shows higher median and variation than the previous results, with maximum QE and PE deviations reaching 9.1% and 6.3° from the baseline. The difference between these two groups is likely due to measurement noise and repeatability issues capable of causing differences in localisation performance. Large errors are observed when using the non-individual data contained in group 3, with QE and PE reaching differences as high as 20.5% and 15.0° from the baseline, and approaching chance rate corresponding to random guessing. This level of deviation is reported to significantly hinder localisability, realism and externalisation of a rendered VR scene [7]. The photogrammetric DTFs, showing QE and PE deviations from the baseline of 16.4% and 12.0°, relate to a localisation performance similar to non-individual data, even though they are computed on the same underlying dummy head geometry. This is attributed to the geometrical difference between photogrammetric and laser meshes reported in Section 2.3, especially at the ear cavities, which are critical for generating the spectral HRTF characteristics used in elevation perception [22]. High increases in QE and PE up to 6% and 12° are observed in [15] between DTFs computed on reference and photogrammetric scans of a different dummy head. While the increase in PE is found to be in line with the

Table 4. Groups of DTFs defined in relation to their similarity to the numerical laser KU100 DTFs.

Group ID	Description	ID
1	indiv. accurate	{ <i>aco1</i> , <i>aco4</i> , <i>aco5</i> , <i>aco7</i> }
2	indiv. inaccurate	{ <i>aco2</i> , <i>aco3</i> , <i>aco6</i> , <i>aco8</i> }
3	non-indiv.	{ <i>nh2</i> , <i>nh5</i> , <i>nh8</i> , <i>nh10</i> }
4	photogrammetric	{ <i>pho</i> }

**Figure 10.** Sagittal plane localisation error with template laser KU100 DTFs for different target DTFs. The dots indicate baseline error of the laser DTFs (*las*) and error obtained with target photogrammetric DTFs (*pho*). The envelopes represent the 25th and 75th percentile of the errors obtained with the DTFs in the groups of Table 4 as targets. The black line corresponds to chance rate, i.e. random guessing.

current outcome, the QE metric shows a smaller increase in that study, which could be attributed to the use of a different dummy head, potentially relating to a different baseline QE.

The MSD and ISSD are evaluated with equations (7) and (8) between the laser and the other DTFs presented in Table 4. The MSD between the laser data and the DTFs in group 1 shows a median value of 2.3 dB, while values of 4.0 dB, 5.1 dB and 4.4 dB are observed for group 2, 3 and 4, respectively. Median ISSD values for group 1–4 are 0.9 dB², 3.3 dB², 9.8 dB² and 4.5 dB², respectively. Correlation analysis is carried out between the resulting MSD, ISSD and the sagittal localisation metrics. The outcome shows a high Pearson’s correlation coefficient of around 0.82 between MSD and QE and 0.84 for MSD and PE. Lower correlation values of around 0.75 are reported in [25]; this discrepancy might arise due to the limited amount of common ϕ selected in the current analysis. Higher correlation values of 0.90 and 0.91 are seen between ISSD and QE, and ISSD and PE, respectively, which might relate to the use of gammatone filtering in the ISSD metric, which is also applied in the sagittal localisation model but not in the MSD calculation. MSD and ISSD show a correlation coefficient of 0.88, while a value of 0.98 is observed between QE and PE.

The same metrics employed in the previous analyses are used to assess the effect of misalignment in the simulated results. The outcome is reported in Table 5, where the HRTFs computed on the laser mesh to which rotation in each axis is applied (*las* misaligned) [46] are compared to the HRTFs obtained on the reference laser mesh alignment. Results of the comparison between the photogrammetric

HRTFs computed on the manual alignment (*pho* misaligned) [46] and the reference alignment, obtained using the ICP algorithm to minimise the distance from this mesh to the reference laser mesh, are also displayed. Although the deviation is generally low in all metrics, higher values are seen in those relating to the horizontal plane, i.e. ITD, ILD and ρ_H . The maximum ITD and ILD deviations appear for lateral directions, with both values being smaller than those observed between repeated acoustic measurements. However, these deviations are found to generally be above the anechoic JND; therefore, they are assumed to cause perceivable differences. While ρ_M shows deviation below 1% from the reference, ρ_H reaches values above 5%. However, this value appears at the contralateral side, where the symmetry coefficient is generally low, and the ρ_H of repeated acoustic measurements reaches a deviation up to 15.8%. The values of PE, QE, MSD and ISSD, being considerably smaller than the deviation in these metrics between baseline laser DTFs and data contained in the individual accurate group, seem to indicate that the effect of misalignment has a minimal influence on sagittal plane localisation performance. Therefore, given the relation of QE and PE to sound perception in VR scenes [7], it is assumed that potential misalignment errors in the computed numerical HRTFs have a negligible impact on elevation perception.

4 Analysis of pinna geometry influence

4.1 Mesh modification and HRTF assessment

Although the laser scan only approximates the dummy head geometry, the results of Section 3.7 show that the HRTFs computed on it are similar in spectral metrics and modelled elevation perception to several acoustic KU100 HRTFs. Therefore, this mesh and the related HRTFs are used as a reference to further analyse the photogrammetric data. The poor localisation performance obtained with the photogrammetric results is attributed to the geometrical difference between this and the laser scan. However, it is unclear at which parts of the pinna this inaccuracy has the strongest effect. Thus, an assessment is done by modifying the laser mesh, including errors at individual ear structures, and analysing the HRTFs computed on the modified geometries. Parts of the laser (*las*) right ear mesh are removed and substituted with geometries extracted from the photogrammetric scan (*pho*) with optimal alignment to the laser mesh. These photogrammetric mesh parts are merged to the laser geometry and remeshing to a uniform edge length of 0.6 mm is applied. The modified meshes are created starting from the laser mesh, swapping the concha (*l-con*), cymba (*l-cym*), fossa (*l-fos*) and scapha (*l-sca*), since these ear structures present the largest differences and are indicated as acoustically relevant anatomical parts for sound localisation [16]. Moreover, since the largest geometrical difference is observed at the cymba, an additional mesh is created merging the laser right ear cymba to the photogrammetric mesh (*p-cym*), to test if a different cymba geometry in the photogrammetric mesh relates to numerical HRTFs closer to the laser ones. The resulting meshes with swapped

Table 5. Deviation from reference of objective and perceptual metrics of KU100 HRTFs, computed on six misaligned laser meshes (*las* misaligned) and one misaligned photogrammetric mesh (*pho* misaligned). The reference HRTFs are computed on the optimally aligned laser (for *las* misaligned) and photogrammetry (for *pho* misaligned) scanned KU100 meshes.

ID	Max. deviation from reference							
	ITD [μs]	ILD [dB]	ρ_M [%]	ρ_H [%]	QE [%]	PE [$^\circ$]	MSD [dB]	ISSD [dB^2]
<i>las</i> misaligned	25.0	4.2	0.6	5.8	0.8	0.5	1.1	0.2
<i>pho</i> misaligned	10.4	2.3	0.9	2.0	0.6	0.8	1.3	0.2

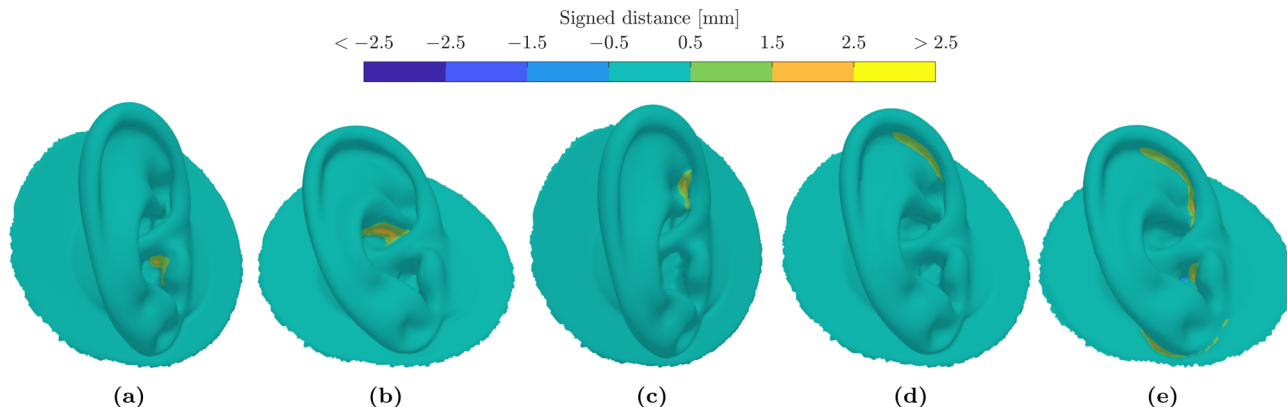


Figure 11. KU100 right ear meshes with swapped anatomical features between laser and photogrammetric meshes. Laser mesh with swapped cavum conchae (a, *l-con*), cymba conchae (b, *l-cym*) fossa triangularis (c, *l-fos*) and scapha (d, *l-sca*), taken from the photogrammetric mesh. Photogrammetric mesh with swapped cymba conchae (e, *p-cym*), taken from the laser mesh. The colours show the signed distance from the laser mesh, cropped at ± 2.5 mm.

pinna structures are presented in Figure 11, alongside the signed distance to the laser geometry.

The same processing steps as in Section 2.1 are applied to the modified geometries, obtaining right ear graded meshes on which the HRTFs are evaluated [46]. Frequency scaling using the average α obtained in Section 3.7 between laser and acoustic data is applied, and the HRTFs are low-pass filtered and converted to DTFs and CTFs. Given that the photogrammetric error shows similar trends but different details between the two ears, the right ear HRTFs are mirrored to the left to focus on the effect of the inaccuracy rather than its asymmetry. Potential deviation between HRTFs arising from this and the remeshing procedure is further analysed in Section 4.2. The metrics presented in Section 2 and 3 are evaluated on the modified geometries and related HRTFs.

4.2 Results

The geometrical metrics computed on the front of the right ear, using the laser mesh as the reference, are presented in Table 6. The results of the photogrammetric mesh are also included for comparison. Between the swapped meshes, *p-cym* shows the worst results in the majority of the metrics, which is due to the fact that the base geometry of this ear is the photogrammetric scan, and only the cymba is modified to reduce its distance from the laser mesh. Between the meshes in which the laser geometry is used as the base, the worst results are generally observed for *l-cym*, which is attributed to the incompleteness and high

noise in the photogrammetric scan of this ear part. The only exception is for **Acc** and **Avg**, showing similar or worse results for *l-con*, which could be due to the fact that the concha has a bigger size than the other swapped ear structures; hence, it has a higher impact on **Avg** and **Acc**, focusing on the overall error distribution from scanned to reference points. The fact that **Max** is the same for *pho* and *l-cym* indicates that the maximum geometrical difference between *pho* and *las* is located at the cymba.

The difference between the metrics computed on the modified meshes HRTFs and on the original laser results is calculated; Table 7 presents the outcome of this analysis, also reporting the results obtained on the *pho* data for comparison. Deviations in all metrics tend to be the highest for the *pho* results. ITD deviations are below the JND for all cases, suggesting that this interaural feature, mainly related to low frequency components, is correctly captured in the simulated photogrammetric HRTFs. Conversely, the ILD deviation shows values above the JND. To better understand the cause of the deviation, this is compared to the effect of the remeshing and mirroring procedure, evaluated by recomputing the right ear HRTFs on the unmodified laser geometry, remeshed to a slightly different uniform edge length, i.e. 0.7 mm, prior to applying the mesh grading algorithm. The ITDs and ILDs evaluated on the mirrored HRTFs computed on this geometry show a deviation of 10.4 μs and 1.7 dB from the reference results. Therefore, the ILD deviation of *pho*, *l-con* and *l-cym* is considered to be due to the inaccurate geometry, given its value higher than the one stemming from remeshing and mirroring.

Table 6. Geometrical error metrics of KU100 right ear meshes with modified ear structures, relating to the geometries shown in Figure 11. Analysis limited to the ear front. The KU100 laser right ear mesh is used as reference.

ID	Acc [mm]	Cmp [%]	Avg [mm]	Max [mm]	CD [mm ²]
<i>pho</i>	0.87	90.5	0.19	2.28	0.50
<i>l-con</i>	0.26	98.8	0.05	1.82	0.07
<i>l-cym</i>	0.15	96.1	0.05	2.28	0.19
<i>l-fos</i>	0.02	98.3	0.02	1.83	0.09
<i>l-sca</i>	0.07	98.0	0.02	1.64	0.08
<i>p-cym</i>	0.47	94.5	0.13	1.83	0.28

Table 7. Deviation from reference of objective and perceptual metrics of KU100 HRTFs computed on meshes with modified ear structures, relating to the geometries shown in Figure 11. The reference HRTFs are computed on the laser scanned KU100 mesh.

ID	Max. deviation from reference					
	ITD [μ s]	ILD [dB]	QE [%]	PE [$^{\circ}$]	MSD [dB]	ISSD [dB ²]
<i>pho</i>	6.3	3.7	16.3	12.0	4.4	4.5
<i>l-con</i>	8.3	2.5	2.0	1.7	2.5	1.5
<i>l-cym</i>	10.4	3.1	13.1	9.1	3.9	3.2
<i>l-fos</i>	10.4	1.5	4.9	3.2	2.7	1.6
<i>l-sca</i>	10.4	1.7	2.1	2.0	2.3	1.5
<i>p-cym</i>	8.3	1.6	6.9	3.7	3.0	1.9

The proximity of the ILD deviations of the other HRTFs to this value makes it unclear if the effect is due to the inaccurate geometry or the remeshing and mirroring procedure. The QE and PE metrics, evaluated using the laser DTFs as template and the modified meshes DTFs as targets, show the maximum deviations for *pho*. Within the modified laser meshes results, the highest deviations are seen for *l-cym*, followed by *l-fos*. Inaccuracies at the concha and scapha seem to have a low influence; their localisation error deviation is close to the one due to remeshing and mirroring, relating to QE and PE values of 2.1% and 1.8 $^{\circ}$, respectively. The results of *p-cym* show that the cymba geometry has a large impact on the modelled localisation performance, since a reduction in QE and PE deviations from the baseline condition by around 9% and 8 $^{\circ}$ from *pho* to *p-cym* are observed. The MSD and ISSD show results similar to the localisation error analysis. The highest values are obtained between *pho* and the reference data. A great influence of the error at the cymba is also observed, while all the other modifications of the reference mesh yield similar results, found to be close to the value of 2.0 dB in MSD and 1.5 dB² in ISSD due to remeshing and mirroring. Furthermore, decreases of around 1.4 dB and 2.6 dB² from *pho* to *p-cym* are observed in MSD and ISSD, respectively.

The influence of the geometrical error at the various pinna structures on the localisation performance is also displayed in Figure 12. In this plot, the localisation error obtained with the modified meshes DTFs used as targets is compared to the localisation performance in baseline conditions of the laser DTFs, and with the expected errors evaluated on the groups defined in Table 4. The results summarise the previous findings, showing that localisation errors with the photogrammetric concha and scapha geometries tend to be slightly higher than baseline values, with a deviation similar to that of individual accurate

DTFs. Given the low deviation, it is unclear if this is due to scanning errors or to the effect of remeshing and mirroring. Differences in fossa geometry tend to have a stronger influence, showing a higher localisation error. Modifications of the cymba lead to QE and PE tending to values relating to modelled localisation performance with non-individual filters. The *p-cym* results are seen to be in line with the localisation error of individual inaccurate DTFs.

Two studies on parametric pinna models have identified several parameters as the most influential for the HRTFs. Specifically, in [24] d_3 and d_4 (concha width and fossa height) are indicated as the most influential parameters, while d_2 (cymba height) is reported to have minor importance. In comparison to the current study, the swapped area in *l-cym* relates to d_2 and d_3 , while *l-fos* relates to d_4 . The outcome of the current study agrees with the significance of d_3 since, although relating to the concha, this parameter can modify the geometry of the upper wall of the cymba, in proximity of the area where the geometrical error in *l-cym* appears. The fact that errors in *l-fos* are found to impact some of the perceptual metrics also matches with the identification of d_4 as a relevant parameter. However, the low influence of d_2 , directly linked to the cymba dimensions, seems to disagree with the current results. The most influential control points found in [25] are, in order of importance, CPd1, CPd4, CP3 and CP13 (root of helix relief, antihelix relief, crus of antihelix and upper antihelix). CPd1 and CPd4 relate to the depth of concha (along with cymba) and fossa, respectively, while CP3 and CP13 relate to the upper wall of the cymba. These results are also considered to be in agreement with the current outcome since the reported control points relate to locations where errors in *l-cym* and *l-fos* appear. However, CP2 (crus of helix), which is related to d_2 and *l-cym*, is reported to have a low influence. The disagreement between the previous and current results

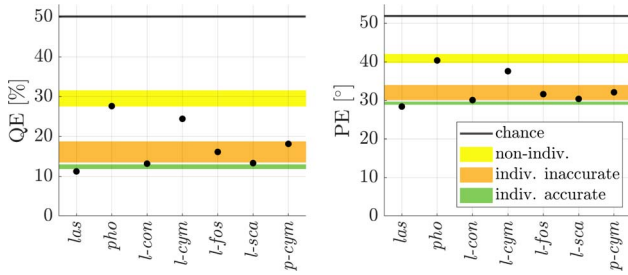


Figure 12. Sagittal plane localisation error with template laser KU100 DTFs for different target DTFs. The dots indicate baseline error of the laser DTFs (*las*), error obtained with photogrammetric DTFs (*pho*) and modified meshes DTFs as target, relating to the geometries shown in Figure 11. The envelopes represent the 25th and 75th percentile of the errors obtained with the DTFs in the groups of Table 4 as targets. The black line corresponds to chance rate, i.e. random guessing.

on d_2 and CP2 could be related to the high modifications of the cymba due to differences between the photogrammetric and lasers scans, reaching maximum values above 2 mm and affecting the entire surface of the cymba, while d_2 and CP2 are modified by around 2 mm from their mean values, i.e. around 6 mm. Conversely, CPd1 is modified by maximum 6 mm from its mean value, which could relate to the high sensitivity indicated for this control point [25], especially when compared to the maximal deviation seen in *l-fos* in the current study, below 2 mm.

To better assess the effect of modifying the cymba geometry, DTFs at a representative elevation angle on the median plane, and CTFs obtained on *l-cym* and *p-cym*, are plotted against the *las* and *pho* results in Figure 13. The acoustic DTFs and CTFs distributions are also displayed for comparison. Similarity is noticed between the *las* and *p-cym* DTFs, with their spectral features showing relatively low discrepancies in centre frequency and magnitude. A similar observation can be made for *pho* and *l-cym*. Furthermore, this trend is also noticed in the CTFs, although above 12 kHz the divergence tends to increase. It is worth mentioning that the data within these two pairs originates from different base meshes, i.e. laser and photogrammetric scans, but share a common cymba geometry. By tracking the N1 centre frequency of *l-cym* and *p-cym* on the median plane, and comparing it to that of *las*, an average shift towards higher frequencies of 1.3 kHz is observed for the first, while the latter shows a downward shift of 0.1 kHz. This effect is attributed to the large geometrical differences between photogrammetric and laser meshes, mainly located at the cymba and fossa. Their important role in the creation of the HRTFs spectral features, mainly N1, P2 and P3, results in a high effect of the error at these concave ear structures. The match in the cymba geometry of *p-cym* and *las* relates to a better agreement in their spectral features, most notably in N1, leading to a decrease in modelled perceptual deviation between them. In relation to the acoustic DTFs, the *las* results tend to show the best match, while larger discrepancies are observed for the numerical DTFs of the modified meshes. The acoustic CTFs show high

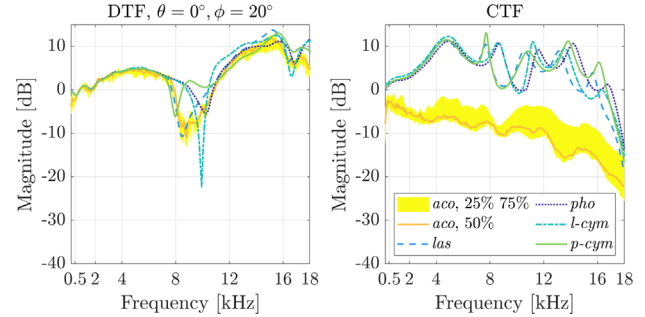


Figure 13. Magnitude of acoustic (*aco*), numerical laser (*las*), photogrammetric (*pho*), laser with swapped cymba conchae (*l-cym*) and photogrammetric with swapped cymba conchae (*p-cym*) KU100 right ear DTFs for frontal source, and CTFs. The *aco* results are shown as an envelope representing the 25th and 75th percentile of the measured DTFs and CTFs magnitude, the 50th percentile is also displayed. The modified ear DTFs and CTFs relate to the geometries shown in Figure 11.

divergence from the numerical ones; potential reasons for this are discussed in Section 3.7.

These analyses indicates that, on the tested ear meshes, the effect of geometrical errors at the cymba tend to have the greatest impact on the computed HRTFs. This could be related to the large photogrammetric scanning error in relation to the small dimensions of this ear structure. Nonetheless, by including the laser cymba geometry in the photogrammetric scan, the modelled perceptual deviation between the HRTFs computed on this mesh and the laser results tends to decrease to levels observed with individual inaccurate data. Therefore, for the acquisition of perceptually accurate DTFs on scanned geometries, high precision is needed at the ear cavities, especially at the cymba, which tends to be the location where the scan is the most challenging due to occlusion.

Correlation analysis is carried out between the geometrical metrics evaluated on the meshes with swapped ear structures, and the objective and perceptual metrics evaluated on their DTFs, specifically QE, PE, MSD and ISSD. This is done in order to identify the most relevant geometrical metrics in terms of similarity of numerical HRTFs computed on the evaluated meshes to reference data. All the geometrical metrics show high correlation coefficients above 0.92 between them, with the exception of **Max**, relating to coefficients below 0.67 with the others. Most notably, high correlation is observed between **Acc** and **Avg**, showing a value of 0.98, and **Cmp** and **CD**, with a value of -0.99 and the negative coefficient deriving from **Cmp**, for which higher values relate to better results. **CD** shows the highest correlation with the HRTFs assessment metrics, with coefficients ranging from 0.85 to 0.88, while comparable but negative values ranging from -0.84 to -0.88 are seen for **Cmp**. **Acc** and **Avg** show slightly lower correlation coefficients with the acoustic metrics, spanning the range from 0.70 to 0.77, while **Max** shows values ranging from 0.64 to 0.81. This analysis identifies **CD** and **Cmp** as the best predictors of deviation between HRTFs computed on target and reference scans. However, these

results are limited by the low number of seven observations employed in this analysis, deriving from the photogrammetric mesh, the five swapped meshes and the remeshed and mirrored laser mesh HRTFs.

5 Discussion

The results of the geometrical deviation analysis presented in Section 2.3 indicate that, with the employed photogrammetric scanning procedure, the best results are obtained on the optically treated dummy head. The largest difference between photogrammetric and laser scans is noticed at the concave pinna structures, where noise and holes are present due to the complex pinna geometry generating occlusion. The outcome is found to generally be in line or better than the results reported for photogrammetric scans of other dummy heads [15, 19] or human subjects [14], also adopting optical treatment of the scanned surfaces. Discrepancies between the current and previous results could stem from differences in the acquisition and processing of the geometry, e.g. the number of pictures used for the reconstruction, or the post-processing steps used to convert point clouds in meshes. While in [18] it is reported that the photogrammetric scanning outcome is statistically independent from the tested number of pictures used for the reconstruction, i.e. 30, 60 and 90, the metrics used in that study are targeted to outer ear scanning for prostheses fabrication, and tend to be less strict than those related to scans for HRTF computation [14]. A similar analysis using geometrical metrics more relevant to HRTFs could be used to select the optimal number of pictures and sampling directions for the scans to possibly maximise the accuracy of the outcome or minimise the processing time, given that using fewer pictures relates to faster reconstructions [18].

The results of Section 3.7 indicate that the experimentally measured HRTFs present high variability in interaural features, also between repeated measurements on the same dummy head, reaching values above the anechoic JND for both ITDs and ILDs, as reported in previous studies [27, 28]. The numerical laser and photogrammetric ITDs and ILDs are close to each other and show small deviations from the median results of the measured data, although also above the JND; hence, likely perceivable in an anechoic case. This discrepancy could stem from differences between the simulated and real scenario, e.g. a different alignment of the dummy head, or the choice of using rigid boundary conditions in the computations, which ignores potential damping effects of the dummy head material [42]. It should also be mentioned that the ILD results are averaged on a linear frequency scale; hence, emphasising high frequency spectral components, at which deviation is more likely to occur between measurements and simulations, given the higher effect of mismatch in geometry or other simulation parameters due to the small wavelength. These mismatches could also have an influence on the symmetry coefficients of the numerical HRTFs, which are seen to be generally lower than those related to the measured data.

The perceptually inspired analysis carried out on the full HRTFs dataset seems to indicate that the modelled localisation error with the numerical laser DTFs is close to the one obtained with some of the measured results, while the photogrammetric DTFs lead to a modelled localisation error only slightly lower than the one obtained with non-individual data. This outcome is also noticeable in the MSD and ISSD metrics, given the high correlation between them and the QE and PE results. The hindering effect in modelled localisation performance with the photogrammetric DTFs is attributed to the differences between photogrammetric and laser geometries, mainly located at the pinna cavities and generating discrepancies in HRTFs spectral features above 8 kHz, e.g. in the centre frequency of N1. It should however be stressed that the current outcome is based on the clustering and threshold value applied on the sagittal localisation model results. This model is only an approximation of the complex perceptual factors related to spatial hearing [3]. Although its outcome is reported to transfer well to sound perception in VR scenes [7], additional analyses entailing perceptual experiments on human subjects are needed to validate the assumptions and results of the current study.

To focus on the effect of the photogrammetric error, the scan is optimally aligned and scaled to reduce its deviation from the laser mesh, similarly to previous studies focusing on photogrammetric head and ears scanning [14, 15, 19]. Although this can lead to improved results compared to a manual alignment and scaling approach, in the current study the misalignment and scaling errors observed in repeated mesh alignments seem to have a negligible effect on sagittal plane localisation. However, they relate to ITD and ILD deviations above the anechoic JND; hence, potentially perceivable. Therefore, as discussed in [14], scaling and alignment of the scanned geometry should be correctly addressed given that they can bias the results. However, this is not trivial in an application scenario where no reference mesh is available. Methods to optimise the scaling and rotation of HRTFs to match reference data have been developed [13]. A potential alignment approach, applicable when no reference is available, is to compute ITDs and ILDs, and align their minimal values to the median plane assuming symmetry of the HRTFs. Nevertheless, in the case of asymmetric behaviour, this procedure cannot correct the misalignment [8].

The study of the scanning error influence at various ear cavities, presented in Section 4.2, shows that the geometrical deviation at the cymba is the most significant for the analysed photogrammetric HRTFs. These results generally agree with the outcome of previous studies indicating that parameters relating to the concha, especially those in the proximity of the cymba and fossa, are the most significant [24, 25]. Some discrepancies are also found, i.e. in parameters directly relating to the cymba height. This could be attributed to the different ranges in which morphological features are modified, based on the inter-subject variability observed on a population in the parametric pinna models [24, 25], and on the estimated photogrammetric error in the current study. As particular ear shapes might emphasise the

influence of different pinna structures [24], additional analyses on different dummy heads and real pinnae would be of added value, but are out of scope of the current manuscript. It should be mentioned that the reference geometry, derived from the laser scanned data, is considered to be the best approximation of the dummy head shape available in the current study. Although a good match is seen between the numerical laser HRTFs and several acoustic ones in different metrics, the use of a more precise reference, e.g. obtained through an accurate volumetric scan, should be preferred since it would increase the reliability of the analyses.

While several metrics have been defined to compare the outcome of a scan to a reference geometry, their correlation with objective and perceptually inspired HRTFs metrics shows variable results. The highest positive and negative correlation coefficients are seen for **CD** and **Cmp**, respectively. Notably, these are the only two geometrical metrics in which the distance from reference to target scan is considered, while the other metrics focus only on the distance from target to reference. Nonetheless, it can be seen that although the geometrical metrics computed on *p-cym* show worse results compared to *l-cym*, the opposite is observed in the HRTFs metrics related to these two geometries. This is attributed to the fact that some locations on the ear are more sensitive to error, and the geometrical metrics do not take this into account. Applying a weighting function dependent on the location on the pinna in which the error occurs and its significance on the HRTFs could improve the correlation between geometrical and acoustic metrics.

The deviation between laser and photogrammetric geometries is attributed to the scanning error observed in the raw point clouds obtained with the latter method, requiring extensive processing for the creation of meshes suitable for numerical computation. Applying photogrammetry on a real subject also entails additional sources of error which can further decrease the scanning accuracy. Results obtained in [14] show average values of 80% **Cmp** on the scan of real subjects with ears treated with black water colour, and lower values reaching maximum 70% in an untreated scenario. Although no HRTFs have been computed on these scans, based on the outcome of the current analysis, it is assumed that these meshes would relate to HRTFs showing deviation in high frequency spectral features from the reference, which could hinder the elevation perception. In [19], photogrammetry is used to scan a real subject, and this geometry is used as the reference and to 3D print a plastic head replica which is rescanned to create a target mesh for evaluation. The analyses conducted on this scan and the related HRTFs show a typical error of 1 mm and similar patterns of horizontal plane HRTFs with good agreement below 7 kHz. Further inspection reveals geometric differences exceeding 3 mm on parts of the concha and cymba, and deviation exceeding 10 dB in HRTFs magnitude for frontal directions. Moreover, the use of photogrammetry for the reference scan in that study is questionable, given the low accuracy of this method. Similarly to the current study, the reported error at the

pinna cavities is assumed to lead to discrepancies in median plane HRTFs spectral features which might hinder the elevation perception when using the computed HRTFs for localisation tasks. Therefore, numerical computations of HRTFs on head and ear meshes directly originating from photogrammetric scans are not thought to relate to perceptually accurate individual results, even in the tested optimal case of optically treated dummy head ear scans. Additional post-processing or alternative techniques to replace or complement photogrammetry are required.

A scanned geometry close to the reference can lead to a good match between the HRTFs; this could be achieved for example by denoising photogrammetric scans using algorithms targeted to image-based 3D reconstruction [30] or deep-learning approaches trained for point cloud denoising [40]. Initial attempts in this direction have been recently made by the authors, showing that deep-learning methods used for denoising of photogrammetric pinna scans could partially reduce the geometrical difference between these and reference scans, relating to an improved similarity of the computed HRTFs to reference data [54]. However, the improvement is only marginal, likely due to the deviation of the photogrammetric error from random noise on which the algorithm is originally trained [40]. Thus, additional analyses are needed to assess if the outcome relates to perceptually accurate HRTFs. Alternative techniques could also be employed for the 3D scanning, e.g. affordable depth sensors, as tested in [15]. Although depth sensor scans show a better outcome than the photogrammetric results presented in that study, the maximum difference from the reference reaches values of 2.5 mm at the antihelix, in the proximity of the cymba; therefore, the outcome is thought to relate to similar problems as those faced in the current analysis. Recent studies have proposed to morph a high accuracy parametric pinna mesh to a target ear geometry [16]. The employed target mesh is a high accuracy scan, which is not available in an application scenario. Using a photogrammetric scan as the target could result in an improvement over a direct meshing of the noisy scan. Nevertheless, the observed photogrammetric error might hinder this approach. Morphing a parametric pinna model to an individual pinna geometry by using a limited set of measured morphological parameters could be a promising approach to create reliable individual pinna geometries; however, additional studies are needed to determine if this technique is capable of producing meshes relating to perceptually accurate individual HRTFs [25].

6 Conclusion

A KU100 dummy head mesh, obtained through photogrammetry on the dummy head treated with scanning spray, is compared to a laser scan. The geometrical differences between them, mainly located at the pinna cavities, reach maximum values around 3 mm.

The BEM is used to numerically compute HRTFs on the scanned meshes. The photogrammetric HRTFs are

compared to reference acoustic data, extracted from a database of KU100 measurements, and numerical data obtained on the laser scan; both objective and perceptually inspired metrics are used for the assessment. Given the variability observed in the measured data, making it difficult to define a valid reference, the photogrammetric results are compared to the entire dataset of measured and simulated HRTFs.

Interaural differences computed on numerical laser and photogrammetric HRTFs, although showing deviation above the anechoic JND, present less variability than in repeated acoustic measurements. The symmetry coefficients of the simulated HRTFs tend to be generally lower than those computed on the measured data, except on the contralateral side. An analysis employing a sagittal plane elevation perception model shows that the laser and several measured HRTFs relate to a similar localisation performance, while the photogrammetric results show a localisation error comparable to non-individual data, extracted from a database. The MSD and ISSD computed between the HRTFs present high correlation to the outcome of the perceptual model. These analyses validate the results computed on the laser scan and expose the limitations related to the photogrammetric acquisition of ear shapes for individual HRTFs computation, which are thought to stem from the differences between this and the laser scanned ear geometry, capable of significantly altering the HRTFs spectra above 8 kHz. Analyses on HRTFs computed on misaligned meshes in ranges extracted from repeated alignment attempts indicate that, within the tested range, discrepancies slightly above the JND can be seen in horizontal plane interaural features, while negligible effects are seen on metrics related to elevation perception. Therefore, the discrepancies observed between laser and photogrammetric HRTFs are attributed to geometrical differences between the scanned results rather than misalignment.

By including the photogrammetric geometry of different pinna cavities in the laser mesh, the influence of the scanning error at individual anatomical ear structures on the simulated HRTFs is assessed. The results show that inaccuracies at the cymba conchae relate to the highest deviation in the majority of the analysed metrics. Furthermore, including the laser scanned cymba concha in the photogrammetric mesh, the deviation of these HRTFs to the original laser data is decreased, especially in terms of metrics relating to the modelled elevation perception.

Correlation analysis between the geometrical and acoustic related metrics identifies Chamfer Distance and completeness as the most relevant metric for evaluating a scan in terms of similarity of HRTFs to those computed on a reference mesh.

The results emphasise the importance of an accurate pinna geometry for the correct acquisition of numerical HRTFs. However, the analyses focus on a particular pinna geometry and should be further extended to different artificial and human ears for generality purposes. Furthermore, perceptual experiments with real subjects are needed to assess the validity of the findings, which are currently based on a model that only approximates the perceptual factors related to sound localisation.

Conflict of interest

The authors declare no conflict of interest.

Data availability statement

The research data associated with this article are available in KU Leuven Research Data Repository (RDR), under the reference [46]. Scanning and mesh data are available on request from the authors.

Acknowledgments

The European Commission is gratefully acknowledged for their support of the VRACE research project (GA 812719). The research of S. van Ophem (fellowship no. 1277021N) is funded by a grant from the Research Foundation – Flanders (FWO). Internal Funds KU Leuven are gratefully acknowledged for their support. We would like to thank the KU Leuven Manufacturing Metrology Group for the assistance in carrying out the laser scanning.

References

1. J. Blauert: Spatial hearing: the psychophysics of human sound localization, revised edn., The MIT Press. 1996.
2. M. Burge, W. Burger: Ear biometrics in computer vision, in: Proceedings 15th International Conference on Pattern Recognition. IEEE, 2000, pp. 822–826.
3. R. Baumgartner, P. Majdak, B. Laback: Modeling sound-source localization in sagittal planes for human listeners. *Journal of the Acoustical Society of America* 136 (2014) 791–802.
4. X.-L. Zhong, B.-S. Xie: Head-related transfer functions and virtual auditory display, in: H. Glotin (Ed.), Chapter 6: Soundscape Semiotics – Localisation and Categorisation, Vol. 1, IntechOpen, 2014, pp. 99–134.
5. F. Asano, Y. Suzuki, T. Sone: Role of spectral cues in median plane localization. *Journal of the Acoustical Society of America* 88, 1 (1990) 159–168.
6. E.M. Wenzel, M. Arruda, D.J. Kistler, F.L. Wightman: Localization using nonindividualized head-related transfer functions. *Journal of the Acoustical Society of America* 94, 1 (1993) 111–123.
7. C. Jenny, C. Reuter: Usability of individualized head-related transfer functions in virtual reality: empirical study with perceptual attributes in sagittal plane sound localization. *JMIR Serious Games* 8, 3 (2020) 1–15.
8. S. Li, J. Peissig: Measurement of head-related transfer functions: a review. *Applied Sciences* 10, 14 (2020) 1–40.
9. R. Barumerli, M. Geronazzo, F. Avanzini: Round robin comparison of inter-laboratory HRTF measurements – assessment with an auditory model for elevation, in: IEEE 4th VR Workshop on Sonic Interactions for Virtual Environments. IEEE, 2018, pp. 1–5.
10. B.F.G. Katz: Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation. *Journal of the Acoustical Society of America* 110, 5 (2001) 2440–2448.
11. H. Ziegelwanger, P. Majdak, W. Kreuzer: Numerical calculation of listener-specific head-related transfer functions and sound localization: microphone model and mesh discretization.

- Journal of the Acoustical Society of America 138, 1 (2015) 208–222.
12. F.R. Ospina, M. Emerit, B.F. Katz, The threedimensional morphological database for spatial hearing research of the BiLi project, in: Proceedings of Meetings on Acoustics. Acoustical Society of America, 2015, pp. 1–17.
 13. C.T. Jin, P. Guillon, N. Epain, R. Zolfaghari, A. Van Schaik, A.I. Tew, C. Hetherington, J. Thorpe: Creating the Sydney York morphological and acoustic recordings of ears database. IEEE Transactions on Multimedia 16, 1 (2014) 37–46.
 14. A. Reichinger, P. Majdak, R. Sablatnig, S. Maierhofer, Evaluation of methods for optical 3-D scanning of human pinnas, in: Proceedings – 2013 International Conference on 3D Vision. IEEE, 2013, pp. 390–397.
 15. M. Dinakaran, F. Brinkmann, S. Harder, R. Pelzer, P. Grosche, R.R. Paulsen, S. Weinzierl: Perceptually motivated analysis of numerically simulated head-related transfer functions generated by various 3D surface scanning systems, in: ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing – Proceedings, IEEE, 2018, pp. 551–555.
 16. K. Pollack, W. Kreuzer, P. Majdak, Perspective chapter: modern acquisition of personalised head-related transfer functions – an overview, in: Advances in Fundamental and Applied Research on Spatial Audio, IntechOpen, 2022.
 17. Y. Kahana, P.A. Nelson: Numerical modelling of the spatial acoustic response of the human pinna. Journal of Sound and Vibration 292, 1–2 (2006) 148–178.
 18. M.T. Ross, R. Cruz, T.L. Brooks-Richards, L.M. Hafner, S. K. Powell, M.A. Woodruff: Comparison of three-dimensional surface scanning techniques for capturing the external ear. Virtual and Physical Prototyping 13, 4 (2018) 255–265.
 19. A. Mäkivirta, M. Malinen, J. Johansson, V. Saari, A. Karjalainen, P. Vosough, Accuracy of photogrammetric extraction of the head and torso shape for personal acoustic HRTF modeling, in: 148th Audio Engineering Society International Convention. Audio Engineering Society, 2020, pp. 1–8.
 20. V.R. Algazi, R.O. Duda, D.M. Thompson, C. Avendano: The CIPIC HRTF database, in: Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics. IEEE, 2001, pp. 99–102.
 21. W.W. Hugeng, D. Gunawan: Improved method for individualization of head-related transfer functions on horizontal plane using reduced number of anthropometric measurements. Journal of Telecommunications 2, 2 (2010) 31–41.
 22. H. Takemoto, P. Mokhtari, H. Kato, R. Nishimura, K. Iida: Mechanism for generating peaks and notches of head-related transfer functions in the median plane. Journal of the Acoustical Society of America 132, 6 (2012) 3832–3841.
 23. J. Fels, M. Vorländer: Anthropometric parameters influencing head-related transfer functions. Acta Acustica united with Acustica 95, 2 (2009) 331–342.
 24. S. Ghorbal, T. Auclair, C. Soladié, R. Séguier: Pinna morphological parameters influencing HRTF sets, in: Proceedings of the 20th International Conference on Digital Audio Effects (DAFx-17). DAFx-17, 2017, pp. 353–359.
 25. P. Stitt, B.F.G. Katz: Sensitivity analysis of pinna morphology on head-related transfer functions simulated via a parametric pinna model. Journal of the Acoustical Society of America 149, 4 (2021) 2559–2572.
 26. M. Vorländer: Past, present and future of dummy heads, in: Acústica. S. Hirzel Verlag, 2004, pp. 1–6.
 27. A. Andreopoulou, D.R. Begault, B.F. Katz: Inter-laboratory round robin HRTF measurement comparison. IEEE Journal on Selected Topics in Signal Processing 9, 5 (2015) 895–906.
 28. R. Greff, B.F.G. Katz: Round robin comparison of HRTF simulation results: preliminary results, in: 123rd AES Convention. Audio Engineering Society, 2007, pp. 1–5.
 29. P. Mokhtari, H. Takemoto, R. Nishimura, H. Kato: Computer simulation of KEMAR’s head-related transfer functions: verification with measurements and acoustic effects of modifying head shape and pinna concavity, in: Principles and Applications of Spatial Hearing, World Scientific Publishing Company, 2011, pp. 205–215.
 30. K. Wolff, C. Kim, H. Zimmer, C. Schroers, M. Botsch, O. Sorkine-Hornung, A. Sorkine-Hornung: Point cloud noise and outlier removal for image-based 3D reconstruction, in: Proceedings – 2016 4th International Conference on 3D Vision, 3DV. IEEE, 2016, pp. 118–127.
 31. R. Struck, S. Cordoni, S. Aliotta, L. Pérez-Pachón, F. Gröning: Application of photogrammetry in biomedical science, in: Chapter 10: Biomedical Visualization, Advances in Experimental Medicine and Biology, vol. 1, Springer, 2019, pp. 121–130.
 32. J.L. Schönberger, J.-M. Frahm: Structure-from-motion revisited, in: Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016, pp. 4104–4113.
 33. J.L. Schönberger, E. Zheng, M. Pollefeys, J.M. Frahm: Pixelwise view selection for unstructured multi-view stereo, in: European Conference on Computer Vision (ECCV). Springer, 2016, pp. 1–15.
 34. P. Cignoni, M. Callieri, M. Corsini, M. Dellepiane, F. Ganovelli, G. Ranzuglia: MeshLab: an open-source mesh processing tool, in: Sixth Eurographics Italian Chapter Conference. The Eurographics Association, 2008, pp. 129–136.
 35. M. Kazhdan, H. Hoppe: Screened poisson surface reconstruction. ACM Transactions on Graphics 32, 3 (2013) 1–13.
 36. J. Möbius, L. Kobbelt: OpenFlipper: an open source geometry processing and rendering framework, in: Proceedings of the 7th international conference on Curves and Surfaces. Springer, 2010, pp. 488–500.
 37. P. Liepa: Filling holes in meshes, in: Eurographics Symposium on Geometry Processing. The Eurographics Association, pp. 200–205, 2003.
 38. H. Ziegelwanger, W. Kreuzer, P. Majdak: A priori mesh grading for the numerical calculation of the head-related transfer functions. Applied Acoustics 114 (2016) 99–110.
 39. P. Cignoni, C. Rocchini, R. Scopigno: Metro: measuring error on simplified surfaces. Computer Graphics Forum 17 (1998) 167–174.
 40. M.-J. Rakotosaona, V. La Barbera, P. Guerrero, N.J. Mitra, M. Ovsjanikov: PointCleanNet: learning to denoise and remove outliers from dense point clouds. Computer Graphics Forum 39, 1 (2019) 185–203.
 41. H. Ziegelwanger, W. Kreuzer, P. Majdak: MESH2HRTF: an open-source software package for the numerical calculation of head-related transfer functions, in: Proceedings of the 22nd International Congress on Sound and Vibration. International Institute of Acoustics and Vibration, 2015, pp. 1–8.
 42. A.B. Dobrucki, P. Plaskota: Computational modelling of head-related transfer functions. Archives of Acoustics 32, 3 (2007) 659–682.
 43. F. Brinkmann, A. Lindau, S. Weinzierl, S. Van De Par, M. Müller-Trapet, R. Opdam, M. Vorländer: A high resolution and full-spherical head-related transfer function database for different head-above-torso orientations. Journal of the Audio Engineering Society 65, 10 (2017) 841–848.
 44. J.C. Middlebrooks, D.M. Green: Directional dependence of interaural envelope delays. Journal of the Acoustical Society of America 87, 5 (1990) 2149–2162.
 45. ClubFritz HRTFs Database: <https://sofacooustics.org/data/database/clubfritz/> [Accessed on January 29, 2021].

46. Replication Data for: Analysis of Photogrammetric Scanning Error Significance on Numerical Head-Related Transfer Functions of a Dummy Head [Accessed on March 20, 2023]. <https://doi.org/10.48804/MLQ90Q>.
47. P.L. Sondergaard, P. Majdak: The auditory modeling toolbox, in: J. Blauert (Ed.), *The Technology of Binaural Listening*, Springer, 2013, pp. 33–56.
48. P. Majdak, R. Baumgartner, B. Laback: Acoustic and non-acoustic factors in modeling listenerspecific performance of sagittal-plane sound localization. *Frontiers in Psychology* 5, 319 (2014) 1–10.
49. A. Andreopoulou, B.F. Katz: Subjective HRTF evaluations for obtaining global similarity metrics of assessors and assessees. *Journal on Multimodal User Interfaces* 10, 3 (2016) 259–271.
50. ARI HRTFs Database: <https://sofacoustics.org/data/database/ari/> [Accessed on April 20, 2021].
51. B.F. Katz, R. Nicol: Binaural spatial reproduction. in: *Sensory Evaluation of Sound*, CRC Press, Taylor & Francis Group, 2018: 349–388.
52. A. Andreopoulou, B.F.G. Katz: Identification of perceptually relevant methods of inter-aural time difference estimation. *Journal of the Acoustical Society of America* 142, 2 (2017) 588–598.
53. S. Klockgether, S. van de Par: Just noticeable differences of spatial cues in echoic and anechoic acoustical environments. *Journal of the Acoustical Society of America* 140, 4 (2016) 352–357.
54. F. Di Giusto, F. Lluís Salvadó, S. van Ophem, W. Desmet, E. Deckers: Deep learning for photogrammetric ear point clouds denoising, in: *Proceedings of DAGA 2022*. German Acoustical Society, 2022, pp. 146–149.

Cite this article as: Di Giusto F. van Ophem S. Desmet W. & Deckers E. 2023. Analysis of laser scanning and photogrammetric scanning accuracy on the numerical determination of Head-Related Transfer Functions of a dummy head. *Acta Acustica*, 7, 53.