DOI: 10.61782/fa.2025.0356

FORUM ACUSTICUM
EURONOISE 2025

EVALUATING DEEP LEARNING APPROACHES FOR PHOTOGRAMMETRIC EAR SCAN DENOISING IN HEAD-RELATED TRANSFER FUNCTION COMPUTATION

Fabio Di Giusto[1,2*] **Francesc Lluis**[3] **Sjoerd van Ophem**[4] **Elke Deckers**[5,2]
[1] Department of Mechanical Engineering, KU Leuven, Belgium
[2] Flanders Make@KU Leuven, Belgium
[3] Bang & Olufsen A/S, Denmark
[4] Institute of Sound and Vibration Research, University of Southampton, United Kingdom
[5] Department of Mechanical Engineering, Campus Diepenbeek, KU Leuven, Belgium

ABSTRACT

Accurate Head-Related Transfer Functions (HRTFs) are key elements for virtual audio rendering and can be numerically computed from detailed 3D scans of a subject's head and ears. While high-precision scanning techniques yield reliable geometries, their accessibility is limited. Photogrammetry, leveraging widely available camera sensors, offers a more accessible alternative, though it introduces scanning errors affecting HRTF fidelity.

This study investigates the use of Deep Neural Networks (DNNs) to denoise photogrammetric ear scans. Various DNN architectures are tested against a classical denoising approach. To improve performance, the DNNs are modified and fine-tuned on pinna samples with synthetic errors replicating those in photogrammetric dummy head scans. Comparing HRTFs from original and denoised scans with reference data shows that the best-performing DNN reduces HRTF deviations, though improvements remain limited. Correlation analyses between geometric features and HRTF metrics, computed on scanned point clouds and corresponding HRTFs, identify key measures for assessing deviations from reference scans. These findings provide insights for refining denoising techniques and enhancing photogrammetric scans for HRTF computation.

*__Corresponding author__: *fabio.digiusto@kuleuven.be.*

Keywords: *head-related transfer functions, photogrammetry, point cloud denoising*

1. INTRODUCTION

Spatial sound perception relies on the brain's interpretation of auditory cues in signals reaching a subject's eardrums [1]. The Head-Related Transfer Functions (HRTFs) contain these binaural and monaural cues. The latter, largely influenced by pinna morphology, are crucial for elevation perception [2]. HRTF individualisation is essential for realistic binaural audio rendering. However, standard acquisition methods, i.e. measurements or simulations via numerical techniques such as the Finite Element Method (FEM) on a subject's scan, require specialised equipment, limiting large-scale application [1, 3].

This study focuses on numerical HRTF computation, leveraging efficient simulation techniques and accessible scanning methods. A key challenge of this approach is capturing the pinna geometry, requiring a resolution of $1\,\mathrm{mm}$ for accurate HRTFs [1]. While laser and structured-light scanners provide sufficient resolution, their widespread use is limited by the costly equipment. Photogrammetry, employing common camera sensors, offers high scalability. However, it tends to introduce scanning errors due to self-occlusion, particularly in acoustically relevant concave pinna regions [4]. These errors can severely impact the HRTF spectral features, compromising the elevation cues. Thus, denoising photogrammetric point clouds or using alternative acquisition methods is necessary to obtain accurate individual HRTFs [3].

Point cloud denoising has recently gained attention, especially through Deep Neural Network (DNN) techniques, often outperforming classical methods based on filtering or optimisation [5]. DNNs are typically trained on datasets of generic shapes corrupted with synthetic error, usually modelled as unstructured Gaussian noise. While effective on such noise types, their performance tends to decline when error characteristics differ significantly from those encountered during training [6].

This paper evaluates the effect of DNN-based denoising on photogrammetric scans of dummy head ears and its potential to enhance the accuracy of computed HRTFs. Two DNNs are compared against a classical denoising method using various geometric metrics. To improve denoising performance, the DNNs are trained or fine-tuned on a dataset of ear geometries corrupted by synthetic photogrammetric error replicating experimental data. Modifications are applied to one DNN to further enhance its effect. HRTFs computed on meshes derived from original and denoised scans are assessed using objective and perceptually inspired metrics. Correlations between geometric and HRTF metrics are analysed to provide insight into the most effective geometric measures in relation to HRTF accuracy. These findings can better guide the selection of relevant metrics and loss functions for developing HRTF individualisation techniques based on individual scans.

2. METHODS

2.1 Denoising methods

Let $Y = \{y_i, i = 1, \ldots, N_Y\}$ be a noisy point cloud of N_Y points y_i, often approximated as $y_i = x_i + e_i$, where e_i is an error affecting the clean point cloud $X = \{x_i, i = 1, \ldots, N_X\}$. Denoising is formulated as $\widetilde{y}_i = \mathcal{D}(y_i)$, where $\mathcal{D}$ is a function mapping Y to its denoised version $\widetilde{Y}$. $\mathcal{D}$ can represent classical denoising techniques or DNNs usually trained in a supervised manner on noisy-clean data pairs [5]. To evaluate the denoising effect, the deviation between $\widetilde{Y}$ (or Y) and X is measured using metrics such as the Chamfer Distance (CD) [6]:

$$\mathrm{CD} = \frac{1}{N_X} \sum_{x_i \in X} \min_{y_i \in Y} \|x_i, y_i\|^2 + \frac{1}{N_Y} \sum_{y_i \in Y} \min_{x_i \in X} \|y_i, x_i\|^2, \tag{1}$$

or the symmetric Hausdorff Distance (HD) [5]:

$$\mathrm{HD} = \max \left\{ \sup_{x_i \in X} \inf_{y_i \in Y} \|x_i, y_i\|, \sup_{y_i \in Y} \inf_{x_i \in X} \|y_i, x_i\| \right\}. \tag{2}$$

Completeness (Cmp) has also been used to assess ear scans [4], representing the percentage of reference points in X within 1 mm of the nearest point in Y. Additionally, a Mesh Distance (MD) is defined using a distance function $d_M(y, M)$, measuring the deviation between y and the nearest face of a reference mesh M [5]:

$$\mathrm{MD} = \frac{1}{N_Y} \sum_{y_i \in Y} d_M(y_i, M)^2. \tag{3}$$

In this work, the tested DNNs for point cloud denoising are PointCleanNet (*PCN*) [6] and DMRDenoise (*DMR*) [7]. A classical polynomial filtering method (*POL*) is used as a benchmark [8]. These models operate on local point cloud patches, i.e. neighbouring points within a radius r from a query point (r-balls), with r typically set as a small fraction of the point cloud bounding box diagonal (l) [5].

PCN processes an input r-ball centred at a query point ($\widehat{y}_i$) through a spatial transformer network, rotating patch points to a canonical orientation. Subsequently, a feature extractor and symmetric operator compute order-invariant patch features, which are passed to a regressor predicting an optimal displacement vector for $\widehat{y}_i$. Supervised training is performed on noisy-clean point patch pairs extracted from generic shapes with Gaussian noise at varying levels. Training patches consist of several randomly sampled r-balls with $r = 5\,\%$ of l. The loss function is formulated as a weighted combination of two terms: $\mathcal{L}_\mathrm{s}$, measuring proximity between X and $\widetilde{Y}$, and $\mathcal{L}_\mathrm{r}$, promoting a uniform point distribution in $\widetilde{Y}$. These are defined as the minimum and maximum squared distances between the denoised $\widehat{y}_i$ and the corresponding clean patch in X, respectively [6].

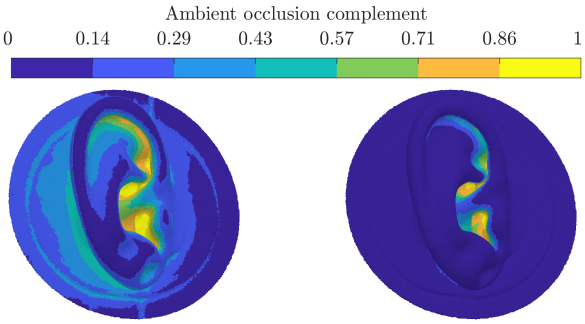
DMR employs an auto-encoder architecture comprising a representation encoder and a manifold reconstruction decoder. The encoder features a differentiable down-sampling unit selecting points closer to the underlying geometry. The decoder maps each sampled point and its neighbours onto a patch manifold, enabling up-sampling of the denoised point cloud. Supervised training is done similarly to *PCN*, using CD as the loss function [7].

POL operates by fitting an implicit plane to an r-ball and approximating the point patch with a 2D polynomial of arbitrary order. The estimated surface normal is then used to project the query point onto the polynomial surface [8]. In this study, r is set to 3 mm, and a 2$^{\mathrm{nd}}$ order polynomial approximation is applied.

2.2 Ear dataset

DNNs require extensive datasets for effective training. Thus, multiple scans are conducted on three dummy heads: the Neumann KU100 (KU1), the Neutrik COR-TEX MK2 (COR), and the GRAS KEMAR 45BB (KEM) with its original ears and two sets of custom 3D-printed pinnae [9]. Photogrammetric scans are performed on the optically treated manikins, using white matte scanning spray to enhance surface reconstruction [4]. Following the methodology in [3], the scans yield ear point clouds of approximately $100\,000$ points. A single scan is conducted for COR and the three KEM configurations, whereas the KU1 scan is repeated four times after reapplying the surface treatment. Precise reference geometries are also required. For KU1, this is taken from a previous scan, where a uniform triangular mesh with an Average Element Length (AEL) of $0.6\,\mathrm{mm}$ is acquired through a laser scanner [3]. The KEM geometry is derived from its CAD model, meshed with an AEL of $0.6\,\mathrm{mm}$, also replacing its original pinnae with the custom ear geometries [9]. For COR, a structured-light scan from a study on binaural room impulse responses is used [10], featuring an AEL of $3.5\,\mathrm{mm}$ at the head and below $1.5\,\mathrm{mm}$ at the ears.

Given the low amount of scanned data, the properties of the Real Photogrammetric Error (RPE) are extracted and used to design a Synthetic Photogrammetric Error (SPE) replicating the real one. This is achieved by aligning the scans with the corresponding reference mesh and computing the Signed Distance Function (SDF) between them. The SDF represents the Euclidean distance, with a positive sign when the point lies outside the underlying surface and vice versa. Fitting the best distribution to the SDF for each scanned ear shows that it is generally well approximated by a t-distribution ($\mathcal{T}$) with median parameters: shape $\nu = 1.95$, location $\mu = 0.02\,\%$ of l, and scale $\sigma = 0.20\,\%$ of l, where $l \approx 117\,\mathrm{mm}$. Furthermore, Cmp and maximum deviation (Mxm) are determined. The former, relating to holes in scanned point clouds, has a median value of $\mathrm{Cmp} = 93.45\,\%$, while $\mathrm{Mxm} = 3.78\,\%$ of l. Ear regions exhibiting the largest RPE are approximated using the Ambient Occlusion (AO) [4], computed as the fraction of unobstructed directions over a hemisphere centred at a surface point, ranging from 0 to 1. Instead of AO, its complement, i.e. $\mathrm{AOc} = 1 - \mathrm{AO}$, is used to obtain larger values at the most occluded regions, where errors are concentrated. To further emphasise concave pinna structures, AOc is raised to an integer power p, found by maximising the Pearson's correlation coefficient between AOc^p and the absolute SDF at each point, yielding a median value of $p = 3$ across all scans. A visualisation of AOc and its raised version with $p = 3$ is shown in Fig. 1.



(a) $p = 1$. (b) $p = 3$.

Figure 1: KU1 right ear point cloud from the laser scan. The colour shows the ambient occlusion complement, raised to the integer power p.

Based on the RPE parameters, SPE realisations are generated and used to corrupt individual meshes sourced from the HUTUBS HRTF database [11]. The processing, mainly done in Pymeshlab [1], starts by uniformly remeshing a head mesh to an AEL of $0.3\,\mathrm{mm}$ and computing the AO at its vertices. A clean left ear point cloud X, with $N_X \approx 100\,000$, is obtained by extracting mesh vertices and related AO within a cylinder of $40\,\mathrm{mm}$ radius aligned with the interaural axis. A corresponding noisy point cloud Y is created by subsampling X to introduce incompleteness, and generating an additive error $E = \{e_i, i = 1, \cdots, N_Y\}$ following a scaled t-distribution $\widehat{\mathcal{T}}$, used to displace the subsampled X according to:

$$y_i = x_i + e_i = x_i + \widehat{\mathcal{T}}_i \cdot n_i, \tag{4}$$

where n_i denotes the point normal at x_i. The procedure is detailed in Alg. 1, where weighted random choice algorithms (RChoice) are employed to subsample X and assign larger $|E|$ values to the most concave pinna regions using a weight w proportional to AOc^p. Different noise levels are generated by defining a range of σ values (σ_{range}) between $0.1\,\%$ and $0.5\,\%$ of l. Cmp is selected based on σ, varying from $\mathrm{Cmp}_{\mathrm{min}} = 80\,\%$ to $\mathrm{Cmp}_{\mathrm{max}} = 98\,\%$. These values approximate the observed σ and Cmp ranges in the RPE.

A comparison between the RPE and SPE at the right ear of the KU1 is shown in Fig. 2. While similarities are

[1] https://zenodo.org/records/14974687

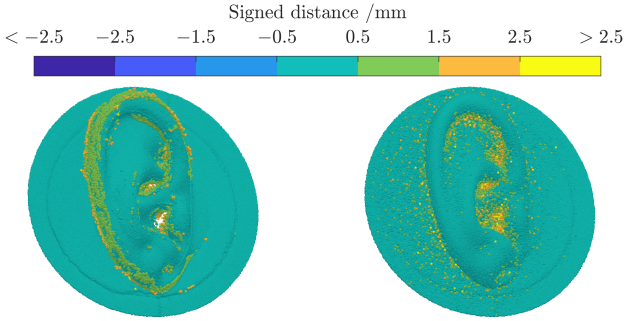
Algorithm 1 Synthetic photogrammetric error

Input: clean point cloud X, N_X, l
error parameters ν, μ, σ_{range}, p
weight $w = \text{AOc}^p$
Output: noisy point cloud Y

for $\sigma \in \sigma_{\text{range}}$ **do**
 # Subsample X
$\eta \leftarrow (\sigma - \sigma_{\min})/(\sigma_{\max} - \sigma_{\min})$
$\text{Cmp} \leftarrow \text{Cmp}_{\max} * (1 - \eta) + \text{Cmp}_{\min} * \eta$
$N_Y \leftarrow \text{Integer}(N_X * \text{Cmp})$
$i_S \leftarrow \text{RChoice}(i_X, \text{size}=N_Y, \text{weight}=1 - w)$
$X_S \leftarrow X[i_S]$ *# Subsampled X*
 # Generate scaled $\mathcal{T}$ samples
$\mathcal{T} \leftarrow \text{Generate}\mathcal{T}(\nu, \mu * l, \sigma * l, \text{size}=N_Y)$
for $\mathcal{T}_i \in \mathcal{T}$ **do**
if $|\mathcal{T}_i| > \text{Mxm}$ **then**
$\widehat{\mathcal{T}}_i \leftarrow \mathcal{T}_i / \max|\mathcal{T}_i| * \text{Mxm}$
else
$\widehat{\mathcal{T}}_i \leftarrow \mathcal{T}_i$
end if
end for
 # Displace Y based on assigned $\widehat{\mathcal{T}}$ samples
$i_{\widehat{\mathcal{T}}} \leftarrow \text{SortIndex}(|\widehat{\mathcal{T}}|, \text{order=descending})$
$i_{Y_w} \leftarrow \text{RndChoice}(i_Y, \text{size}=N_Y, \text{weight}=w)$
$E[i_{Y_w}] \leftarrow \widehat{\mathcal{T}}[i_{\widehat{\mathcal{T}}}]$ *# Sorted error samples*
$Y \leftarrow \text{AddNoise}(X_S, E)$ *# Eqn. (4)*
end for

seen, particularly at the most concave pinna structures, some differences are noted. Specifically, the RPE is higher than the SPE at the most convex parts. However, errors in these regions have a low influence on the HRTFs [3]. As the SPE distribution is designed to follow that of the RPE, objective metrics computed on both yield similar values.

The left ears of 20 HUTUBS subjects are selected as training data, with 6 SPE levels applied, i.e. σ ranging from 0 % to 0.5 % of l in steps of 0.1 %, with the 0 % value included to train the DNNs to preserve the clean geometry [6]. Validation and testing datasets consist of 10 left ears each, with 3 error levels, i.e. $\sigma = 0.1\,\%,\ 0.3\,\%$ and $0.5\,\%$. This results in 120 noisy-clean data pairs for training, and 30 each for validation and testing, consistent with the amount of data used in the original training of the DNNs [6, 7]. Additionally, a dataset is created using the dummy head scans, including both left and mirrored right ears, resulting in 16 samples from 8 dummy head scans. This is done to evaluate the denoising effect on the RPE.



(a) Real error. (b) Synthetic error.

Figure 2: KU1 right ear point clouds with real and synthetic photogrammetric error. The colour shows the SDF from the laser scan, cropped at $\pm 2.5\,\text{mm}$.

2.3 HRTF computation and assessment

Graded meshes suitable for HRTF computation are obtained from the scans as in [3]. Similarly, meshes are generated after merging the denoised ears with the head scan. The adaptive order FEM approach is leveraged for the HRTF computations [12], using the fine accuracy level, and employing an automatically matched layer to impose free-field conditions [13]. The results of this method are validated against measured data in [14]. Given the unknown acoustic impedance of the manikin materials and its low effect on HRTFs [15], rigid boundary conditions are applied. The results are compared through the Inter-Subject Spectral Difference (ISSD) [3]:

$$\text{ISSD} = \frac{1}{N_\Psi} \sum_\Psi \text{Var}_f \left(\tilde{H}_i(\Psi, f) - \tilde{H}_j(\Psi, f) \right), \quad (5)$$

with Var_f denoting variance between the dB amplitude of two gammatone-filtered HRTFs, $\tilde{H}_i$ and $\tilde{H}_j$, averaged across frequency f and number N_Ψ of incidence angles Ψ. Since low deviations are seen in binaural metrics between reference and photogrammetric KU1 HRTFs in [3], these are omitted. A sagittal plane localisation model is used as a perceptually inspired metric [2], implemented in the Auditory Modeling Toolbox 1.5 [2]. This model compares a template and a target HRTF through a virtual sound localisation experiment, predicting localisation results in terms of Quadrant Error (QE) and Polar Error (PE), relating to percentage of confusion and local angular precision between target and localised sound source, respectively.

[2] `https://amtoolbox.org/`

3. RESULTS AND DISCUSSION

3.1 Denoising

Different trainings of the DNNs are tested, including pretrained models (*PRT*) on the original dataset of generic shapes with unstructured noise [6, 7], and refined versions (*RFN*) through fine-tuning on the ear dataset with SPE. Additionally, *PCN* is further modified to better tackle scanning errors by adjusting its loss function's proximity term to $\widehat{\mathcal{L}_\mathrm{s}} = \mathcal{L}_\mathrm{s} \cdot \widehat{\mathrm{AOc}_{y_i}}$, where $\widehat{\mathrm{AOc}_{y_i}}$ represents AOc at y_i normalised by its average value on the point cloud. This model, denoted as *PCN* modified (*PCNm*), is retrained from scratch (*RTR*) on the ear dataset with SPE and further fine-tuned on a subset of dummy head scans with RPE. For this, the COR and KMR scans are used for training and validation, while the KU1 geometries are reserved for testing. While other modifications and trainings of the DNNs have been tried, their results are not included in the following analyses given their lower performance. A summary of the tested DNN models is listed in Tab. 1.

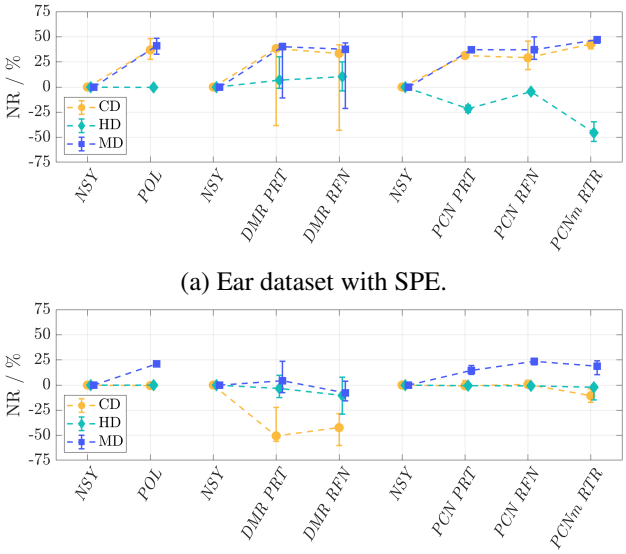
Table 1: Summary of training and fine-tuning datasets for tested DNN models.

DNN ID	Training data	Fine-tuning data
DMR PRT	original *DMR*	/
DMR RFN	original *DMR*	ears with SPE
PCN PRT	original *PCN*	/
PCN RFN	original *PCN*	ears with SPE
PCNm RTR	ears with SPE	ears with RPE

The denoising effect is evaluated on ear geometries affected by SPE and RPE. For the latter, only the KU1 samples are tested, as the others are used for refining *PCNm*. The CD, HD, and MD, defined in Eqn. (1), (2), and (3), respectively, are used for assessment. To facilitate comparisons, the Noise Reduction (NR) is computed:

$$\mathrm{NR} = \frac{\mathcal{M}(Y, X) - \mathcal{M}(\widetilde{Y}, X)}{\mathcal{M}(Y, X)}, \tag{6}$$

where $\mathcal{M}$ denotes a distance metric between clean, noisy, and denoised point clouds, i.e. X, Y, and $\widetilde{Y}$, respectively. Effective denoising corresponds to positive values in Eqn. (6), while negative NR indicates increased error in $\widetilde{Y}$. The resulting NR for data with RPE and SPE is shown in Fig. 3, plotted separately for each denoising method, starting from the initial noisy photogrammetric point cloud (*NSY*), having $\mathrm{NR} = 0$.



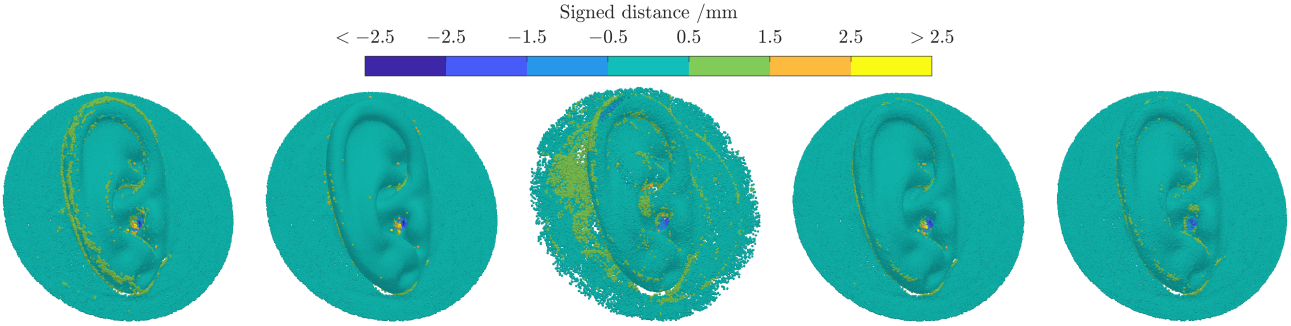
(a) Ear dataset with SPE.

(b) Ear dataset with RPE.

Figure 3: Noise reduction of denoising algorithms on different datasets. The bars show the interquartile range, while the markers show the median.

The results in Fig. 3a for ear data with SPE show that all models can reduce CD and MD, though *DMR* exhibits high variability. However, their impact on HD is minimal or even negative, e.g. for *PCN*. Since HD is influenced by outliers, i.e. single points at a large distance from X, this suggests that the denoising methods struggle to address them. Refinement or retraining appears beneficial only for *PCN*. Indeed, *PCNm RTR* is the only DNN outperforming *POL* in CD and MD, reaching $\mathrm{NR} \approx 50\,\%$. The other models show limited or no improvement over their pretrained versions. For RPE, the results in Fig. 3b indicate a weaker denoising effect. Only *POL*, *PCN RFN*, and *PCNm RTR* achieve $\mathrm{NR} > 20\,\%$ for MD, while other models and metrics show limited or adverse effects. This can relate to differences between SPE and RPE, hindering generalisation of the models refined on the SPE alone. Although *PCNm RTR* is fine-tuned on a small set of samples with RPE, its denoising performance remains modest, probably due to the low amount of data for refinement.

A visualisation of the denoising effect of the various models on a photogrammetric scan (*PHO*) of the KU1 right ear is provided in Fig. 4. Only samples denoised with



(a) *PHO.* (b) *POL.* (c) *DMR RFN.* (d) *PCN RFN.* (e) *PCNm RTR.*

Figure 4: KU1 right ear point cloud from the photogrammetric scan *PHO*, denoised with *POL* and refined *RFN* or retrained *RTR* DNNs (*PCN*, *DMR*). The colour shows the SDF from the laser scan, cropped at $\pm 2.5\,\mathrm{mm}$.

POL and the refined or retrained versions of the DNNs are displayed. It is noticeable that different algorithms yield different outcomes. *POL*, enforcing smoothness in the geometry, is effective in removing overall noise, although outliers persist in the most concave pinna regions. *DMR* tends to reduce errors in some pinna locations while introducing deviations in others, likely due to the down-sample up-sample technique, which may be hindered by the complex pinna shape. *PCN RFN* and *PCNm RTR* exhibit similar overall denoising, with a notable difference in the most concave locations, where *PCNm RTR* specifically targets denoising. Indeed, fewer outliers are observed in these regions in the *PCNm RTR* results.

3.2 HRTFs

The denoising effect on the point clouds is further evaluated by computing HRTFs on the KU1 geometries and assessing them in terms of their ISSD, defined in Eqn. (5), and sagittal plane localisation error, in relation to the HRTFs of the reference KU1 laser scan (*LAS*). The initial photogrammetric scans *PHO* and denoised samples with *POL* and the refined or retrained DNNs are compared. The results are summarised in Tab. 2. It is notable that the only model capable of consistently reducing the deviation of the *PHO* HRTFs from the reference is *PCNm RTR*, though by a small amount. The resulting deviation is slightly higher than that between reference and accurately measured HRTFs in [3], considered to relate to low differences in localisability in binaural renderings. Conversely, *POL* tends to increase the HRTF deviation, while the other models show little to no benefit, with improvements in only some HRTF metrics. Although a clear explanation for these results is not trivial, the differences might be linked to the specific denoising effect of each method, as displayed in Fig. 3. For instance, *POL* and *PCN RFN* are more effective on the overall point clouds but leave errors in the most critical pinna regions, which are better handled by *PCNm RTR*. Conversely, *DMR* might over-smooth the geometry, introducing additional discrepancies.

Table 2: Median deviation of objective and perceptually inspired metrics for scanned and denoised KU1 HRTFs from the KU1 laser scan HRTF.

HRTF ID	Median deviation from reference		
	ISSD /dB^2	QE /%	PE /°
PHO	1.54	8.53	4.83
POL	2.05	9.81	5.70
DMR RFN	1.58	8.48	3.96
PCN RFN	1.50	8.71	4.42
PCNm RTR	1.28	5.58	3.21

The QE and PE results are further summarised in Fig. 5, where the median and interquartile range of the localisation errors obtained on the computed KU1 HRTFs are compared to the expected error ranges with individual accurate, inaccurate, and non-individual KU1 HRTFs [3]. While the median values show that the deviation is the lowest with *PCNm RTR*, it is notable that all results exhibit a large spread, similar to that of the original *PHO* data, which the denoising models struggle to reduce. This large variability, ranging from values close to the deviation be-

tween reference and accurately measured data to levels exceeding the deviation observed with individual inaccurate results affected by measurement error, is attributed to different amounts of scanning errors in the initial photogrammetric data, and necessitates further analyses.

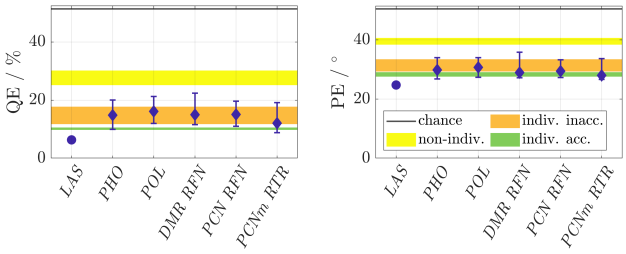


Figure 5: Sagittal plane localisation error with template KU1 laser scan HRTF and different scanned and denoised target HRTFs. The bars show the interquartile range, while the markers show the median. The envelopes show the interquartile range of several individual and non-individual measured HRTFs. The black line corresponds to chance rate.

3.3 Correlation

Evaluating the correlation between geometric and HRTF metrics computed on ear point clouds and their corresponding HRTFs provides insights into the most relevant metrics for assessing ear geometries. These findings can inform the selection of an appropriate loss function to optimise the denoising models. Thus, the Pearson's correlation coefficient is computed between all tested geometric metrics and the ISSD, QE, and PE calculated on the KU1 geometries. The resulting median values are shown in Fig. 6 as a function of the integer power p applied to $\widehat{\mathrm{AOc}}$ for weighting the distance from the reference geometry before computing the geometric metrics.

The correlation between geometric and HRTF metrics remains low when no weighting is applied, i.e. for $p = 0$. In this case, the highest correlation, observed for MD, reaches 0.3, while CD exhibits values close to 0; hence, the latter is ineffective at predicting deviations between HRTFs computed on the tested geometries. All correlation coefficients increase with p, suggesting that the distance weighting emphasises points in acoustically relevant pinna regions. All metrics achieve coefficients above 0.5 at $p = 3$, with CD and MD exceeding 0.7. Thus, employing these weighted metrics is recommended to better

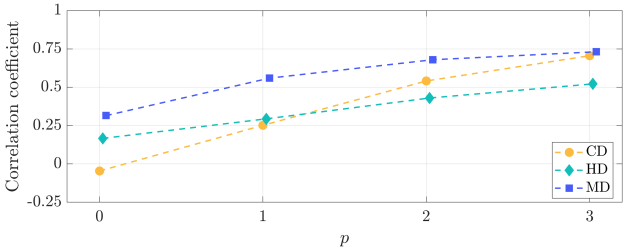


Figure 6: Median correlation coefficients between geometric and HRTF metrics computed on the scanned and denoised KU1 point clouds and related HRTFs, as a function of the integer power p used to raise the AOc for distance weighting.

assess ear point clouds and as a potential loss function in DNN methods to improve denoising.

As an alternative approach, a Parametric Pinna Model (PPM) can be used to obtain a reliable mesh from measured or scanned data of a target ear. This method morphs a precise ear mesh to match the target geometry by tuning a set of parameters [1]. DNNs can be used to predict optimal PPM parameters from pinna images, yielding promising outcomes [16]. While the results of that study are assessed with a metric similar to HD, the current findings suggest that other metrics, such as the weighted CD or MD, may offer a more reliable quantitative assessment of the match between pinna geometries.

4. CONCLUSION

Denoising of photogrammetric pinna scans to enhance HRTF accuracy is explored using both DNNs and a classical approach. Two DNNs are fine-tuned on ear data corrupted with synthetic errors, with one model further modified to better target errors in concave pinna regions. Geometric metrics comparing the original, denoised, and reference scans indicate that fine-tuning is not always beneficial; only one DNN outperforms the classical method on synthetic data and performs comparably on real scans. This likely reflects that the synthetic error only partially represents real noise, highlighting the need for more data to improve the former or enable training on real scans.

HRTF comparisons show that only the modified DNN consistently reduces deviations from reference data, though the improvements remain marginal and may benefit from a more suitable loss function. Correlation analyses reveal that the MD is an effective metric to assess ear

point clouds in relation to their HRTFs, and that applying a distance weighting factor proportional to AOc further enhances the correlation. These findings offer valuable insights for refining denoising techniques and improving HRTF individualisation based on scanned data.

5. ACKNOWLEDGMENTS

The European Commission is gratefully acknowledged for their support of the VRACE research project (GA 812719). Internal Funds KU Leuven are gratefully acknowledged for their support.

6. REFERENCES

[1] K. Pollack, W. Kreuzer, and P. Majdak, "Modern Acquisition of Personalised Head-Related Transfer Functions - An Overview," in *Advances in Fundamental and Applied Research on Spatial Audio*, IntechOpen, 2022.

[2] R. Baumgartner, P. Majdak, and B. Laback, "Modeling sound-source localization in sagittal planes for human listeners," *The Journal of the Acoustical Society of America*, vol. 136, no. 2, pp. 791–802, 2014.

[3] F. Di Giusto, S. van Ophem, W. Desmet, and E. Deckers, "Analysis of laser scanning and photogrammetric scanning accuracy on the numerical determination of Head-Related Transfer Functions of a dummy head," *Acta Acustica*, vol. 7, no. 53, pp. 1–21, 2023.

[4] A. Reichinger, P. Majdak, R. Sablatnig, and S. Maierhofer, "Evaluation of Methods for Optical 3-D Scanning of Human Pinnas," in *Proc. of International Conference on 3D Vision*, (Seattle, USA), pp. 390–397, 2013.

[5] L. Zhou, G. Sun, Y. Li, W. Li, and Z. Su, "Point cloud denoising review: from classical to deep learning-based approaches," *Graphical Models*, vol. 121, no. C, pp. 1–16, 2022.

[6] M.-J. Rakotosaona, V. La Barbera, P. Guerrero, N. J. Mitra, and M. Ovsjanikov, "PointCleanNet: Learning to Denoise and Remove Outliers from Dense Point Clouds," *Computer Graphics Forum*, vol. 39, no. 1, pp. 185–203, 2019.

[7] S. Luo and W. Hu, "Differentiable Manifold Reconstruction for Point Cloud Denoising," in *Proc. of the 28th ACM International Conference on Multimedia*, (Seattle, USA), pp. 1330–1338, 2020.

[8] F. S. Egner, L. Sangiuliano, R. F. Boukadia, S. van Ophem, W. Desmet, and E. Deckers, "Polynomial filters for camera-based structural intensity analysis on curved plates," *Mechanical Systems and Signal Processing*, vol. 193, no. 110245, pp. 1–22, 2023.

[9] D. Sinev, F. Di Giusto, J. Peissig, S. van Ophem, and E. Deckers, "Individual Ear Replicas with Complete Ear Canals Compatible with an Artificial Head," in *Proc. of DAGA 2023*, (Hamburg, Germany), pp. 166–169, 2023.

[10] M. Rychtáriková, T. Van den Bogaert, J. Wouters, and G. Vermeir, "Binaural sound source localization in real and virtual rooms," *Journal of the Audio Engineering Society*, vol. 57, no. 4, pp. 205–220, 2009.

[11] F. Brinkmann, M. Dinakaran, R. Pelzer, P. Grosche, D. Voss, and S. Weinzierl, "A cross-evaluated database of measured and simulated HRTFs including 3d head meshes, anthropometric features, and headphone impulse responses," *Journal of the Audio Engineering Society*, vol. 67, no. 9, pp. 705–718, 2019.

[12] H. Bériot, A. Prinn, and G. Gabard, "Efficient implementation of high-order finite elements for Helmholtz problems," *International Journal for Numerical Methods in Engineering*, vol. 106, pp. 213–240, 2016.

[13] H. Bériot and A. Modave, "An automatic perfectly matched layer for acoustic finite element simulations in convex domains of general shape," *International Journal for Numerical Methods in Engineering*, vol. 122, no. 5, pp. 1239–1261, 2020.

[14] F. Di Giusto, S. van Ophem, and E. Deckers, "Investigating Adaptive Order Schemes for Efficient Head-Related Transfer Function Computation," in *Proc. of DAGA 2025*, (Copenhagen, Denmark), 2025.

[15] F. Di Giusto, D. Sinev, K. Pollack, S. van Ophem, and E. Deckers, "Analysis of Impedance Effects on Head-Related Transfer Functions of 3D Printed Pinna and Ear Canal Replicas," in *Proc. of Forum Acusticum 2023*, (Turin, Italy), pp. 323–330, 2023.

[16] F. Pausch, F. Perfler, N. Holighaus, and P. Majdak, "Comparison of deep-neural-network architectures for the prediction of head-related transfer functions using a parametric pinna model," in *Proc. of Forum Acusticum 2023*, (Turin, Italy), pp. 2329–2334, 2023.