Growth modelling of children's ear canals using implicit neural distance representations

Karen Guldhammer-Skov[1,2], Kristine Aavild Sørensen[2,3], Paula López Diez[4], Christian Palama[5], Erik F. Kjærbøl[6], Gerard Encina-Llamas[6,7], Rasmus R. Paulsen[2], and Miguel A. González Ballester[1,8]

[1] BCN Medtech, Department of Engineering, Pompeu Fabra University, Barcelona, Spain (karen.guldhammer@upf.edu)
[2] DTU Compute, Technical University of Denmark, Kongens Lyngby, Denmark
[3] Novo Nordisk A/S, Søborg, Denmark
[4] Eriksholm Research Centre, Demant A/S, Smørum, Denmark
[5] Department of Otorhinolaryngology, Zealand University Hospital, Køge, Denmark
[6] Copenhagen Hearing and Balance Center, Rigshospitalet University Hospital, Copenhagen, Denmark
[7] Catalan Hearing and Balance Sciences Group, Faculty of Medicine, University of Vic - Central University of Catalonia, Vic, Catalonia – Spain
[8] ICREA, Barcelona, Spain

Abstract. Rapid ear canal growth in infancy and childhood poses a challenge for maintaining accurate hearing aid fitting during critical developmental stages. A method to predict subject-specific ear canal growth has the potential to optimize paediatric ear mould preparation, enabling timely refitting while reducing the need for frequent visits to specialised audiological clinics. We present a framework for modelling ear canal growth using implicit neural representations. Our approach builds on unsigned distance fields and uses latent vectors for capturing subject-specific growth trajectories. The contributions of this work are threefold: (i) A novel use of implicit neural representations for growth modelling, benefiting from their inherent continuity across both space and time; (ii) a robust scheme for handling variable-length data sequences and irregularly sampled time points; and (iii) an approach that accommodates open surface meshes, enabling the modelling of open anatomical structures such as the ear canal without the need to artificially close the geometry. We evaluate the method on a longitudinal dataset of 912 3D ear canal scans from children aged 1 month to 10.5 years, and show that it robustly models anatomical growth, achieving a mean Chamfer distance of 0.65 mm in predictions of ear canal shapes at future time points.

Keywords: Spatio-temporal modelling · Latent space embedding · Unsigned distance fields · Computational audiology · Hearing impairment · Paediatric longitudinal data

1 Introduction

Newborn hearing screening programs are widely implemented in Western countries to enable early detection of hearing impairment and timely intervention [11]. Infants diagnosed with hearing loss are typically fitted with hearing aids (HAs) using custom ear moulds made from silicone-based ear canal impressions, which are scanned during manufacturing, giving the digital 3D surface mesh data used in the present work.

The human ear canal exhibits substantial inter-subject variability and undergoes gradual and clinically relevant geometric changes over time (see examples in Fig. 1). This is particularly pronounced in infants, whose rapid ear canal growth causes HA fitting mismatch and sound leakage within weeks to months of the initial fitting [14, 21]. This period is especially critical for language development, and inadequate HA fitting can compromise effective acoustic stimulation during a key stage of sensory maturation. Accurate computational modelling of ear canal shape and its temporal evolution is therefore of critical interest. A method capable of predicting subject-specific ear canal growth from a limited number of observations could reduce the need for frequent visits to specialised audiological clinics to obtain new physical impressions, enable timely HA refitting using model-predicted moulds, and ease the associated burden on healthcare resources.

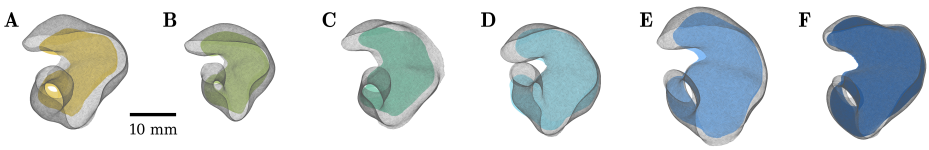

Fig. 1: Six patients' first (coloured) vs. last (grey) ear canal shapes.
A: 1 mo vs. 2 yr 10 mo **B**: 2 mo vs. 2 yr 0 mo **C**: 5 mo vs. 4 yr 8 mo **D**: 2 yr 2 mo vs. 4 yr 7 mo **E**: 4 yr 7 mo vs. 8 yr 5 mo **F**: 8 yr 0 mo vs. 10 yr 5 mo.

Capturing and describing anatomical variation has been extensively done using statistical shape models (SSMs) [1, 5, 22], also of the ear canal [8, 16, 17, 20]. SSMs construct a compact, low-dimensional linear subspace of shape deformations from a training population, typically via principal component analysis across aligned surfaces. Notable restrictions, however, include the requirement for point correspondence across shapes and the difficulty of incorporating longitudinal data.

Recent advances in geometric deep learning have introduced implicit neural representations (INRs) as a promising method for modelling complex three-dimensional geometries without requiring fixed topology. In this paradigm, a neural network (NN) learns a continuous function that maps spatial coordinates to a distance to the underlying surface. Initially proposed using signed distance fields (SDFs) [15, 18], the method has since been adapted to use unsigned distance fields (UDFs) for open surfaces [12] and binary occupancy values for closed

shapes [2]. Instead of training a separate network per shape, the implicit function can be conditioned on a low-dimensional latent code, embedding individual shapes or even sequences of shapes, into a continuous latent space that compactly captures population-level variability [9, 12, 19].

Despite their increasing adoption in static shape modelling, INRs have been less explored in the context of spatio-temporal growth modelling. In the image domain, temporally consistent 4D fetal atlases have been constructed from MRI volumes using INRs [4]. Subject-conditioned implicit neural atlases have since been applied to longitudinal anatomical modelling of the fetal brain across gestational ages [6] and extended to the perinatal period with multimodal and pathology conditioning in latent space [7]. All approaches, however, operate exclusively on volumetric intensity images.

Work on longitudinal modelling of 3D surface meshes is sparser. An INR-based surface mesh method has been extended to the time dimension to model the cyclic movement of the beating heart [19], building on earlier work in open surface INRs [12]. This method, however, is designed for equally spaced temporal sampling and cyclic motion. Ear canal growth differs fundamentally, as the number of scans per patient varies, temporal spacing is irregular, and the deformation is non-cyclic. Our problem therefore requires extrapolation of future shapes from past observations, rather than interpolation within a known cyclic regime. The present work addresses these gaps and contributes novelty by: **(i)** Using an INR framework for growth modelling in paediatrics, benefitting from INRs being continuous in both space and time; **(ii)** handling non-uniform sequence lengths and irregular temporal sampling; and **(iii)** accommodating open surface meshes representing open anatomies such as the ear canal, without imposing restrictions such as artificially closing the geometry.

2 Methodology

Three-dimensional surfaces can be represented explicitly (as point clouds, meshes or voxel grids) or implicitly by e.g. neural distance fields. We adopted an implicit neural distance representation based on UDFs, where a neural network maps a spatio-temporal coordinate to its unsigned distance to the nearest surface point.

2.1 Overall framework

A spatio-temporal neural distance field is defined as a neural function that maps an arbitrary space-time coordinate $\mathbf{x} = (\mathbf{p}, t)$, comprising the 3D spatial coordinate $\mathbf{p}$ and time t, to an unsigned distance $\hat{d} = f_\theta(\mathbf{x})$ to the nearest location on the surface at time t.

To learn the temporal UDF for each shape in each patient sequence, the network is trained on a set $\mathcal{X}$ of space-time coordinates paired with their corresponding ground-truth UDF values, formally defined as $\mathcal{X} := \{(\mathbf{x}, d) : \mathrm{UDF}(\mathbf{x}) = d\}$,

where $d \geq 0$. The network parameters θ are optimized by minimizing the following loss between the predicted and the ground-truth distance values:

$$\underset{\theta,\, \{\mathbf{z}_n\}_{n=1}^{N}}{\arg\min} \sum_{n=1}^{N} \left[\sum_{k=1}^{K_n} \mathcal{L}(f_\theta(\mathbf{z}_n,\, \mathbf{x}_{n,k}),\, d_{n,k}) + \frac{1}{\sigma^2} \|\mathbf{z}_n\|_2^2 \right] \tag{1}$$

where $\mathbf{z}_n$ is the patient-sequence specific latent vector and $\mathcal{L}(f_\theta(\mathbf{z}_n, \mathbf{x}_{n,k}), d_{n,k}) = |\text{clamp}(f_\theta(\mathbf{z}_n, \mathbf{x}_{n,k}), \delta) - \text{clamp}(d_{n,k}, \delta)|$ is the clamped L1 loss. The clamping threshold δ defines the spatial extent around the surface within which an accurate distance field is expected to be learned, and $\frac{1}{\sigma^2}\|\mathbf{z}_n\|_2^2$ is a regularization term that encourages a compact latent space by penalizing the magnitude of the latent vectors, required to converge to well-formed solutions. σ^2 controls the trade-off between the L1 loss and the regularization term. N is the number of patient sequences, and K_n is the number of point samples in the sequence for patient n. With this scheme, the K_n points are flexibly aggregated by stacking a variable number of a patient's input shapes at irregular time intervals.

To model multiple anatomical sequences with a single NN, we associate each patient sequence $n \in \{1, \ldots, N\}$ with a learnable latent vector $\mathbf{z}_n$, following the auto-decoder paradigm [15]. This latent vector is concatenated with the space-time coordinate and passed as input to f_θ, such that the network receives $(\mathbf{x}, \mathbf{z}_n)$ and outputs the corresponding predicted unsigned distance value $\hat{d} = f_\theta(\mathbf{x}, \mathbf{z}_n)$. The latent vector thus conditions the neural function on the identity of each anatomical sequence, enabling f_θ to learn a distinct distance field and, consequently, a distinct surface, recovered at the local minima of $f_\theta(\mathbf{x}, \mathbf{z}_n)$ for each shape in the sequence n.

Training: The decoder parameters θ and the individual latent vectors $\{\mathbf{z}_n\}_{n=1}^{N}$ are jointly optimized by backpropagating the loss in Equation 1 (see Fig. 2).

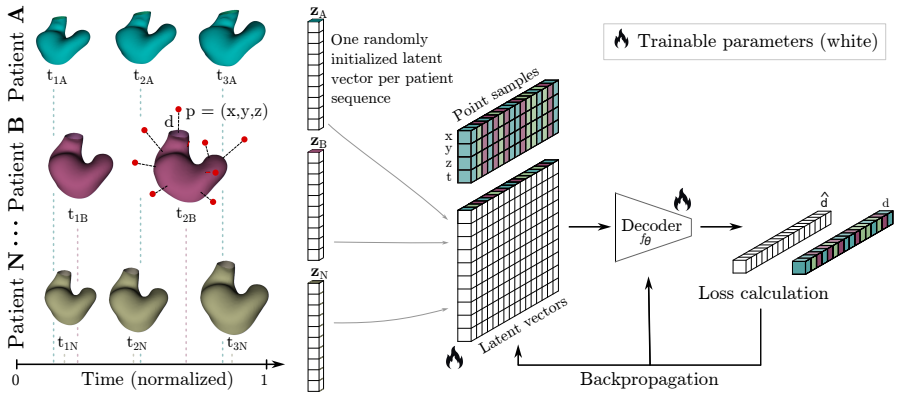

Fig. 2: Training schematic. White elements are trainable while colour elements are fixed or known parameters.

Test-time optimization: During testing (Fig. 3), the optimal latent vector for an unseen sequence of shapes can be found by freezing the network parameters θ and minimizing the following loss:

$$\hat{\mathbf{z}} = \arg\min_{\mathbf{z}} \sum_{k=1}^{K} \mathcal{L}(f_\theta(\mathbf{z}, \mathbf{x}_k), d_k) + \frac{1}{\sigma^2}\|\mathbf{z}\|_2^2 \tag{2}$$

where $\mathbf{x}_k$ and d_k denote the time-space coordinates and corresponding unsigned distance values sampled from the test sequence of shapes. The latent vector is optimized using an arbitrary number of shapes in a sequence, for instance, excluding only the final shape when the objective is to predict future growth. The final forward pass through the trained decoder returns the predicted UDF from which the mesh is reconstructed using the method described in Section 2.2.

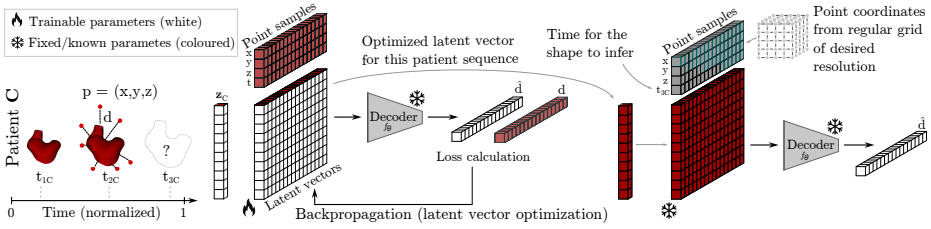

Fig. 3: Test-time optimization. The latent vector for the prior shape sequence is optimized, and the UDF for the time to predict is computed in a forward pass. White elements are trainable; colour elements are fixed or known parameters.

Implementation details: The NN was implemented as a fully connected network with 8 hidden layers, each consisting of 512 units, followed by a final single-unit output layer that produces the predicted unsigned distance value. Rectified linear unit (ReLU) activations are used throughout, as the network is only required to predict non-negative distance values. The concatenation of the space-time coordinate $\mathbf{x}$ and the latent vector $\mathbf{z}_n$ is provided as input to the first layer and reintroduced via a skip connection at the fifth layer. To stabilize and regularize training, weight normalization and a dropout probability of 0.2 are applied across all hidden layers. The latent vector dimension used was 128.

2.2 Surface extraction

To predict the distance from each coordinate to the surface, the latent vector obtained from the test-time optimization was concatenated with coordinates from a grid of size 256^3 along with the target time point, and passed through the network. The extraction of the mesh from this predicted UDF is based on [10]. Given the UDF and a parameter $r>0$, an iso-surface at iso-value r is extracted using the Marching Cubes algorithm [13], resulting in a double-layered mesh. This is then

refined by minimizing a loss combining a distance term that pulls vertices toward the zero level-set and a Laplacian regularization term that promotes geometric stability and uniform vertex distribution, yielding a single-layered shell.

2.3 Data, preprocessing and evaluation metrics

The data come from Region Zealand University Hospital, Denmark, with data usage permission EMN-2024-04703. The dataset consists of 912 raw 3D scans of ear canal impressions of children aged 1 mo-10.5 yrs, with 1-15 impressions per patient, skewed towards younger patients, who have overall more scans as well as longer data sequences (Fig. 4).

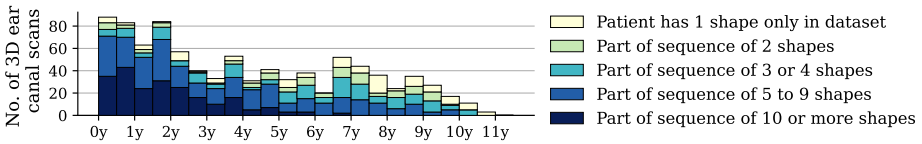

Fig. 4: Distribution of ear canal scans by patient age, colour coded by the length of the sequence the scan belongs to. The data is skewed towards younger children, having more scans overall as well as longer sequences for individual patients.

The meshes were manually cropped to remove superfluous silicone impression material. Left ears were mirrored to a consistent right-ear coordinate frame, and shapes were rigidly aligned using iterative closest point (ICP) [3]. All shapes in a patient sequence were then anchored at a central point at the junction between the ear canal opening and concha cavum, to model growth inwards and outwards from this point. The dataset was scaled by $1/20.25$ mm (set by the largest mesh) to lie within the unit sphere. Approx. 250,000 points were sampled on the mesh and perturbed with two zero-mean Gaussian noise kernels (variance 0.05 and 0.1 mm) to capture fine mesh details, plus 25,000 points sampled randomly within the unit sphere for broader spatial coverage. For each point, the shortest distance, d, to the mesh was computed, along with the patient's age normalized to the full age span covered in the model, t, giving an (x, y, z, d, t) entry per point.

Reconstructed meshes were rigidly aligned to their ground truth via ICP before evaluation, since the model aims to learn geometry, not pose. Prediction and reconstruction performance was assessed using symmetric Chamfer distance (CD) and Hausdorff distance (HD). To contextualize results against a random baseline, we computed a random size-matched measure (termed here "Random same-size Chamfer distance" (rssCD) and "Random same-size Hausdorff distance" (rssHD)), obtained by randomly pairing ground-truth meshes that have surface areas within 2% of each other and originate from different patients, and calculating CD and HD per pair across the dataset.

3 Experiments and results

To account for the highly heterogeneous nature of the data sequences, model training and evaluation was performed with a five-fold cross-validation with an 80/20 train/test split, balancing each fold on 1) total number of sequences, 2) length of sequences, and 3) main age group. To prevent data leakage, left and right ear scans from the same patient were kept within the same fold. Test-set predictions were performed by optimizing the latent vector on all prior shapes in the sequence (e.g., for a patient with a sequence of 3 ear canal scans, shape 2 was predicted based on shape 1, and shape 3 was predicted based on shapes 1 and 2). This was chosen to strictly predict only shapes within a clinically realistic time-step framework. The zero-level set extraction was performed at $r = 0.0125$ mm (normalized scale), corresponding to 0.25 mm in original scale.

Overall, the test-set predictions reproduced the ground-truth dataset growth trajectory, correctly demonstrating rapid growth in younger ages, levelling off at older ages, as seen in Fig. 5. Logarithmic curves were fitted to the ground-truth and prediction data groups. The ground-truth curve falls entirely within the 95% confidence interval of our predicted curve, demonstrating strong statistical agreement. The meshes shown are examples of predictions at different age ranges, coloured by distance to their corresponding ground-truth mesh.

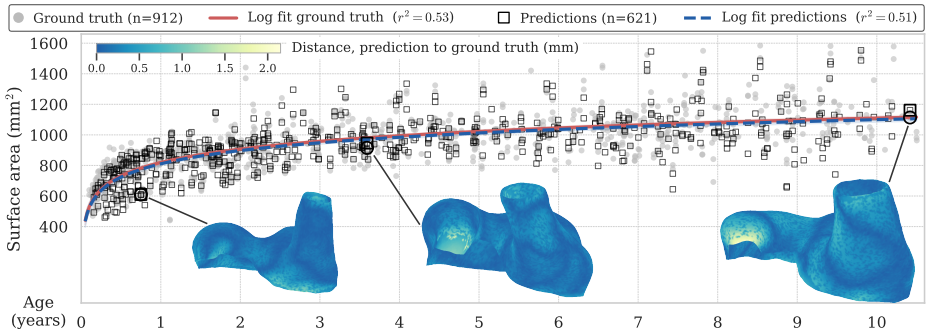

Fig. 5: Surface areas across age for ground-truth and test-set prediction meshes. The inserts are ear canal predictions of patients C, D and F from Fig. 1.

Across all test-set predictions, CD $= 0.65 \pm 0.24$ mm and HD $= 2.58 \pm 1.0$ mm ($n = 621$). Fig. 6 shows CD with respect to age, coloured by prediction time step, alongside the number of prior shapes. HD trends are similar but omitted for brevity. The substantially lower prediction CD relative to rssCD reflects considerable explanatory power of the model (rssCD = 2.12 mm, rssHD = 3.91 mm; metric defined in Section 2.3). Neither age, number of prior shapes, nor prediction time step had a statistically significant effect on CD values (within the time-step scheme present in the clinical dataset). Outlier cases are discussed in Section 4.

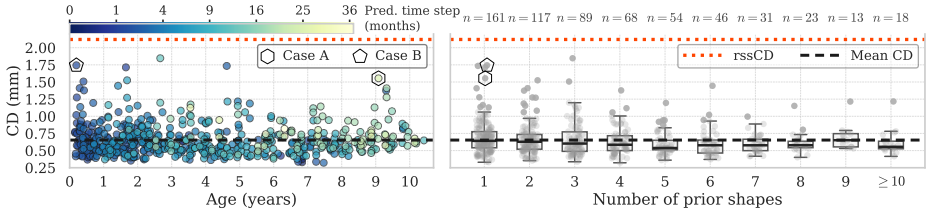

Fig. 6: Chamfer distances across age and number of prior shapes. Case A and B are shown in Fig. 7 and are further described in Section 4.

4 Discussion and Conclusion

This study reports an INR-based modelling framework capable of predicting patient-specific ear canal geometry at future time points. To our knowledge, this is the first work to predict ear canal growth in children, and no direct comparison with prior results is therefore possible. The model shows high precision with a mean CD of 0.65 ± 0.24 mm, well below the rssCD $= 2.12$ mm obtained from random pair comparison between same-size meshes. Whether the predicted shapes are at a clinically acceptable level of accuracy cannot be determined from geometric metrics alone, and a meaningful validation will require audiological assessments such as feedback testing and real-ear measurements.

While the model performs well across most of the cohort, prediction precision decreases under two specific conditions: A) For children close to the dataset maximum age with only one prior shape and a large prediction time step (>1.5 years), leading in some cases to incomplete mesh reconstruction, and B) in sequences in which the last prior shape and the ground-truth shape to be predicted are not geometrically similar, resulting in elevated CD and HD values. Examples of both cases are shown in Fig. 7. Case A failures are attributed mainly to data scarcity at older age ranges (see Fig. 4). Full meshes *can* be reconstructed by increasing r in the mesh extraction, but at the cost of blurring details. Case B is a particularly noteworthy behaviour: A prediction resembling the prior is made, but the calculated precision is low because the ground-truth shape is error-prone or anomalous. Interestingly, this behaviour can be reinterpreted as a clinically valuable property rather than a model failure. This sensitivity to growth inconsistency effectively enables the model to act as an outlier detector, with direct practical relevance for real-time quality control during impression-taking and for flagging pathological anatomies, being potentially transferable to other anatomical structures such as cardiac geometries. This behaviour also accounts for the absence of statistical improvement with additional prior shapes; even long sequences yield imprecise predictions when the most recent prior shape is itself flawed (Fig. 6 outliers).

Within the time-step scheme present in the clinical dataset, where sampling is in most cases naturally adapted to growth velocity at each age, no statistically significant effect of prediction time step was found. However, when extrapolating well beyond the intervals present in the data, time step *does* have a significant

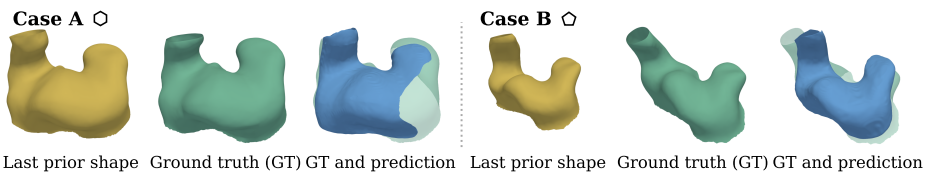

Fig. 7: Examples of deviating predictions. **A**: Incomplete mesh reconstruction due to data scarcity at older ages and large prediction time step, CD=1.55 mm, HD=10.1 mm. **B**: Anomalous ground-truth shape, CD=1.74 mm, HD=3.86 mm.

impact. Full trajectory reconstruction from a single baseline shape, as demonstrated by [19], is not achievable in our model; it targets an unbounded prediction problem, being more demanding than reconstructing a cyclic, bounded motion, while relying on a smaller cohort of sparsely and irregularly spaced shapes.

In the future, we expect to improve prediction accuracy and reduce the number of outlier predictions by integrating the proposed outlier detection, filtering out or down-weighting imprecise meshes. Finally, the strong influence of mesh quality underscores the critical importance of careful data preparation prior to modelling, which is a non-trivial challenge for the complex ear canal geometry.

Acknowledgements

This work was supported by a grant from the William Demant Foundation (23-4995).

Bibliography

[1] Ambellan, F., Lamecker, H., von Tycowicz, C., Zachow, S.: Statistical Shape Models: Understanding and Mastering Variation in Anatomy. In: Biomedical Visualisation, vol. 1156, pp. 67–84. Springer, Cham (2019)

[2] Amiranashvili, T., Lüdke, D., Li, H.B., Zachow, S., Menze, B.H.: Learning continuous shape priors from sparse data with neural implicit functions. Med. Image Anal. **94**, 103099 (2024)

[3] Besl, P., McKay, N.D.: A method for registration of 3-D shapes. IEEE Trans. Pattern Anal. Mach. Intell. **14**(2), 239–256 (1992)

[4] Chen, L., Wu, J., Wu, Q., Wei, H., Zhang, Y.: Continuous Longitudinal Fetus Brain Atlas Construction via Implicit Neural Representation. In: Perinatal, Preterm and Paediatric Image Analysis, vol. 13575, pp. 38–47. Springer Nature Switzerland, Cham (2022)

[5] Cootes, T., Taylor, C., Cooper, D., Graham, J.: Active Shape Models-Their Training and Application. Comput. Vis. Image Underst. **61**(1), 38–59 (1995)

[6] Dannecker, M., Kyriakopoulou, V., Cordero-Grande, L., Price, A.N., Hajnal, J.V., Rueckert, D.: CINA: Conditional Implicit Neural Atlas for Spatio-Temporal Representation of Fetal Brains. In: MICCAI 2024, vol. 15009, pp. 181–191. Springer Nature Switzerland, Cham (2024)

[7] Dannecker, M., Sideri-Lampretsa, V., Starck, S., Mihailov, A., Milh, M., Girard, N., Auzias, G., Rueckert, D.: CINeMA: Conditional Implicit Neural Multi-Modal Atlas for a Spatio-Temporal Representation of the Perinatal Brain. IEEE Trans. Med. Imaging **45**(2), 738–750 (2026)

[8] Darkner, S., Larsen, R., Paulsen, R.R.: Analysis of Deformation of the Human Ear and Canal Caused by Mandibular Movement. In: MICCAI 2007, vol. 4792, pp. 801–808. Springer Berlin Heidelberg, Berlin, Heidelberg (2007)

[9] Erkoç, Z., Ma, F., Shan, Q., Nießner, M., Dai, A.: HyperDiffusion: Generating Implicit Neural Fields with Weight-Space Diffusion. In: ICCV 2023. pp. 14254–14264. IEEE, Paris, France (2023)

[10] Hou, F., Chen, X., Wang, W., Qin, H., He, Y.: Robust Zero Level-Set Extraction from Unsigned Distance Fields Based on Double Covering. ACM Trans. Graph. **42**(6), 1–15 (2023)

[11] Joint Committee on Infant Hearing: Year 2019 position statement: Principles and guidelines for early hearing detection and intervention programs. Journal of Early Hearing Detection and Intervention **4**(2), 1–44 (2019). https://doi.org/10.15142/fptk-b748

[12] Juhl, K.A., Morales, X., de Backer, O., Camara, O., Paulsen, R.R.: Implicit Neural Distance Representation for Unsupervised and Supervised Classification of Complex Anatomies. In: MICCAI 2021, vol. 12902, pp. 405–415. Springer International Publishing, Cham (2021)

[13] Lorensen, W.E., Cline, H.E.: Marching cubes: A high resolution 3D surface construction algorithm. In: SIGGRAPH 1987. pp. 163–169. ACM (1987)

[14] Martinez, J.C., Hwee, G.Z., Yap, L., De Chua, K.W., Kamath, S., Chung, C.K.R., Teo, W.Y.B., Tan, C.K.L., Dritsas, S., Simpson, R.E.: Lexicon for classifying ear-canal shapes. Sci. Rep. **13**(1), 11866 (2023)
[15] Park, J.J., Florence, P., Straub, J., Newcombe, R., Lovegrove, S.: DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation. In: CVPR 2019. pp. 165–174. IEEE, Long Beach, CA, USA (2019)
[16] Paulsen, R., Larsen, R., Nielsen, C., Laugesen, S., Ersbøll, B.: Building and Testing a Statistical Shape Model of the Human Ear Canal. In: MICCAI 2002, vol. 2489, pp. 373–380. Springer, Berlin, Heidelberg (2002)
[17] Paulsen, R.R.: Statistical Shape Analysis of the Human Ear Canal with Application to In-the-Ear Hearing Aid Design. PhD Thesis, Technical University of Denmark, Kgs. Lyngby (2004)
[18] Sitzmann, V., Martel, J.N.P., Bergman, A.W., Lindell, D.B., Wetzstein, G.: Implicit Neural Representations with Periodic Activation Functions. In: NeurIPS 2020. vol. 33. Curran Associates, Inc. (2020)
[19] Sørensen, K., Diez, P., Margeta, J., Youssef, Y.E., Pham, M., Pedersen, J.J., Kühl, T., Backer, O.d., Kofoed, K., Camara, O., Paulsen, R.: Spatio-temporal neural distance fields for conditional generative modeling of the heart. In: MICCAI 2024, vol. 15003, pp. 422–432. Springer Nature Switzerland, Cham (2024)
[20] Unal, G., Nain, D., Slabaugh, G., Fang, T.: Customized Design of Hearing Aids Using Statistical Shape Learning. In: MICCAI 2008, vol. 5241, pp. 518–526. Springer Berlin Heidelberg, Berlin, Heidelberg (2008)
[21] Voss, S.E., Remenschneider, A.K., Farrar, R.M., Myoung, S., Horton, N.J.: Comprehensive Measurements and Analyses of Ear Canal Geometry From Late Infancy Through Late Adulthood: Age-Related Variations and Implications for Basic Science and Audiological Measurements. Trends Hear. **29** (2025)
[22] Zhang, M., Golland, P.: Statistical shape analysis: From landmarks to diffeomorphisms. Med. Image Anal. **33**, 155–158 (2016)