Non-rigid pinna deformation for the calculation of head-related transfer functions

A dissertation submitted for the degree of
Doctor of Philosophy (Ph.D)

in the curriculum for
Sound and Music Computing (UV 094 750)

at the
University of Music and Performing Arts Graz

Dipl.-Ing. Katharina Pollack
Enrolment number 01130273

First supervisor: O.Univ.Prof. Mag.art. DI Dr.techn. Robert Höldrich
Institut für elektronische Musik und Akustik, Universität für Musik und darstellende Kunst Graz

Second supervisor: DI Dr.rer.nat. Privatdozent Piotr Majdak
Institut für Schallforschung, Österreichische Akademie der Wissenschaften

External evaluator: Ao.Univ.Prof. Mag. Dr. Wolfgang Birkfellner
Center for Medical Physics and Biomedical Engineering, Medizinische Universität Wien

Graz, December 17, 2025

kunst uni graz

institut für elektronische musik und akustik

Acknowledgements

In tiefer Dankbarkeit für die Unterstützung widme ich meine Dissertation meiner Familie, insbesondere meinen verstorbenen Großeltern Friedrich, Maria und Michael, meiner Urgroßmutter Johanna und meiner Urgroßtante Maria.

Index of abbreviations

AEL average edge length

AR augmented reality

BCPD Bayesian formulation of coherent point drift

BEM boundary element method

BezierPPM parametric pinna model based on Beziér curves

BVA binaural virtual acoustics

CPD coherent point drift

DTF directional transfer function

FMM fast-multipole method

FPFH fast point-feature histograms

GBCPD geodesic-based Bayesian formulation of coherent point drift

GMM Gaussian mixture model

GT ground truth

HRIR head-related impulse response

HRTF head-related transfer function

ITD interaural time difference

MCT motion coherence theory

NRR non-rigid registration

PPM parametric pinna model

PR photogrammetrical reconstruction

PRGLS preserving global and local structures

VR virtual reality

Kurzfassung

Personenspezifische kopfbezogene Außenohrübertragungsfunktionen (engl. head-related transfer functions, HRTFs) sind von entscheidender Bedeutung für personalisierte binaurale Audiowiedergabe. Bisher gibt es allerdings noch keine zugängliche Methode, die eine HRTF-Erfassung ohne hochspezialisierte Einrichtungen und Geräte ermöglicht. Fotogrammetrische Rekonstruktion (FR) kann die Geometrieerfassung vereinfachen, sodass personenspezifische HRTFs berechnet werden können, aber die notwendige geometrische Genauigkeit von menschlichen Ohrmuscheln (lat. Pinnae) kann nur mit großem Aufwand erreicht werden. In dieser Arbeit wird ein Ansatz zur Verbesserung der Qualität von fotogrammetrisch rekonstruierten Pinnae vorgeschlagen; Dessen Prinzip besteht darin, die Geometrie einer Ohrmuschel so zu verformen, dass sie mit einer personenspezifischen Ohrmuschel übereinstimmt. Es werden zwei Methoden zur Verformung beschrieben: zum Einen nichtparametrische, nichtrigide Registrierung und zum Anderen ein parametrisches Pinnamodell. Die Fähigkeit der Methoden, eine Pinnageometrie mit hoher und niedriger Qualität anzupassen, wurde sowohl im geometrischen, akustischen als auch psychoakustischen Bereich bewertet. Die psychoakustische Evaluierung erfolgte mit Hilfe eines auditiven Modells und über ein Experiment zur Schalllokalisierung. Die Ergebnisse deuten darauf hin, dass die Anwendung des parametrischen Modells eine höhere Lokalisationsgenauigkeit ermöglichte. Zukünftige Untersuchungen sind notwendig, zum Beispiel um den notwendigen Parameterraum abzustecken oder mögliche Überschneidungen zwischen einzelnen Parametern zu berücksichtigen.

Abstract

Listener-specific head-related transfer functions (HRTFs) are crucial when it comes to personalised binaural virtual audio. However, there is no method yet enabling an acquisition of personalised HRTFs without the need for highly specialised facilities and equipment. Photogrammetric reconstruction can alleviate the problem of geometry acquisition such that personal HRTFs can be numerically calculated, but the necessary geometric accuracy of the pinnae can be achieved only with great effort. In this thesis, shape deformation is proposed in order to improve the quality of photogrammetrically reconstructed pinnae. The principle of shape deformation is used to deform a high-quality but non-individual template pinna such that it matches a low-quality but individual target pinna. Two proposals are described, namely the non-parametric non-rigid registration and a parametric pinna model. The performance of these proposals was evaluated on both high-quality and low-quality meshes in the geometric, acoustic, and psychoacoustic domain. In the psychoacoustic domain, both an auditory model was used and a sound-localisation experiment was conducted. Results suggest that the parametric pinna model is more likely to achieve lower localisation errors than the selected non-rigid registration algorithm. However, further investigations of the parametric pinna model are required such as the examination of the parameter space and potential overlap between parameters.

Contents

Chapter 1

Introduction

Millions of people use headphones on a daily basis – to listen to music, watch movies, or simply to make a telephone call. Signals played back over headphones are typically perceived within the head, instead of from the actual sound-source position, that is, outside of the head. This phenomenon is commonly accepted, but mostly because of the lack of alternatives. The growing use of augmented reality (AR) and virtual reality (VR) in various systems emphasizes the necessity of using realistic sound reproduction over headphones. This motivates the ultimate goal of this thesis: Investigating new methods to improve realistic headphone playback outside of controlled laboratory conditions.

A realistic headphone playback of sound has been made possible for some time under controlled laboratory conditions. Personalised binaural sounds can be generated enabling a person to perceive the sounds in a natural way – outside their head. This is what is called binaural virtual acoustics (BVA). A major contribution to the personalisation quality of a BVA system comes from the pinna shape, because listener-specific binaural signals are acoustically created by the filtering of the sound with the torso, head, and pinnae, with the latter being strongly listener-specific. This filtering is described as head-related transfer function (HRTF). A binaural pair of HRTF describes the filtering of a sound source considering its position relative to the listener.

Listener-specific HRTFs are essential for BVA systems to replicate natural sound sources convincingly. Unfortunately, classic HRTF acquisition requires highly sophisticated laboratory equipment: Multiple loudspeakers (with a mostly flat frequency response) are placed in an sound-proof chamber, miniature microphones (with a mostly flat frequency response) are placed in the entrances of both ear canals of the listener, and impulse responses from hundreds of different sound-source directions are measured acoustically. This HRTF acquisition method is cumbersome, and thus, one of the reasons as why personalised BVA systems are unavailable to consumers.

There are, however, alternatives to the acoustical measurement of HRTFs: The acoustic influence of the geometry of a listener (torso, head, both pinnae) on the sound field arriving at the entrance of the ear canal can be simulated using a computer. With this method, HRTFs can also be *numerically calculated*. Because it is not necessary to place miniature microphones in ear canals of human listeners, this method is classified as non-invasive.

The numerical calculation of HRTFs – details of which are further elaborated in Sec. 1.3 – is possible given an adequate description of listener-specific head and pinnae geometries. However, one obstacle to access high-quality numericall calculated HRTFs is the 3D geometry acquisition, especially in those of the pinnae, which are most important. Out of all the possible geometry acquisition methods, photogrammetrical reconstruction (PR) is the focus in this thesis, because it is considered easy to access for a wide range of people. In this approach, many photos of listener-specific pinnae are made, from which a 3D pinna model is calculated and represented as point cloud and mesh. PR is most likely to be integrated in a user-friendly system utilising HRTF calculation, because there are basically no high-level requirements or equipment: photos can be made with a customary camera, and the construktion of the point cloud or mesh as well as the HRTF calculation can be executed on a common computer. Section 1.4 provides a detailed overview of state-of-the-art mesh acquisition by means of photogrammetry.

User-friendly PR has been integrated in applications using personalised HRTF. However, the results do not sound convincing, because PR typically results in pinna meshes containing errors, such as missing points (holes) in occluded areas, wrongly assigned points not actually part of the geometry of interest (noise and outliers), e.g., vellus hair or points outside the pinna geometry. Listener-specific HRTFs require a high accuracy of the 3D representation, and a manual sculpting of the pinna meshes requires a time-intense manual postprocessing. Reasons for the faulty pinna meshes may be the complex folds of the pinna which are partially occluded and impossible to be captured using a camera. To my knowledge, there are no reliable state-of-the-art methods for improving the PR of pinnae besides manually sculpting the shapes using 3D computer graphics software. Additionally, it is unclear how potential improvements to meshes, HRTFs, and, lastly, the realism of binaural signals affect AR/VR systems.

In this thesis, several systematic investigations were conducted in order to answer the following research question:

> How can non-rigid deformation methods improve photogrammetric reconstructions for HRTF personalisation?

In order to answer this research question, two non-rigid deformation methods are proposed: Non-rigid registration and a parametric pinna model (see Sec. 1.5). Both proposals use the same principle: The deformation of a high-quality (but generic) pinna template using transformation $\mathcal{T}$ such that it matches a PR-based (but listener-specific) pinna target. *Registration* is the process of deforming a template such that it matches a target. Figure 1.1 shows this principle. The goal of pinna registration was the preservation, or replication, of listener-specific details in the pinna shape, while maintaining the high quality of the template. If these criteria are met, the resulting mesh is ideal for the numerical calculation of personalised HRTFs.

The investigations in this thesis were performed by systematically varying relevant parameters with the focus on the contribution of various pinna regions encoding directional information in HRTFs. Evaluations on the registration success were performed in three different domains: the geometric domain, the acoustic domain, and the psychoacoustic

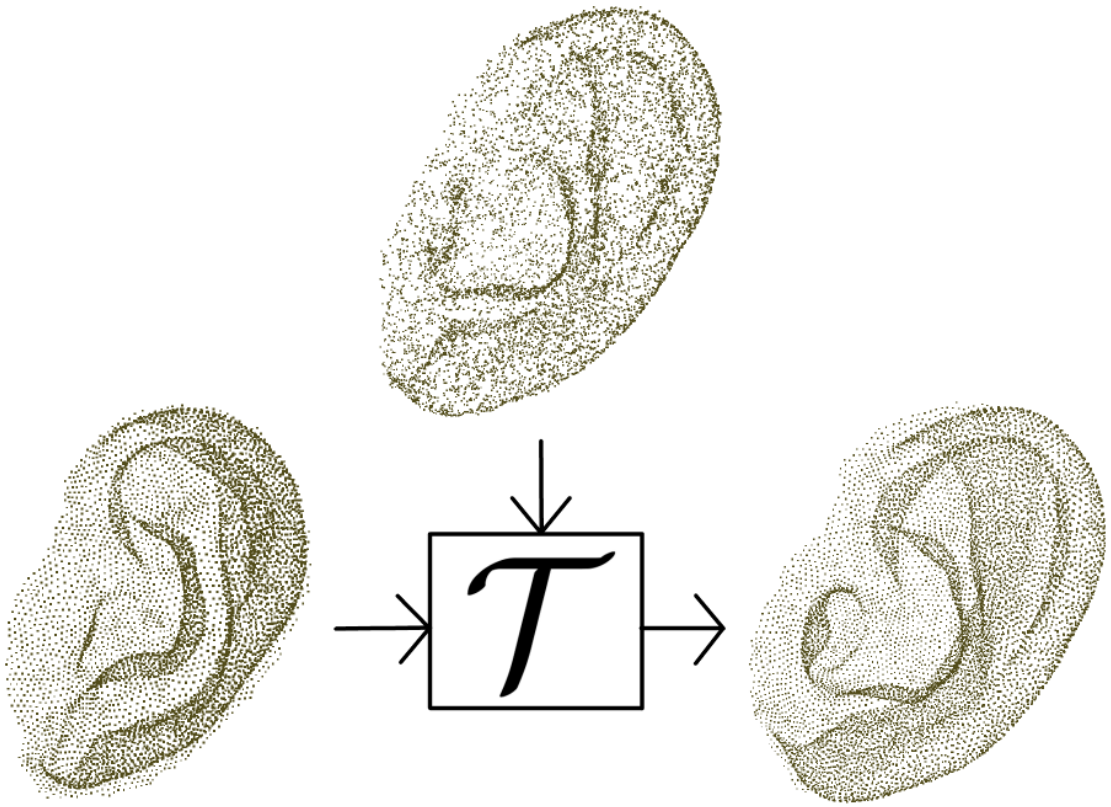


Figure 1.1: Workflow proposal of the dissertation. Deform a high-quality non-individual template ear using a low-quality individual target obtained via photogrammetry such that the output is a high-quality individual ear. $\mathcal{T}$ describes a transformation performed by the two proposals, non-rigid registration (NRR) algorithms and the parametric pinna model (PPM).

domain using both an auditory model and a behavioural experiment for selected conditions. Insight in the applicability of NRR algorithms and the PPM is an essential output of this thesis forming an essential step towards realistic sound reproduction in personalised AR/VR systems.

Section 1.1 gives a detailed overview of the fundamentals of HRTFs, and Sec. 1.2 elaborates on the discretisaion of HRTFs.

1.1 Head-related transfer functions

The sound produced by a sound source arriving at the eardrums is filtered in time, frequency, and distance by the listener's anatomy. Humans are able to locate sounds in the physical space around them, i.e., they can map the filtered sounds arriving at the eardrums to direction and distance of perceived auditory events [1]. This acoustic filtering can be summarised as HRTFs. Humans use HRTFs to build an internal model of their natural

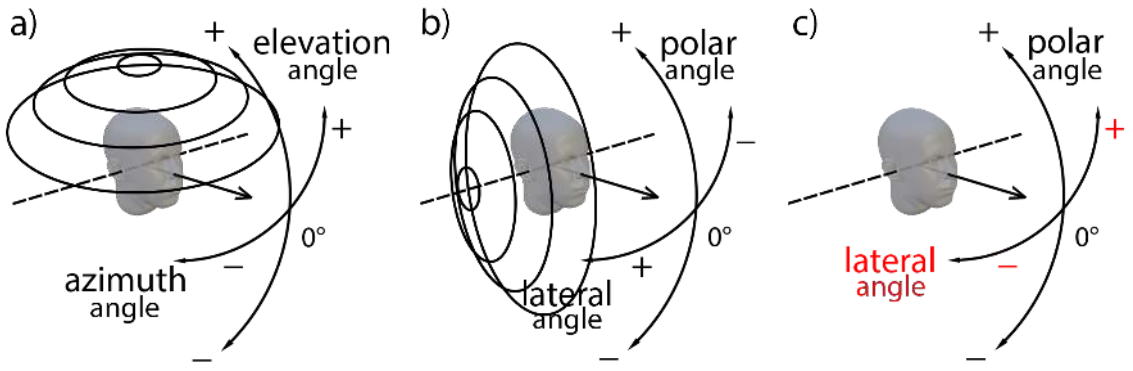


Figure 1.2: Coordinate systems typically used in the HRTF acquisition and representation. The dashed line represents the interaural axis, and the arrow represents the viewing direction. **(a)** Spherical coordinate system with the azimuth and elevation angles. **(b)** Interaural-polar coordinate system with the lateral and polar angles obtained by rotation the poles of the spherical system [3]. **(c)** Modified interaural-polar coordinate system with the lateral and polar angles corresponding to the azimuth angle in the horizontal plane and the elevation angle in the median plane [4].

environment and with it span an auditory space [2].

Despite a spherical coordinate system being ideal to describe positions of sound sources from the listener's perspective in space, it makes much more sense to shift the poles on the top and bottom of the sphere to the left and right instead, where the ears are. Figure 1.2 summarises the transition from the spherical over the interaural-polar to the modified interaural-polar coordinate system, in which the lateral angle corresponds to the azimuth angle in the horizontal plane and the polar angle corresponds to the elevation angle in the sagittal planes. The polar angle describes the position of a sound source along a sagittal plane, i.e., planes that are parallel to the median plane. The lateral angle describes the position of a sound source along the horizontal plane.

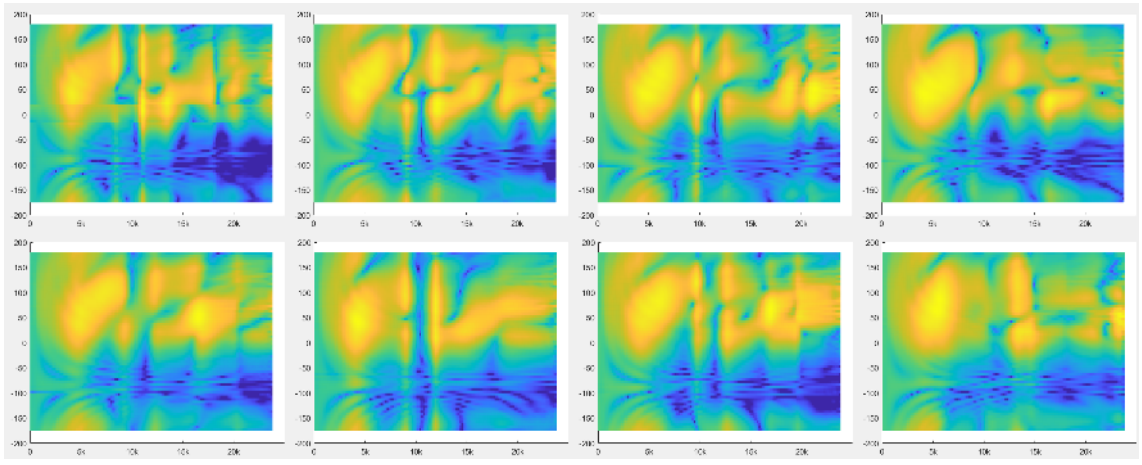
Depending on the usecase, it can be advantageous to use one of the three coordinate system. Various software in research, e.g., the SOFA toolbox [5, 6], binaural renderers such as [7, 8], or the auditory modeling toolbox (AMT) [9, 10], use these coordinate system to visualise HRTFs, render sounds with HRTFs, or use HRTFs in auditory modelling.

Various parts of the human body interact with the incoming sound wave, each affecting HRTFs in different frequency regions. HRTFs can be separated into binaural and monaural cues that each become dominant in different frequency ranges [11]. For lower frequencies up to the equivalent wavelength of the listener's interaural distance, i.e., on average 14.49 cm [12], the dominant binaural cue is the interaural time difference (ITD). The ITD describes the time difference between ipsilateral and contralateral ear and can vary between listeners of up to ± 80 μs due to the head width variation [12]. For frequencies higher than the equivalent wavelength of the interaural distance, the dominant binaural cue is the interaural level difference (ILD). The ILD describes the level difference between the two ears due to shadowing effects of torso and head [13, 14, 15, 16, 17]. Using these two cues, humans already can distinguish sound sources in the horizontal plane [18].

However, there are specific instances of sound source positions which result in the same ITD and ILD, these positions lie on the rim of the so-called *cone of confusion* [19]. The cone of confusion corresponds to a circle on the sagittal planes, see Fig. 1.2b. When using the modified interaural-polar coordinate system, sound sources with a constant lateral angle and various polar angles have (approximately) the same ITD And ILD and their monaural spectra can be compared independent of binaural cues.

The confusion could trivially be resolved dynamically by turning the head such that ITDs and ILDs get nonzero. However, in static conditions, humans are able to resolve the confusion using another – a *monaural* – cue: The pinnae are complex shaped geometries, designed to not only focus the incoming sound into the ear canals, but also spectrally filter it [20, 21, 22]. Additionally to resolving front-back and up-down confusions, these monaural spectral features have been identified as the main contributor to distance perception and vertical localisation. The state-of-the-art understanding is that these monaural spectral cues are responsible for externalisation of sound sources, i.e., the perception of sound sources in their natural position outside of the head [23]. Monaural cues become dominant for frequencies above approximately 4 kHz, for resolving the cone of confusion, and for the perception of elevated sound sources [24, 23, 14, 25, 26, 27]. The pinnae are unique for every person, and can yield inter-individual variation in the HRTF of up to 20 dB [11], thus HRTFs are *listener-specific*, or, *individual* [19, 28, 29, 24, 23, 14, 25, 26, 27, 30, 31]. As an example, Fig. 1.3 shows individual ear shapes and individual HRTFs.

(a)



(b)

Figure 1.3: Pinnae and, thus, HRTFs vary drastically among listeners. **(a)** Pinnae of eight listeners and **(b)** corresponding HRTFs in the median plane. Abscissa shows increasing frequency from left to right (linear scale), ordinate shows the lateral angle from bottom to top between -180 and 180°. Each horizontal line represents an HRTF from a specific direction on the horizontal plane. Colour shows normalised magnitude between -50 and 0 dB.

Listener-specific HRTFs are essential in order to achieve good localisation performance, because HRTFs heavily depend on the individual geometry of a listener [31]. While the

head geometry can be approximated using a sphere [32], the relationship between pinna geometry and HRTF is far more complex and currently subject to research, e.g., [16]. Already small changes of the pinna geometry can result in significant changes in the HRTF. Because of their concave shape, Cavum and Cymba conchae play an important role for localisation, and a high-quality geometry acquisition of these areas are out of the question [17].

HRTFs can be obtained currently with three different approaches: acoustical measurement, personalisation, and numerical calculation [33]. The gold standard in state-of-the-art HRTF acquisition is the acoustical measurement. Traditionally, the acoustical measurement is designed as recording of the impulse responses between source and receiver in an (semi-)anechoic chamber. Each of these impulse responses describe the transmission path from a sound source in a specific position to the listener's ears [11, 34]. Li and Peissig have recently published an extensive overview of various state-of-the-art acoustical measurement techniques [35]. Acoustic HRTFs measurements typically are long procedures, because of the high number of necessary measurement directions (in the range of a few hundred different directions). Using highly sophisticatd equipment, HRTFs can be acoustically measured for *one* direction in about 5 to 20 minutes [36]. However, the acquisition of *multiple* distances requires a tremendous effort [37] and is impractical for commercial usage. For distances of more than approximately 1 m, monaural cues in HRTFs change marginally. In this thesis, similar to typical scenarios when HRTFs are used in applications, only farfield HRTFs are considered.

A second approach for HRTF acuiqisition is *personalisation*, which can be performed in an objective or subjective way, both describing the altering of an existing geometric shape or HRTF. Objective personalisation deals with measuring anthropometric data of a listener and then adapt either a geometric model of ears (and, sometimes, head) [38, 39, 40, 41, 42, 43] or a *non-individual*, sometimes a dummy-head, HRTF to the listener-specific data [31, 44, 45, 46, 47, 48, 42, 49, 50, 51, 52, 47]. Also, a lot of neural network approaches have been applied to solve the problem in both geometric and spectral domain [53, 54, 55, 56, 57, 58, 59, 60, 61, 62]. Subjective personalisation deals with confronting listeners with various HRTFs and selecting the best such that some form of error is minimised given feedback from the listener [28, 63]. This error function is typically a localisation error, i.e., the difference between target (rendered sound source position) and response (perceived sound source position). Objective and subjective personalisation approaches can also be combined, e.g., anthropometric data of a listener can be used to frequency-warp the HRTFs prior to a subjective personalisation [64, 65, 66]. Additionally, psychoacoustic models can help in finding a close-to-individual HRTF, or clustering a set of HRTFs in order to study the spectral nuances between similar HRTFs and further reduce the personalisation procedure duration [67, 68, 63, 69]. A lot of work has been done to find a link between the listener-specific morphology and their corresponding HRTFs [17, 70, 71, 72, 73, 74, 46]. However, this link is not fully understood yet, likely due to the high dimensionality of the problem at hand.

A third HRTF acuiqisition approach, numerical calculation, simulates the acoustical measurement using a sound source grid and a three-dimensional (3D) representation of a listener's geometry. In previous work, I summarised several approaches, i.e., the finite el-

ement method (FEM), the finite discrete-time domain method (FDTD), and the boundary element method (BEM) [75]. In this thesis, I cover the BEM only, as it is the method I am using for the numerical calculation of HRTFs [76]. Three advantages of the numerical calculation are the fact that the listener does not have to travel to a facility with an (semi-)anechoic room and special equipment to get their HRTFs measured, that the listener does not move or fatigues during the measurement, and that the inter-subject levels of HRTFs are in the same amplitude range compared to acoustically measured ones [77]. However, despite previous research showing no perceptual differences by means of localisation errors between acoustically measured and numerically calculated HRTFs [78], the necessary quality of the 3D representation of the listener's geometry can only be achieved when using expensive medical imaging (CT of a cast [79], MRI [40], etc.).

HRTFs are continuous functions in time, frequency, and space, but in order to actually use them in applications, discretising HRTFs is inevitable. The next section focuses on the parts of HRTFs that can be omitted by exploiting physical, theoretical, and perceptual criteria.

1.2 Discretisation of HRTFs

A listener-specific HRTF given a sound-source position $\mathbf{x}$ (encoding distance *and* direction) can, in signal processing terms, be described as

$$H(\mathbf{x}, \mathbf{s}, f) = \frac{p(\mathbf{x}, \mathbf{s}, f)}{p_0(0, f)} \tag{1.1}$$

where H describes the HRTF of one ear, $\mathbf{s}$ describes the listener-specific geometry, f describes the frequency, p describes the sound pressure at the ear canal entrance, and p_0 describes the reference sound pressure, i.e., the pressure in the middle of the interaural axis (inside the head) with the listener being absent in the sound field. An HRTF is typically acquired for both ears in a spatially discretised way. Several properties of auditory perception in humans can be exploited in the discretisaion of HRTFs. First, the *frequency range* can be limited not only to the boundaries of human hearing, i.e., 20 Hz - 20 kHz, but also to between 90 Hz and 16 kHz. The lower boundary can be introduced because previous investigations showed that frequencies lower than 90 Hz do not contribute to sound lateralisation, i.e., left-right localisation within the head [80]. The upper boundary can be introduced because previous investigations showed that frequencies higher than 16 kHz do not contribute to sound localisation outside the head [1]. Second, previous investigations have shown that the *dynamic range* is up to 50 dB [11]. Figure 1.4 shows the magnitude spectra of two listeners' HRTFs in the defined frequency range. For each listener, the left column shows HRTFs of the left ear, and the right column shows HRTFs of the right ear; The top row shows HRTFs in the median plane, and the bottom row shows HRTFs in the horizontal plane. Note the high variation in level between front/back and left/right for frequencies above 4 kHz for each listener and the high variation between the two listeners.

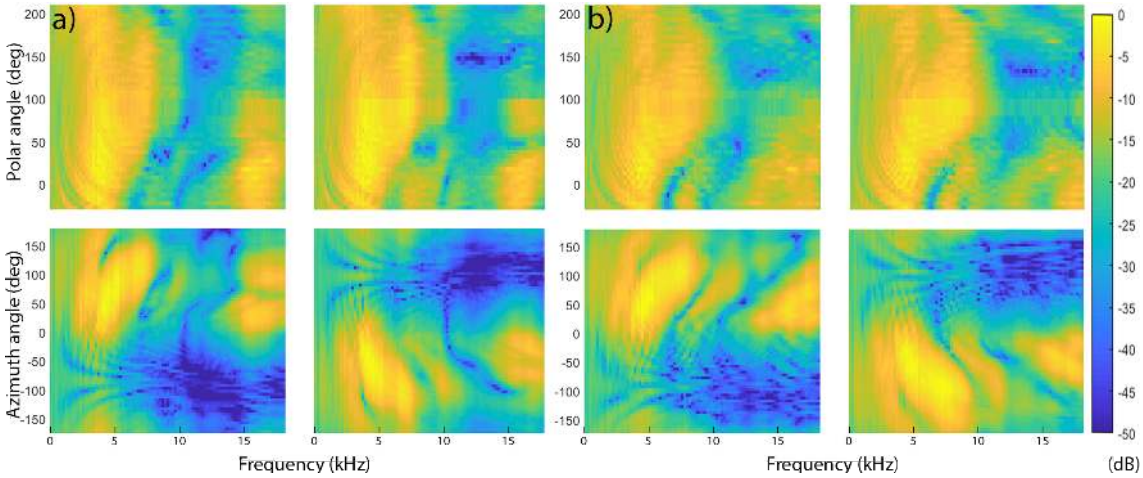


Figure 1.4: Normalised HRTF magnitude spectra for both ears of two example listeners of the ARI database [4] (**a)** NH236 and **b)** NH257). Columns show left and right ear. Top row shows HRTFs in the median plane, bottom row shows HRTFs in the horizontal plane. The HRTF magnitude has been normalised to full scale.

Third, the *temporal duration* can be restricted to 5 ms [81]. Figure 1.5 shows head-related impulse responses (HRIRs) – the time-domain equivalent HRTFs – of the same listeners as Fig. 1.4 in the horizontal domain.

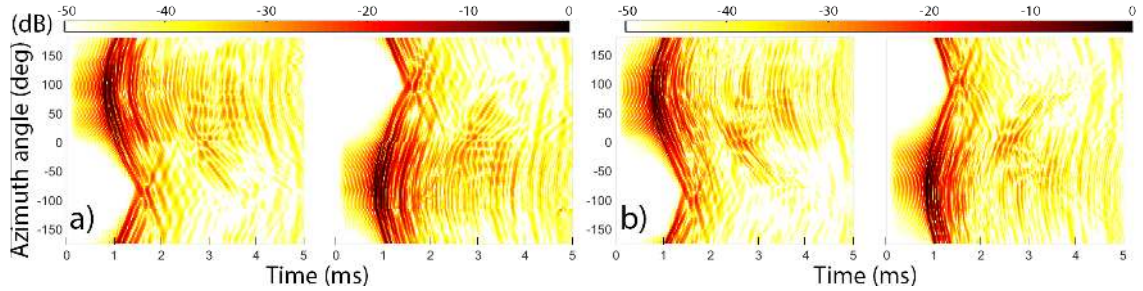


Figure 1.5: Normalised HRIR magnitudes in the horizontal plane for both ears of listeners **(a)** NH236 and **(b)** NH257. Note the decay within the first 5 ms.

Fourth, the *spatial resolution* can be drastically limited exploiting the ability of a listener to evaluate changes in binaural and monaural cues, i.e., the minimum audible angle (MAA) [82]. The MAA depends the type of signal and the sound-source position. Conclusively, broadband signals can be localised with the highest precision of 1-4° in front and behind the listener, and 10-20° for lateral or elevated positions [1, 83, 84, 44, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94]. This thesis only covers static sound-source localisation, where neither the listener nor the sound source move.

1.3 Numerical calculation of HRTFs

Numerical calculation provides a feasible alternative to the acoustical measurement of HRTFs: Using numerical methods, the acoustic sound pressure is calculated at the ear canal entrance of a listener-specific mesh. These numerical methods are widely known and evaluated, thus, do not pose a challenge for practical applications. Already in 2001, the boundary element method (BEM) was used to numerically calculated HRTFs [95], albeit computationally limited to 8 kHz. By coupling the BEM with the fast-multipole method (FMM) [96], the calculation limit could be extended to higher frequency ranges relevant for spatial hearing [97]. Nowadays, HRTFs can be calculated within one hour, if parallel computation is used with Mesh2HRTF [76, 78, 98].

The listener's geometry can be stored in a non-parametric representations such as a discrete and finite set of vertices, further denoted as "point cloud", or a point cloud extended by the relationship between points, i.e., a surface, further denoted as "mesh". To this end, typical 3D dataset formats are used, e.g., OBJ, PLY, or STL. These format are widely used in computer graphics and thus easily accessible by many corresponding applications. It is advantageous if the faces of a mesh are triangular in order to calculate HRTFs using Mesh2HRTF, because the face normal is unambiguous. This is not the case for other polygons typically used in computer graphics, e.g., quadrilaterals.

The requirements on the mesh are restricted by computational and perceptual aspects. Computationally, the BEM typically requires the average edge length (AEL) between vertices to be at least a sixth of the smallest wavelength [99], which corresponds to an AEL of 3.5 mm for frequencies up to 16 kHz. However, from a perceptual point of view, this resolution does not render the pinna geometry sufficiently accurate, the AEL of the elements in the mesh needs to be to be around 1 mm [79]. Figure 1.6 shows the same pinna in all panels, represented by meshes with increasing AELs from left to right.

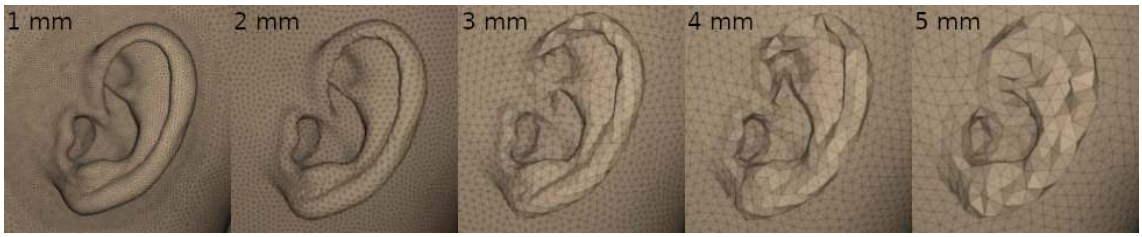


Figure 1.6: Pinna meshes demonstrating the effect of average edge length (AEL) [79]. From left to right: AEL of 1 mm, 2 mm, 3 mm, 4 mm, and 5 mm. Note how the representations of the helix and fossa triangularis degrade with increasing AEL.

In order to reduce the computational cost, the mesh may undergo a so-called *remeshing* [100], which resamples the head mesh dependent on the curvature, thus results in a higher resolution around the pinna regions. In this thesis, the `hrtf-mesh-grading` tool was used for remeshing[1] [101]. Larger anatomical parts like the head and torso can additionally be approximated by simple geometric shapes, e.g., a sphere for the head, a

[1] `https://sourceforge.net/projects/mesh2hrtf-tools/`

cylinder for the neck, and a rectangular cuboid or an ellipsoid representing the torso [32]. However, torso reflections and head shadowing can be applied as part of a low-cost post-processing step to an HRTF [4].

In order to calculate HRTFs with sufficient spectral accuracy, the number of elements needs to be in the range of several tens of thousands, which might be important for the requirements on the computational power. Such large numerical problems usually require large amount of memory being in the range of Gigabytes. Mesh2HRTF calculates HRTFs for each frequency independently, and thus, the calculations can be performed in parallel, especially using large computational clusters. This reduces the calculation time to one hour for HRTFs with 128 frequencies covering the full hearing range based on a mesh of several tens of thousands elements.

In general, the BEM underlying Mesh2HRTF describes a specific numerical method for solving the wave propagation using the Helmholtz equation. By selecting a specific set of ansatz functions, the computation boils down to solving a system of linear equations. The challenging part, however, is the dimension of the problem: For each face midpoint, the sound pressure is calculated. Within Mesh2HRTF, this problem is tackled by two methods: Separating the mesh into clusters of varying size, and exploiting the Helmholtz reciprocity theorem. The clustering is used to simplify the equation system by separating it into clusters. The Helmholtz repciprocity theorem states that the observed sound pressure is not changed when source and receiver positions are switched. Applied to HRTF calculations, this means that the virtual loudspeakers are placed at the blocked ear canals [78] and the real loudspeaker grid around the listener is switched to a grid of virtual microphones. Using these two methods to reduce the dimensionality problem, the computationally expensive part of the BEM, i.e., solving a linear system of equations to calculate the sound pressure at the surface, needs to be done once for each ear.

As a boundary condition, it assumes the 3D geometry to be sound hard, i.e., all surfaces having infinite acoustic impedance. With respect to localisation performance, this boundary condition yields to only small *perceptual* differences to acoustically measured HRTFs [78]. However, the impedance of various regions such as skin and hair may influence the direction-independent HRTF properties and cause changes in the perceived timbre [102, 103, 104, 105].

The BEM has been widely used to calculate HRTFs [97, 95, 106, 107, 108] analysing the computational process from various perspectives. When applied on an accurate and high-resolution representation of the pinna geometry, the BEM can yield similar results to the acoustic HRTF measurements by means of sound localisation performance [78, 76].

1.3.1 Comparison to acoustical measurements

Besides the practical advantages of numerical calculations which include travelling of the listener, listener fatigue and the need for elaborate equipment and facilities, the numerical calculation of HRTFs is briefly compared to its acoustically measured counterpart. This Section cover a brief evaluation whether the numerical calculation of HRTFs is trustworthy and sufficiently accurate [109, 110].

A traditional approach for such an evaluation lies in the comparison of the numerically calculated results with an analytically computated solution. No analytical solution for an HRTF has been found yet due to the complex shape of the pinna. However, several investigations have compared numerical and analytical computations for the pressure distribution on a sphere [98, 111, 110], which yielded a maximum error of 1 dB for the whole frequency range and all positions on the ipsilateral side. For the contralateral side, the magnitude error reached up to 6 dB because the contralateral side shows an average edge length of approximately 5 mm. However, the relatively large error is still negligible as the contribution of the contralateral ear to an HRTF is little [112].

Additionally to the comparison with an analytical solution, Ziegelwanger et al. compared the localisation performance in acoustically measured with numerically calculated HRTFs in a sound-localisation experiment. In this investigation, the AEL of the mesh in the pinna regions was systematically varied and the sound-localisation performance of three participants was compared to the acoustically measured reference. Specifically, AELs of meshes in the pinna region varied between 1, 2, 3, 4, and 5 mm. In this behavioural experiment, the authors found no significant differences between the acoustically measured HRTFs and the numerically calculated HRTFs with 1 mm average edge length [78].

Within the scope of this dissertation, I was involved in the maintenance of the Mesh2HRTF software package at the Acoustics Research Institute. While the BEM core was improved to be able to calculate multiple frequencies or frequency ranges in parallel [98, 113], I rewrote the Matlab script responsible for assembling the calculated sound pressure into an HRTF and formatting it as SOFA file [5, 6].

1.4 Mesh acquisition by means of photogrammetry

A pinna geometry can be captured via numerous approaches [114]: a laser scan [115], a structured-light scan [116], medical imaging techniques such as magnetic resonance imaging (MRI) [117, 118, 119] and computer tomography (CT) of a cast [79], or photogrammetric reconstruction [120]. Laser, MRI, and CT scans yield high-resolution meshes offering a small geometric error, but in turn need a special and often expensive equipment. In the medical imaging approaches, different downsides arise: Acquiring the pinnae geometry via MRI is not a trivial process, because the ears are flattened by the head support. This leads to two separate MRI measurements of each ear, having to be stitched to a head mesh. The anatomy is captured in "slices" that can be stitched together in the postprocessing. The CT captures the anatomy in a similar way, but due to the high radiation exposure, such scans are usually not done with human subjects but with (silicone) mouldings of the listener's ear. The overall procedure may take more time than an acoustic HRTF measurement and require the listener to either manufacture a moulding or meeting rather specific criteria for the scanning equipment (e.g., no tattoos, piercings, or implants). The structured-light scans are based on line-of-sight propagation and are able to measure short distances with an accuracy of up to 0.01 mm. The downside of line-of-sight propagation is that the manifolds of the pinnae are not easy to capture. As an easy-to-use alternative, recent advances have been made for more widely applicable approaches such

as photogrammetry [121, 38]. Photogrammetry is not only non-invasive, but it also can be done with a widely available equipment, e.g., a smartphone or digital camera, without having the listener to travel to a specialised facility. The overall process of the photogrammetrical approach works across applications as follows: a video or a set of photographs from different directions is made for each ear [120, 122, 123], then, the camera positions are estimated by analyzing mutual features across the photographs [124, 125, 126]. A 3D point cloud is constructed, and a 3D mesh is created by connecting the points in the cloud. This method requires no special hardware, because a common photocamera, e.g., from a smartphone, can be used. This method, however accessible, yields inaccurate pinna meshes which require extensive amount of manual postprocessing. Figure 1.7 shows a screenshot of a reconstruction software (Metashape version 1.6.5, Agisoft LLC) containing the reconstructed point cloud surrounded by the estimated positions of each photograph.

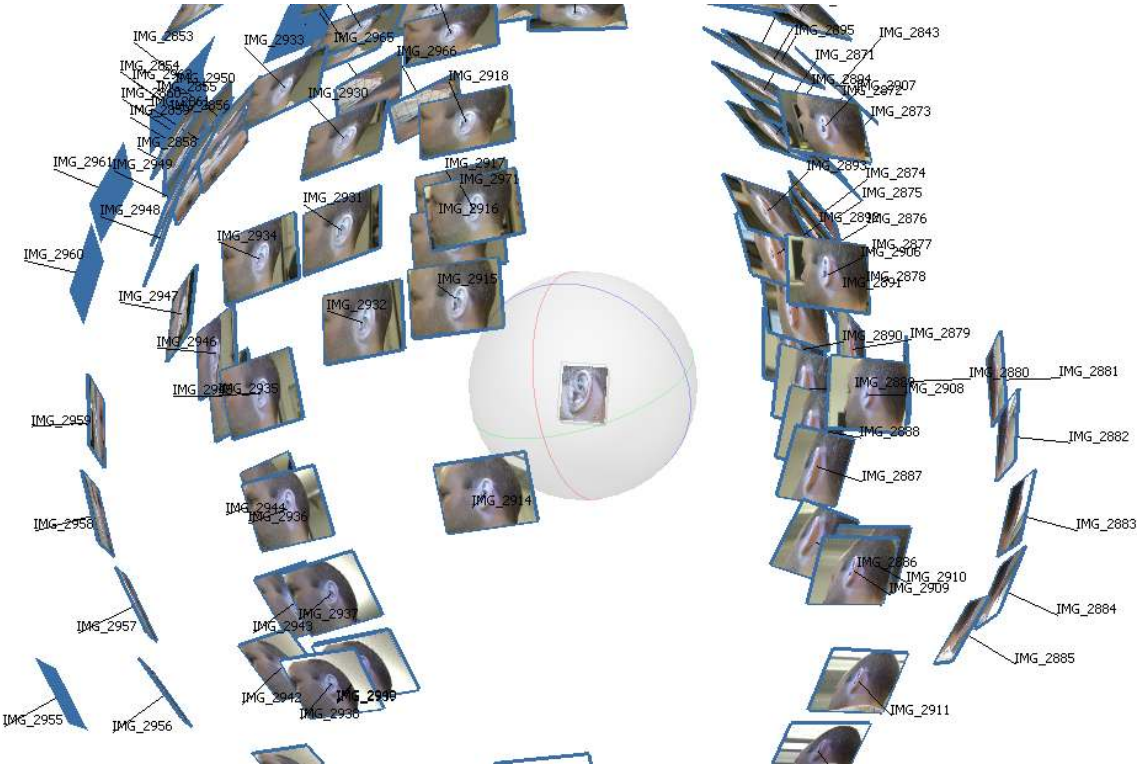


Figure 1.7: From features extracted from photos (shown as `IMG_` followed by a 4-digit number), the original position of the camera is estimated. With the triangulation, a point cloud of the pinna is estimated and its shape is reconstructed. Screenshot from Metashape version 1.6.5 (Agisoft LLC).

Note that the qualities of the reconstruction varies vastly among photogrammetry methods. Approaches using elaborate camera arrays or other professional equipment were discarded in this thesis, because the overall goal of the thesis was to improve photogrammetric reconstruction with minimal cost. Figure 1.8 shows the results of three different PR methods, highlighting the quality divergence.

Currently, manual corrections (e.g. smoothing to reduce noise, filling holes) are still

(a) (b) (c)

Figure 1.8: photogrammetrical reconstruction (PR) varies drastically in quality and depends on the effort of pre- and postprocessing of the point cloud. **(a)** High-quality PR using extensive pre- and postprocessing, **(b)** high-speed PR using a minimum amount of postprocessing, and **(c)** a PR result with no information on the methodology.

required to reach the high quality of the meshes required for accurate HRTF calculations, even for high-quality photogrammetry scans [123]. This is due to various problems, such as subsurface scattering of human skin [127, 128, 129], occlusion in various areas, direction-dependent shadows due to lighting conditions, the caption of vellus hair as part of the geometry, and (quasi-)evenly distributed skin colour.

In order to achieve a high-quality photogrammetry reconstruction without the use of niche or expensive equipment, pre- and post-processing steps can substantially contribute to more detail. The problems of subsurface scattering and (quasi-)evenly distributed human skin colour can be counteracted by applying a matte spray also used in theatre [123, 130]. This additionally adds a stochastic component to the pictures, which support the feature-matching procedure in the photogrammetry algorithm. However, the spray attaches to Vellus hair and has to be smudged after directly applying. Direction-dependent shadows can be avoided when using diffuse lighting or shooting outside when the weather is cloudy. Blurring of the 2D material can be avoided if no video is shot, but rather photographs from multiple angles [114]. The only thing that cannot be tackled is occlusion. The pinna is a complex geometry, and with optical 3D scanning methods such as photogrammetry – even high quality one – a few regions remain where no camera can point. This is where the photogrammetry algorithm either leaves holes or *interpolates*.

Once the mesh has been reconstructed, the occluded areas that have been left open or interpolated, have to be closed or sculpted manually, respectively. The overall surface of the mesh may be smoothed, but at the same time, the personal details of the pinna shape should remain present. This requires a cumbersome process of manual sculpting combined with smoothing, potentially using the photogrammetry source material as reference.

1.5 Shape deformation

This thesis proposes shape deformation as one solution approach to the problem of low quality in high-speed PRs. Two proposals are made, the first one covers non-rigid registration algorithms, with the motivation stemming from medical image processing, and the second one covers the parameterisation of the pinna geometry, with the motivation stemming from computer animation. Both proposals are thoroughly evaluated in this thesis in Chapters 3 and 4.

1.5.1 Proposal I: Non-rigid registration

NRR algorithms are methods to register one point cloud (template) to another point cloud (target) [131]. The registered point cloud (result) should be as close as possible to the target while maintaining the high quality (no noise, outliers, or holes) of the template. NRR is a generalisation of the rigid registration, in which only linear transformations, i.e., translation, rotation, scaling, are allowed. NRR, however, allows additional global changes, i.e., shearing and anisotropic scaling, and local changes, i.e., region-based deformation. Chapter 3 gives an overview of the literature research of state-of-the-art NRR algorithms. These methods play an important role in medical image processing, computer(-assisted) vision, or pattern recognition. One method has to be emphasised here: In coherent point drift (CPD), the template is interpreted as the means of a Gaussian mixture model (GMM), and the target is interpreted as observations from said GMM. In a temporal process, the movement of the means towards the observations is modelled in an alternating fashion between estimating the point correspondence and minimising a cost function while assuming coherence in the movement of proximal points. The velocity coherence allows a smooth transition between single iteration steps of the registration. CPD has been extended and developed further since its first publication in 2010. However, since the applicability of NRR algorithms on pinna geometries has not yet been investigated, several state-of-the-art algorithms were compared.

1.5.2 Proposal II: Parameteric pinna model

The second proposal approaches the deformation a template ear such that it matches a target ear by parameterising the pinna surface. This proposal aims at significantly reducing the dimensionality of the registration problem. In contrast to the first proposal, the parameterisation of the pinna geometry has been investigated thoroughly using various decomposition methods, such as the principal component analysis [40, 132] or active shape models [133, 40, 134, 132]. These previously suggested PPMs based on various decomposition methods have one common disadvantage: They depend on the dataset underlying the decomposition, and thus, these appraoches struggle with several – and potentially unknown – biases. Although other approaches use anthropometric parameters as basis for template deformation [135, 32, 136], they oftentimes face the problem that not all regions considered important for localisation are covered using conventional anthropometric pa-

rameters and thus, the resulting HRTFs yield significantly different sound-localisation performance as the acoustically measured ones.

The proposed PPM tries to unify the knowledge from previous PPMs but implementing them in a creative way, that is, motivated by computer animation. The PPM is then manually deformed to various target pinnae and its registration is evaluated in the geometric and behavioural domains.

Chapter 2

Perceptual relevance of outer ear regions

Evaluating HRTFs is a difficult process: The contribution from the head is typically described by broadband interaural cues, i.e., interaural time and level differences, which can be evaluated using binaural lateralisation models, e.g., [137]. The contribution from the pinna (monaural spectral cues) can be evaluated using a sound-source localisation model for sagittal planes [67]. In this model, sound-localisation errors from the evaluation HRTF set are compared to actual sound-localisation errors. The contribution of interaural spectral cues can be evaluated using externalisation models, e.g., [138, 139]). Non-spatial aspects such as, e.g., differences in timbre of HRTFs are often modelled as amplitude differences (e.g., spectral distortion [140]). However, state-of-the-art research showed that these models are limited in the description of timbre perception (see [49]).

Finally, auditory modelling can only provide a proof of concept, or a preselection of potential HRTF set candidates because not all aspects of spatial hearing are considered in auditory models. Thus, spatial aspects of HRTFs must be evaluated also in behavioural experiments [4]. While realistic scenarios consider controlled dynamic head movements in order to evaluate, e.g., changes in interaural time differences, static conditions (without dynamic head movements) are required for rigorously testing HRTFs. This provides a correct assessment of the quality of spectral cues in an HRTF in a worst-case scenario (see [141]).

Crucial for the success of pinna personalisation is whether the regions important for sound localisation are well-matched to the target. Several investigations covered this topic before, e.g., [12, 136, 71, 135, 52, 134], all resulting in various parameter sets describing a pinna and the importance of specific pinna regions for sound localisation. The pinna is a complex shape and has a variety of regions described with technical terms. Figure 2.1 shows the pinna shape including technical terms. Takemoto and Mokhtari have analysed resonance frequencies in the HRTF in [16, 142, 17] and concluded that the cavum conchae and cymba conchae are the major contributors to the resonances associated with vertical sound-localisation accuracy. Middlebrooks et al. concluded that the ear canal does not contribute to direction-dependent cues which justifies the HRTF being aqcuired

at the blocked ear canal [143]. Most of the studies that numerically calculated HRTFs covered specific frequencies only, either because of high computational costs at that time or because of the limited availability of high-resolution meshes of individual listeners.

In order to establish in which regions of a photogrammetricically reconstructed pinna the geometrical detail should increase in order to improve HRTF personalisation, the following investigations were conducted. First, the effect of isolated deformation of pinna surface regions was investigated. Second, the effect of including the ear canal transfer function was investigated by comparing HRTFs calculated at the blocked ear canal and the eardrum.

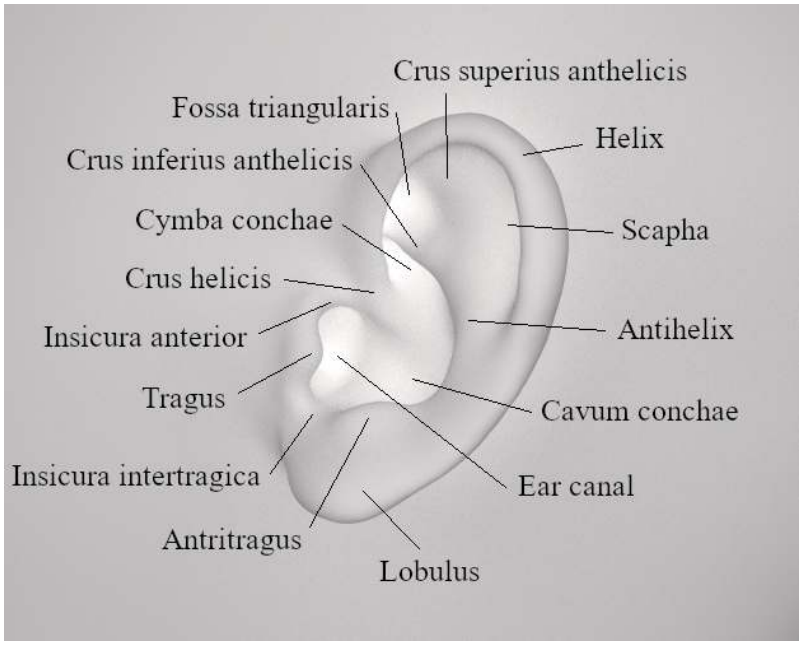


Figure 2.1: The human pinna including technical terms.

2.1 Effect of pinna surface regions

Previous research showed that not all regions of the pinna are equally contributing to an HRTF and whether anthropometric parameters could predict acoustic features [16, 144, 145, 65, 17] or even behavioural performance [135]. The consensus of previous works is that the concave regions – boiling down in literature to cavum conchae and cymba conchae – contribute the most to elevation perception. This aligns with an investigation in which participants got earmolds inserted that decreased the depth of the cavum conchae and cymba conchae, resulting in the collapse of elevation perception and adaptation to heavily altered monaural spectral cues [146, 147]. However, there are three aspects I did not find in the literature. First, other concave regions than the cavum and cymba conchae were not investigated yet – however, there is a lot of individual detail in other concave regions as well, e.g., the Fossa triangularis and the Scapha. Second, no previous investigation covered the *backside* of the pinna, but it is important in the light of photogrammetry

whether or not the backside of the pinna is necessary to be captured in precise resolution. And third, the depth of the cavum conchae and cymba conchae was *decreased* in previous studies, but there was no non-invasive way yet to *increase* depths of pinna regions.

2.1.1 Methods

In order to understand which regions of the pinna mesh are important to be captured in precise resolution, the perceptual relevance was investigated by changing a specific region in the mesh, calculating HRTFs, and evaluating the sound-localisation performance using an auditory model. Using only calculated HRTFs, it was practical to deform the ears as much as possible in a non-invasive way. For this investigation, ten publicly available meshes of the SYMARE databse were used [118], consisting of head (including pinnae) and torso meshes of 10 listeners. The head and pinnae scans without torsi provided enough geometrical information to calculate HRTFs because the effect of the torso is negligible for frequencies above 4 kHz. Several anatomic regions were deformed in a systematic way: The depths of the cavum conchae and cymba conchae were increased, the depths of the fossa triangularis, scapha, and helix, were increased and decreased, and the backside of the pinna was completely flattened. Figure 2.2 shows the geometric change in the pinna. Note that the geometric deformation could not be applied with the same extent for all the listeners, as various regions sometimes were already prominent in some way, and the deformities were applied the left ear only. In the acoustic domain, spectral differences have been calculated.

As error metric in order to evaluate the divergence between two point clouds in the geometric domain, the smallest distance ℓ was calculated between a point $\mathbf{x}_i$ in one point cloud $\mathbf{X}$ and every point $\mathbf{y}_j$ in the other point cloud $\mathbf{Y}$) by applying

$$\ell\left(\mathbf{x}_i, \mathbf{Y}\right) = \min_{\mathbf{y}_j \in \mathbf{Y}} \left\{ ||\mathbf{x}_i - \mathbf{y}_j||_2 \right\}, \tag{2.1}$$

where $||\cdot||_2$ denotes the Euclidean distance between two points. Note that this distance is not commutative, i.e., $\ell\left(\mathbf{x}_i, \mathbf{Y}\right) \neq \ell\left(\mathbf{y}_j, \mathbf{X}\right)$. In this thesis, this distance measure is further denoted as *smallest pointwise distance*. The smallest pointwise distance was computed and visualised on the mesh surfaces in order to emphasize which regions were influenced to which extent. Additionally, $median(\ell)$ was calculated.

Further, HRTFs have been calculated for the reference and deformed meshes for 1,550 positions on a sphere with 1.2 m radius using Mesh2HRTF v1.0 [76, 113]. The HRTFs were compared in the acoustical domain using the log-spectral difference between the HRTF calculated from the reference mesh and the HRTF calculated from the mesh with one region geometrically altered. The log-spectral difference S was calculated as follows:

$$S\left(H_1, H_2\right) = 20 \cdot log_{10}\left(\frac{|H_1|}{|H_2|}\right), \tag{2.2}$$

where $|H|$ denotes the absolute value of the complex pressure of a calculated HRTF. In

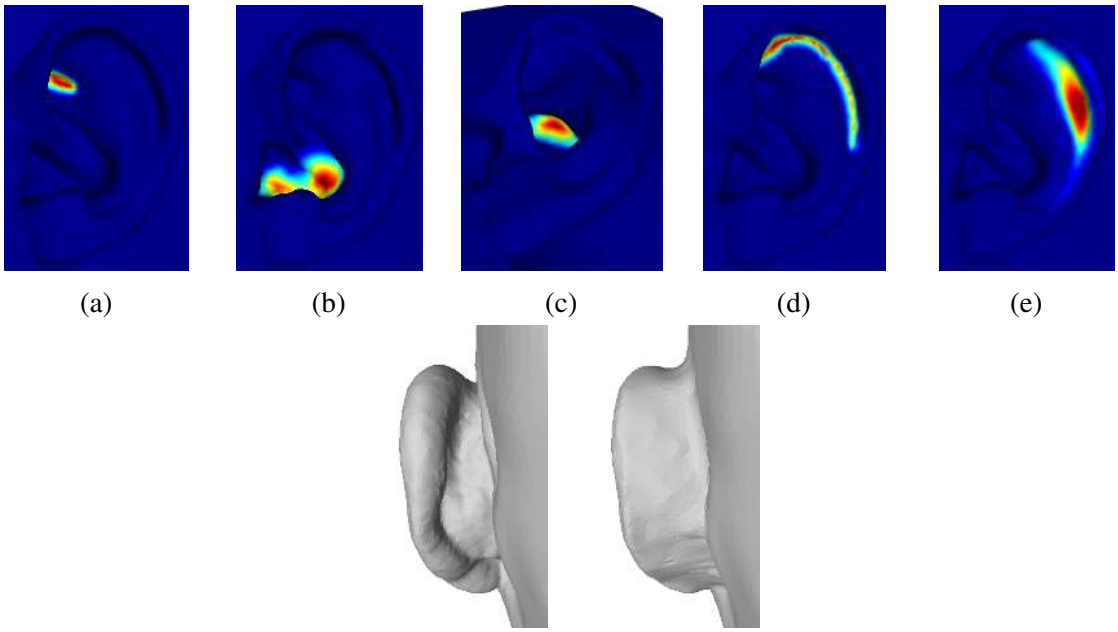


Figure 2.2: Examples of geometric deformation on the left pinnae of the two subjects 1 (**(a)**-**(e)**) and 5 (**(f)**-**(g)**). Coloured geometries show the smallest-distance distribution on the reference geometry from blue (no deformation) to red (maximum deformation). **(a)** Increased depth in fossa triangularis. **(b)** Increased depth in cavum conchae. **(c)** Increased depth in cymba conchae. **(d)** Increased depth in helix. **(e)** Increased depth in scapha. **(f)** Original backside, and **(g)** deformed backside.

this investigation, H_1 denotes the HRTF calculated from the geometrically altered mesh, and H_2 denotes the reference HRTF. The log-spectral difference can be calculated for each position and frequency bin. Typically the log-spectral difference is calculated between directional transfer functions (DTFs), because a direct component of +20 dB would yield a log-spectral difference of 20 dB without changing the spectral cues. However, in this investigation, the log-spectral difference was calculated between HRTFs, which had two reasons. The first reason was that the geometric modifications were negligibly small compared to the overall size of the mesh, hence the direct component of the signals would not change significantly. The second reason was that the geometric modifications might impact direction-independent cues and, hence, might have an influence on the spectrum of the sound-source signal.

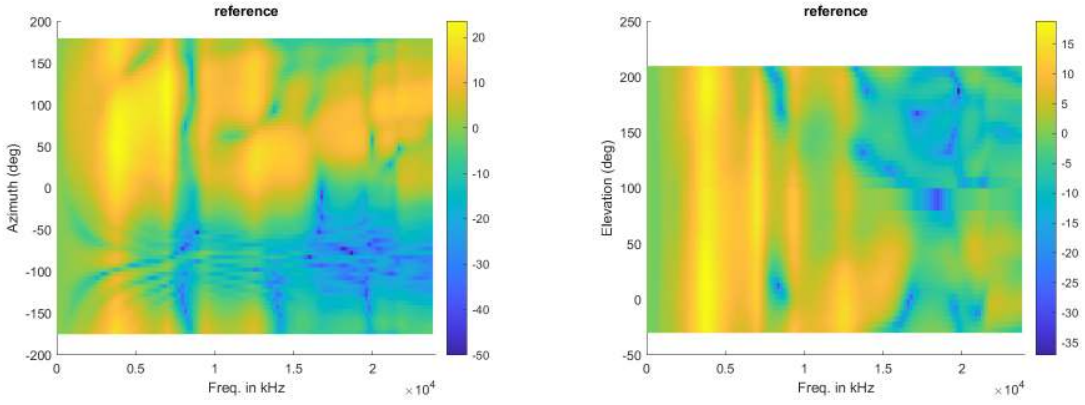
In the psychoacoustic domain, an auditory model that predicts sound-localisation performance in sagittal planes [67] was used to calculate sound-localisation errors, namely the quadrant error rate and local polar RMS error [44]. This model, implemented as the function `baumgartner2014` in the Auditory Modeling Toolbox (AMT) [1] [10], estimates how a listener would perform in a sound-localisation experiment given a template HRTF serving as the internal reference and being confronted with sounds convolved with a target HRTF.

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

Default parameters were used for the model, except a limited frequency range between 4 and 16 kHz. HRTFs were converted to DTFs and transformed into the time domain using the functions `SOFAhrtf2dtf` and `SOFAconvertConventions` from the SOFAtoolbox v2.0. Quadrant error rate and local polar RMS error were calculated from the estimated sound-localisation performance. The quadrant error rate describes the amount of stimuli localised in a wrong quadrant, summarising up-down and front-back confusions. The local polar RMS error, as implemented in the AMT v1.0, describes the root mean square (RMS) polar error for targets whose lateral angles are in the range of $\pm 30^\circ$ around the median plane, excluding the responses already categorised as quadrant error. These localisation errors were calculated because they reflect the major impact the pinna shape has on HRTFs, i.e., the median plane.

2.1.2 Results and Discussion

Figure 2.3 shows the magnitude spectra of the numerically calculated HRTFs of a listener's left ear in the horizontal and median plane. Figures 2.4 to 2.9 show the log-spectral differences between that reference and the left-ear modifications for one listener. In general it can be observed that even the geometric modifications in smaller regions, e.g., the fossa triangularis, have a significant impact on spectral cues from 700 Hz, affecting both direction-independent and direction-dependent cues. For all listeners, log-spectral differences in the range of ± 40 dB were found.



(a) (b)

Figure 2.3: HRTF magnitude spectra for the reference ear of listener 1 in the **(a)** horizontal and **(b)** median plane.

Figures 2.5 and 2.6 show the log-spectral differences for the cavum and cymba conchae, respectively. These regions were the ones modified by silicone molds in other studies, in which the depth was decreased [146, 147]. In this investigation, the depth was increased only, and the log-spectral differences were the highest for these two regions at around ± 40 dB. This indicates that both increasing and decreasing the depth of the relatively large concave regions drastically influences the spectral cues.

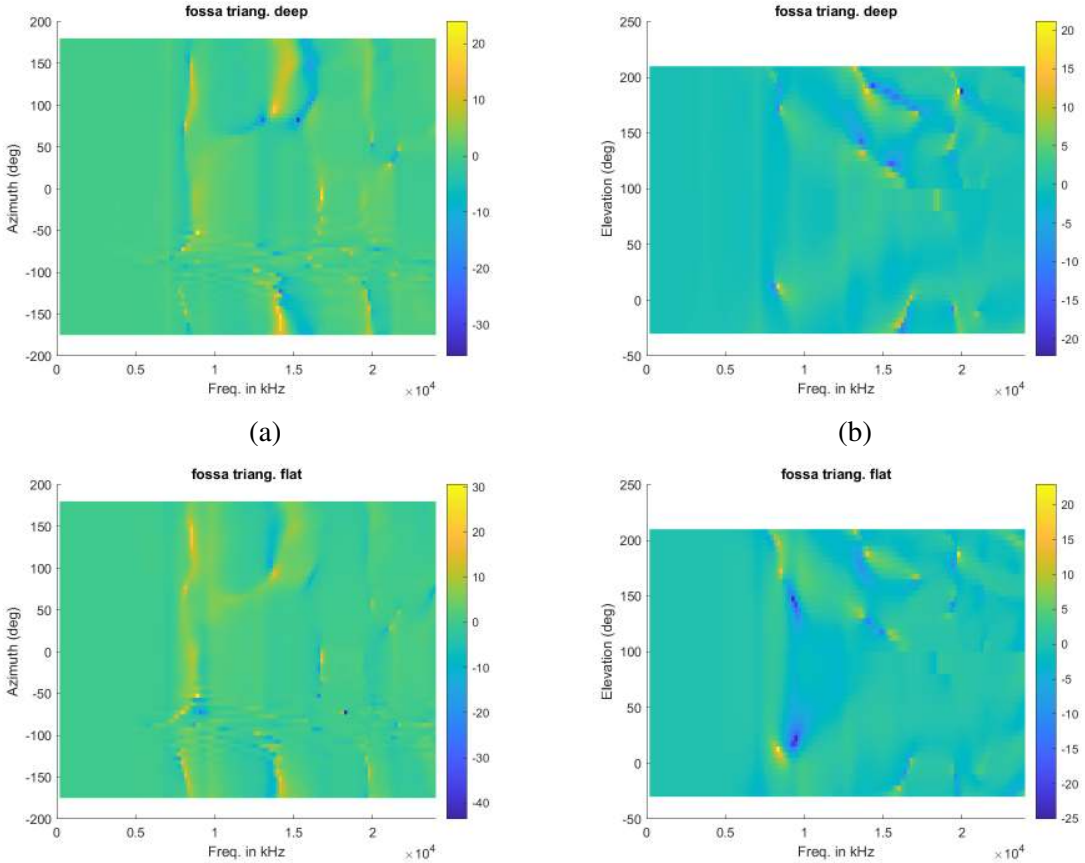


Figure 2.4: Log-spectral differences of HRTF magnitude spectra for fossa-triangularis modifications of listener 1 in the horizontal (left) and median (right) plane. Top row shows resulting differences when increasing the depth and bottom row shows resulting differences when decreasing the depth.

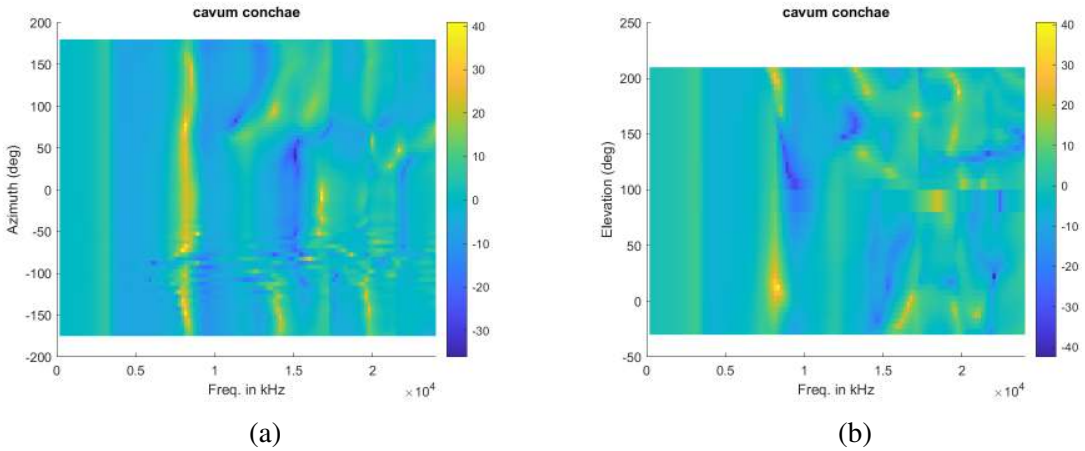
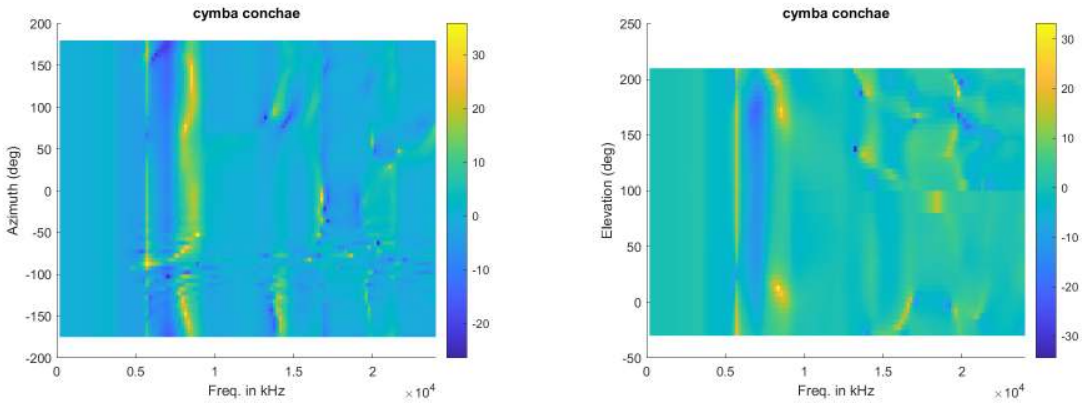
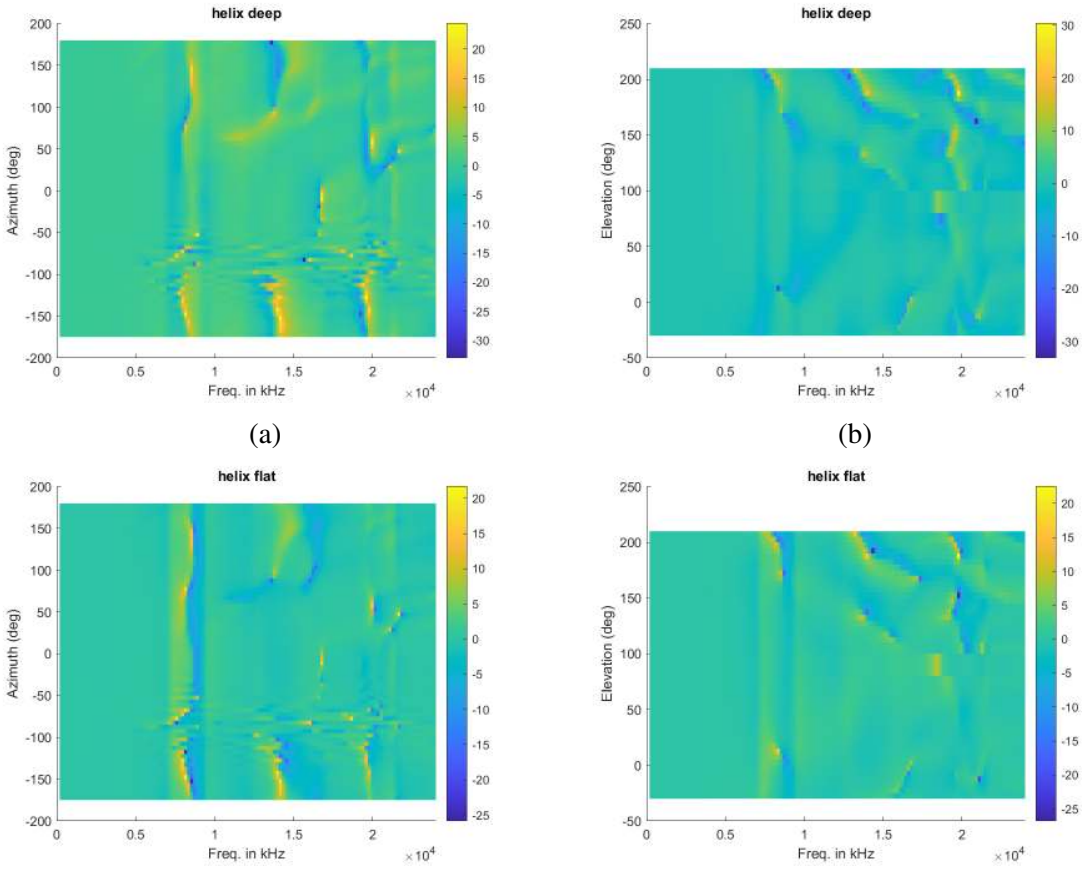


Figure 2.5: Log-spectral differences of HRTF magnitude spectra for cavum-conchae modifications of listener 1 in the **(a)** horizontal and **(b)** median plane.



(a) (b)

Figure 2.6: Log-spectral differences of HRTF magnitude spectra for cymba-conchae modifications of listener 1 in the **(a)** horizontal and **(b)** median plane.



(a) (b)

(c) (d)

Figure 2.7: Log-spectral differences of HRTF magnitude spectra for helix modifications of listener 1 in the horizontal (left) and median (right) plane. Top row shows resulting differences when increasing the depth and bottom row shows resulting differences when decreasing the depth.

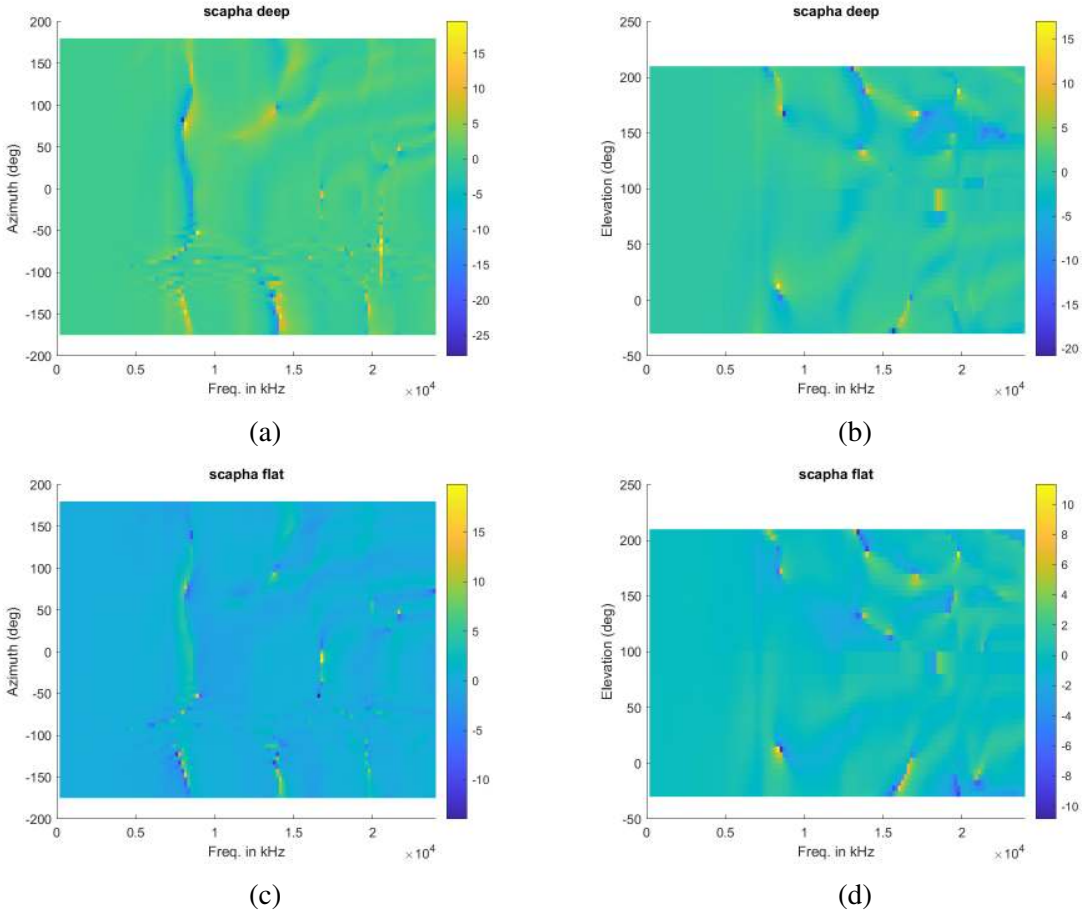


Figure 2.8: Log-spectral differences of HRTF magnitude spectra for scapha modifications of listener 1 in the horizontal (left) and median (right) plane. Top row shows resulting differences when increasing the depth and bottom row shows resulting differences when decreasing the depth.

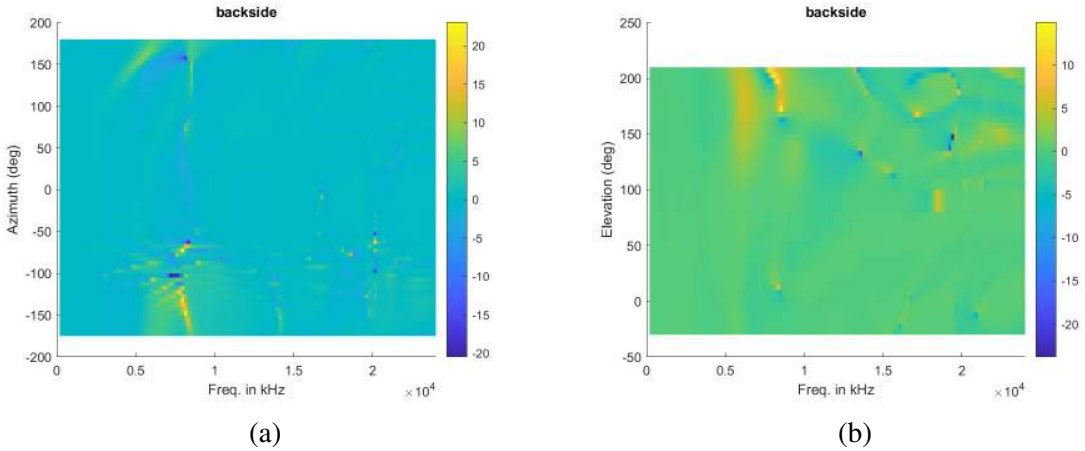
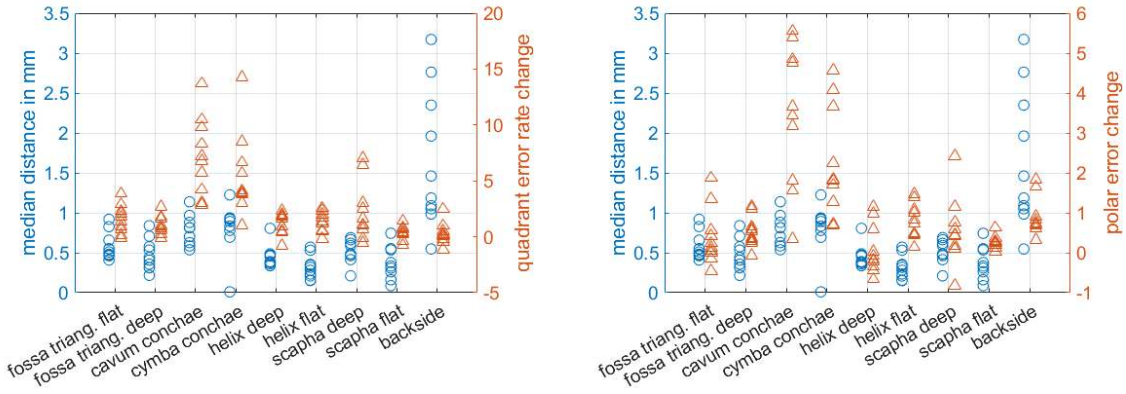


Figure 2.9: Log-spectral differences of HRTF magnitude spectra for decreasing the pinna backside of listener 1 in the horizontal (left) and median (right) plane.

Figure 2.10 shows $median(\ell)$ together with the changes in the sound-localisation errors.



(a) (b)

Figure 2.10: Median geometric error $median(\ell)$ in mm (left y axis, circles) and **(a)** quadrant error rate differences in percentage points and **(b)** polar error differences in deg (right y axis, triangles).

A $max(\ell)$ of up to 13 mm occured for increased depth of cavum conchae, cymba conchae, and the backside. For the first two regions in the concha, these geometrical deformations resulted in large global errors with an increase of the quadrant error rate of up to 14 %, which is in line with literature [16, 17, 147]. For the pinna backside, the geometrical deformations influenced the quadrant error rate in a negligible way (less than 5 %). Regarding the local polar RMS error, none of the geometrical changes had a perceptually relevant influence, as the largest change in the localisation error was less than 6 °, i.e., close to the MAA for static sound-source localisation in normal-hearing listeners in the vertical plane [84].

In conclusion, pinna deformations only affected the quadrant error rates in a perceptually relevant manner. Small increases of the depth of the cavum conchae, cymba conchae, and the scapha resulted in a non-negligible increase in the quadrant error rate for the majority of simulated listeners. Large pinna deformations in the backside had a negligible effect on the quadrant error rate. Changes in other geometric regions, such as, e.g., the fossa triangularis, did not affect the simulated sound-localisation errors much.

2.2 Effect of the ear canal

An HRTF is defined as the sound pressure at the ear canal entrance normalised by the sound pressure in the middle of the head [11]. However, this definition leaves out the transfer path between the ear canal entrance and the eardrum, which is typically simplified as a ¼ resonator [143]. Whether the HRTF changes drastically or whether the ear canal has any direction-dependent component has been studied using acoustic measurement [34, 143] using measurements at the blocked ear canal and 10 mm inside the ear canal. In the

previous study from Middlebrooks, HRTFs of 6 listeners at 356 sound-source positions sampled on a sphere around the listener were used.

The IHA database was released in 2021 containing meshes of ten human listeners including torso, head, and pinnae *including* the ear canal up to the eardrum [119]. This opened up the possibility to numerically calculate an HRTF at the eardrum without an invasive procedure and investigate the directional effects of the complete ear canal and extend Middlebrooks' study by ten virtual listeners and HRTFs with a higher spatial resolution.

2.2.1 Methods

The IHA database was used to calculate HRTFs at the blocked ear canal and at the eardrum of 10 listeners. Torsi were removed from the meshes, because only frequencies above 3 kHz were of interest and the torso geometry would add unnecessary computational cost. For the blocked ear canal condition, the ear canal was removed and blocked. The meshes were processed using Blender v3.1.4 [148], HRTFs were calculated using Mesh2HRTF v1.0 [76, 113][2] and the SOFAtoolbox v2.1 [6][3].

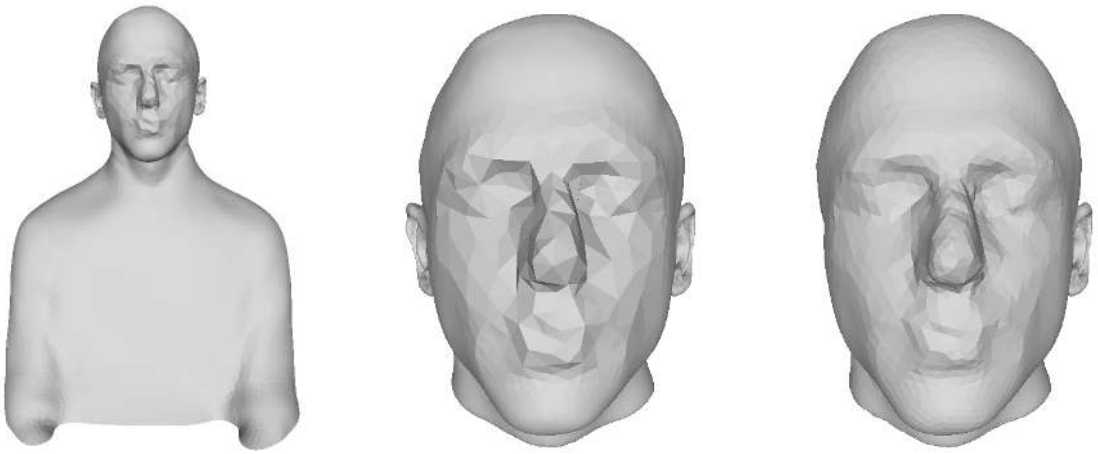
In the acoustic domain, HRTFs were compared using the log-spectral difference between the HRTFs calculated at the blocked ear canal and at the eardrum as in the previous investigation (see Eq. 2.2).

In the psychoacoustic domain, the quadrant error rate and the local polar RMS error were calculated using the auditory model simulating median-plane sound localisation from the AMT v1.6 [67, 10]. The DTFs at the eardrum were selected to be the internal template (reference), because they were considered cloest to "natural" ones.

2.2.2 Results and Discussion

Figure 2.11 shows the stages of mesh preprocessing: the full mesh as in the IHA database, the head mesh (torso removed manually), and the graded head mesh for accelerated HRTF calculation [101][4][5]. Figure 2.12 shows the virtual microphone placement in the two conditions blocked ear canal and at the eardrum.

[2] https://mesh2hrtf.org/
[3] https://github.com/sofacoustics/SOFAtoolbox
[4] https://github.com/cg-tub/hrtf_mesh_grading/
[5] https://mesh2hrtf-tools.sourceforge.io/



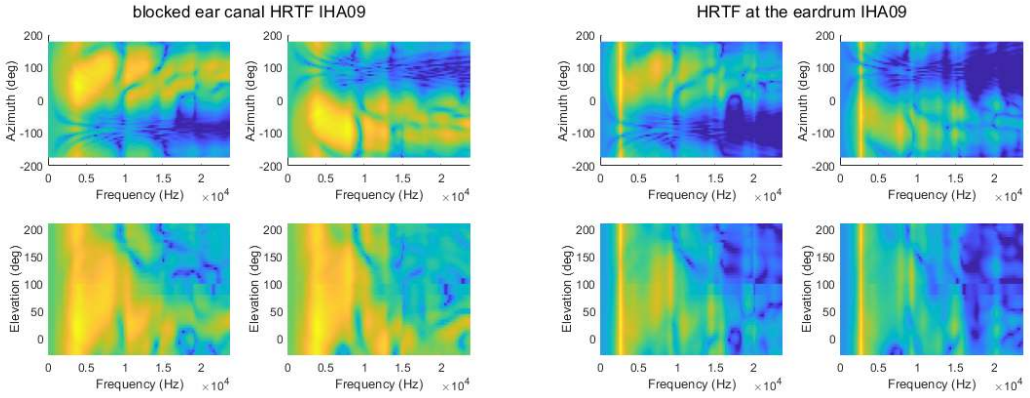
(a) (b) (c)

Figure 2.11: Geometry of listener IHA09: **(a)** raw including the torso, **(b)** without the torso, and **(c)** remeshed for the calculation of the left-ear HRTF.

(a) (b)

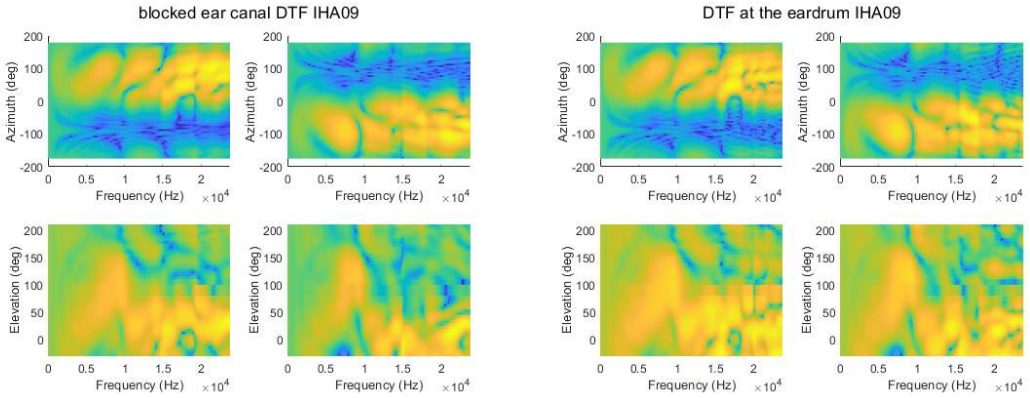
Figure 2.12: Geometry of listener IHA09 from inside the head and the microphone position (red dot). **(a)** Including the ear canal (microphone position in the estimated centre of the eardrum) and **(b)** with a blocked ear canal (microphone position in the estimated centre of the ear canal).

Figures 2.13 and 2.14 show the magnitude spectra of HRTFs and DTFs, respectively, calculated at the blocked ear canal and the eardrum for listener IHA09.



(a) (b)

Figure 2.13: Spectral magnitude of the HRTFs (in dB) calculated **(a)** at the blocked ear canal and **(b)** at the eardrum.



(a) (b)

Figure 2.14: Spectral magnitude of the DTFs (in dB) calculated **(a)** at the blocked ear canal and **(b)** at the eardrum.

Figure 2.15 shows the log-spectral differences for one listener (IHA09). The direction-independent resonances of the ear canal are the dominant differences for frequencies below 12 kHz. Between 15 and 20 kHz, there are various direction-dependent differences, but it is questionable as to how much these cues would influence localisation performance, if at all.

Figure 2.16 shows the sound-localisation errors from a simulated experiment. Indeed, the ear canal seems to contribute only a little in a direction-dependent way, which the model suggests to be negligible, in line with [143].

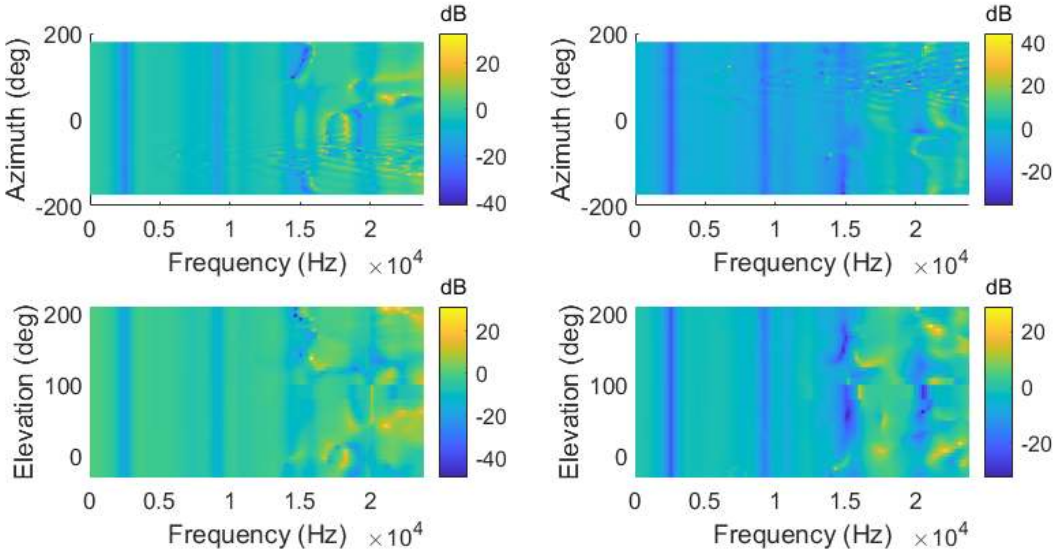


Figure 2.15: Log-spectral differences (in dB) between the HRTFs calculated at the eardrum and the HRTFs calculated at the blocked ear canal. Left and right column show theHRTFs for the left and right ear, respectively, for the horizontal plane (top row) and the median plane (bottom row).

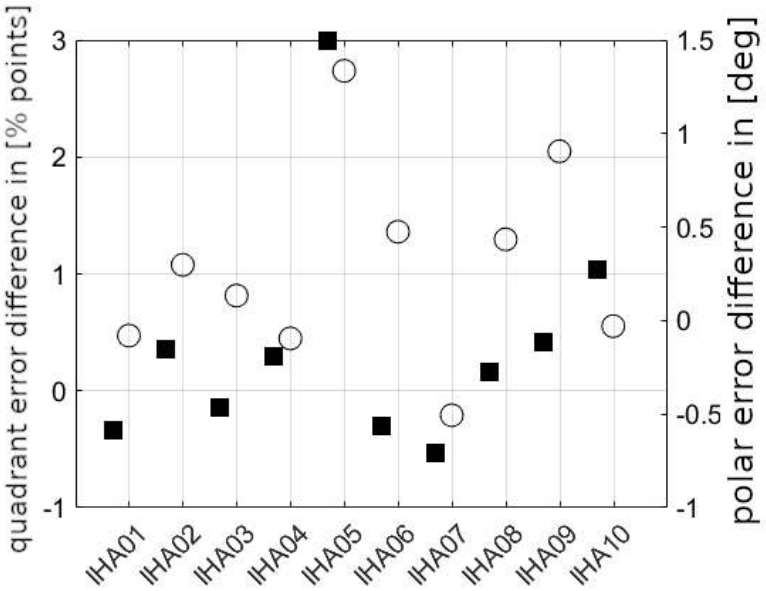


Figure 2.16: Quadrant error rate differences (square, filled) and polar error differences (circle, open) estimated for 10 listeners using the DTFs of the eardrum as the template in the auditory model.

It was apparent that the ear canal is responsible for a high direction-independent resonance at around 2.8 kHz and for an attenuation of higher frequencies, especially for the contralateral ear, leading to high spectral differences in HRTFs [110, 149]. However, when looking at the magnitude spectra of DTFs, there is negligible spectral difference

between the blocked ear canal and eardrum spectrum. Additionally, the model uses a Gammatone filterbank that imposes smaller weights on higher frequency regions. Thus, the model suggests negligible perceptual differences.

In conclusion, the ear canal does contribute to direction-dependent cues in a HRTF such that it would impact sound localisation in sagittal planes, despite its transfer function adding a substantial amount of energy to the $\lambda/4$ resonance.

2.3 Conclusions

In conclusion, not all regions of the pinna affect HRTFs equally [22], but the concave ones, i.e., cavum and cymba conchae, fossa triangularis, and scapha, can be considered most important for vertical sound localisation. Figure 2.17 shows the overview of technical terms of pinna regions from Fig. 2.1 again, this time highlighting the ones important for sound localisation [16, 144, 145, 135, 65, 17]. Basically, the convex curvatures of the pinnae contribute to focusing the incoming sound waves towards the entry of the ear canals, comparable to a satellite dish [75]. The cavum conchae, cymba conchae, scapha and fossa triangularis all contribute to sound-source localisation, because they showed a relatively large change in quadrant error rate for relatively small geometric changes, thus, these regions contribute strongly to resolve front-back confusions.

These findings indicate that the concave regions of a pinna should be captured with precise resolution. Increasing the depth of the scapha showed a higher quadrant error rate than decreasing the depth of the scapha. Even when completely flattened, the backside of the pinna did not seem to contribute substantially to sound localisation, which suggests that its geometry does not need to be captured precisely. Thus, in photogrammetric reconstruction, it may not be necessary to use hairbands in order to reveal the backside of the pinnae. Besides the known direction-independent $\lambda/4$-resonance of the ear canal, direction-dependent differences were found in the spectra of HRTFs. However, these differences are smaller in DTFs, and the auditory model predicting median-plane sound localisation suggested that these differences are negligible.

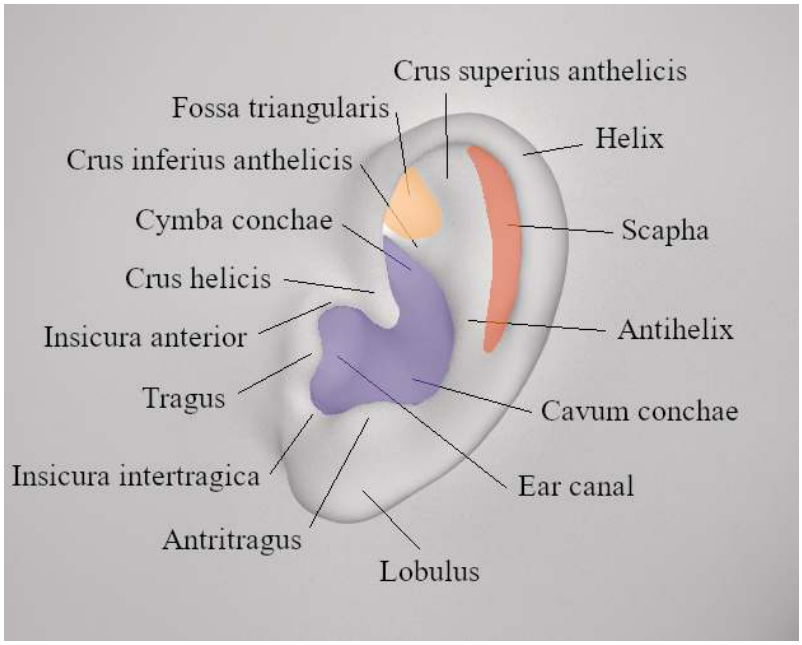


Figure 2.17: Various pinna regions including technical terms. Highlighted regions are considered important for localisation.

Chapter 3

Proposal I: Non-rigid registration

Non-rigid registration (NRR) describes the process of deforming a template point cloud such that it matches a target point cloud by means of a geometric fit. It extends affine registration, that is, translation, rotation, scaling, and shearing, by deforming local areas of a shape independent of its global features. An important assumption underlying the used NRR algorithms is motion coherence theory (MCT), which states that proximal points move together coherently [150]. MCT is an adequate assumption, because it is highly probable that points that are close to each other belong to the same part of a shape and, thus, tend to move in a similar way. Note that, this assumption is sensible if the template and target shape describe the same kind of object, i.e., in this case, both are human pinnae.

The motivation to incorporate NRR algorithms in auditory modelling stems from medical image processing. In medical image processing, such algorithms are often used when it comes to registering organ shapes – mostly complete volumina, not just surfaces – or bones [151]. There are also other fields of study where NRR algorithms play an important role, such as computer graphics [152], animation [153], and personalised shoe design [154]. This chapter investigates whether the application of various NRR algorithms to human ear shapes helps in HRTF personalisation. After several proof-of-concept investigations, however, an additional assumption is introduced, namely that that the template is a high-quality but non-individual pinna and the target is a low-quality but individual pinna. It was investigated whether NRR could preserve the high-quality of the non-individual template while adopting individual features from the low-quality target in the registration.

This chapter is structured as follows: Section 3.1 gives an overview of the literature research on NRR algorithms and explains the concept of four selected algorithms that fit specific criteria. The rest of this chapter deals with investigations in order to answer the main research question of how this proposal can help improving photogrammetric pinna reconstructions for HRTF personalisation.

The effect of high-quality photogrammetry on the pinna geometry was investigated in Sec. 3.2, systematically classifying artefacts appearing in pinna geometries due to PR. The effect of systematically distorting a target geometry using isolated artefacts was in-

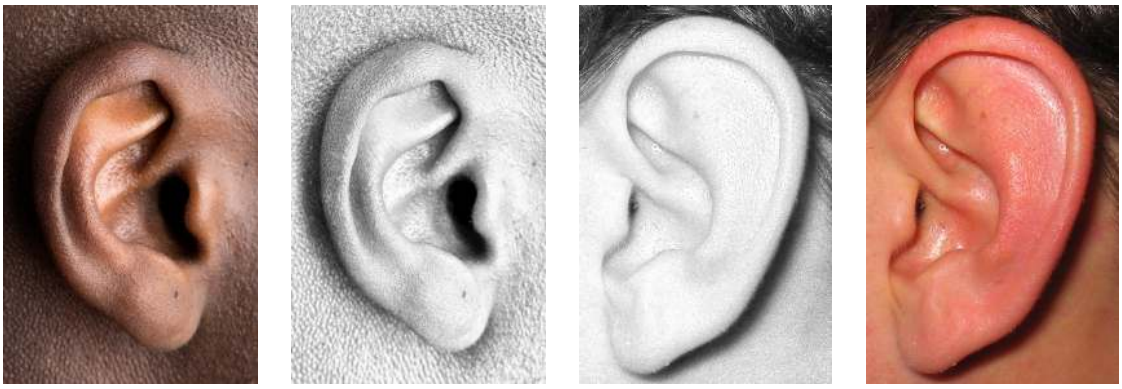
vestigated in Sec. 3.3. These investigations were important in order to evaluate the results of the selected algorithms being applied to clean and synthetically distorted ground-truth point clouds. However, these investigations were performed using isolated conditions which – in real-world scenarios – occur in various combinations. Additionally, the template and target pinnae were taken from the same listener, because the investigations focused on the robustness of the algorithms against artefacts from the PR, not the algorithms' abilities to personalise a pinna. Thus, in Sec.3.4, the effect of the template was investigated, and targets were both ground-truth and photogrammetrically reconstructed point clouds. In Sec. 3.5, it was investigated whether a specific algorithm is able to deal with holes in the target, as this is common in photogrammetrically reconstructed pinnae. Finally, in Sec. 3.6, it was investigated whether additional metadata obtained from the PR can improve the registration.

3.1 Algorithms

The algorithms were selected according to the following criteria: First, the algorithm must be applicable to at least three dimensions, as human pinnae are three dimensional objects. Second, the code of the algorithm must be freely accessible and available. Third, the code must be supported by a scientific publication. And fourth, not essential but definitely to my advantage, the code of the algorithm must be implemented in Matlab and can be edited.

One key algorithm that fit the criteria is coherent point drift (CPD), in which the point clouds are interpreted in a probabilistic way and the registration is formulated as density estimation problem. The other selected algorithms are extensions of CPD, details of which are described in the respective subsection of the algorithm.

There are other extensions of CPD that unfortunately do not fit the criteria. Golyanik et al. extended CPD with correspondence priors in a closed form and optimal sampling [155], similar to Zhou et al. who defined the prior probability as Dirichlet distribution [156]. Savalcalvo et al. extended CPD by utilising colour information of each point [157]. This approach was promising at first, however, when considering, e.g., just the brightness of points in the pinna (in order to be independent of human skin colour): Regions with strong inflexion can still be brighter than regions with low cartilage and high tissue concentration. Figure 3.1 shows two pinnae that fall into this category. From this observation I concluded that neither brightness nor colour necessarily correlate with depth information, thus, this CPD extension was not used.



(a) (b) (c) (d)

Figure 3.1: Two pinnae [158, 159] with disambiguous brightness information. **(a)** & **(d)** Including colour information and **(b)** & **(d)** grey scaled.

Serafin et al. took point normals into account [160], similar to the algorithm described in Sec. 3.1.2. Alexa et al. recently introduced blending whole shape volumes rather than their surfaces [153]. This "as-rigid-as-possible" shape interpolation is rigid by means of preserving local volumes.

Of course, there are NRR algorithms that are not extensions of CPD. Some are based on the iterative closest point (ICP) algorithm [161], as, e.g., Hirshberg et al., who defined a regularisation term for smooth and locally rigid shape deformation in combination with landmarks [162]. Some algorithms tackle the high dimensionality with various interpretations of the problem. One approach is based on the thin-plate spline algorithm with robust point matching (TPS-RPM), in which the shape is projected to a 3D grid, and the transformation of the grid deforms the shape respectively [163]. Peng et al. combined the shape deformation with a principal component analysis (PCA) [164]. In multiple other algorithms, the point clouds are downsampled before the registration is performed, and upsampled afterwards [165, 154, 166]. This has been done in order to minimise the degrees of freedom in the point cloud registration.

In the remainder of this chapter, I use the following notation: $\mathbf{X}$ denotes the target point cloud, $\mathbf{Y}$ denotes the template point cloud, and $\mathcal{T}(\mathbf{Y})$ denotes the non-rigid transformation $\mathcal{T}$ applied to the template in order to register it to the target.

3.1.1 Coherent point drift

The coherent point drift (CPD) algorithm formulates the transformation $\mathcal{T}$ of a template point cloud $\mathbf{Y}$ to a target point cloud $\mathbf{X}$ as a probability density estimation problem. The template is interpreted as the mean values of a GMM with isotropic covariances and the target is interpreted as observations, i.e., data points, from the same GMM. Point correspondences are estimated and the covariance is minimised, i.e., the points of the template are moved with respect to the MCT by updating the GMM's parameter set. These two steps are executed in an alternating fashion, because they depend on one another and

cannot be optimised at the same time. Once at least one criterion is met, the registration is finished. There are three possible criteria: First, the maximum number of iterations is reached, which can be set by the user. I chose a maximum number of 100 iterations. Second, a threshold is met that specifies the difference between movement steps of the template, it can as well be set by the user. I chose this tolerance threshold to be 10^{-6}, which corresponds to 0.1 mm for the point clouds in this thesis. Third, the isotropic covariance falls below a certain threshold, which is fixed at 10^{-7}, which corresponds to 0.01 mm for registrations in this thesis. This criterion can not be changed without editing the programme code.

Additionally, there are three parameters in CPD that can be set by the user: First, the amount of noise, w, present in the point clouds can be defined; It is normalised between 0 (all the noise is in the target) and 1 (all the noise is in the template). Second, the width of a smoothing Gaussian (spatial) filter, β, can be defined. The higher value β is set to, the closer strong curvatures will be fitted. Third, a trade-off parameter, λ, can be defined that balances the accuracy of the maximum likelihood estimation and the regularisation of the registration following the MCT. The regularisation can be interpreted as spatial lowpass filter, as it is defined in the Laplace domain [167]. Per default, $w = 0.1$, $\beta = 2$, and $\lambda = 3$. Empirical testing has shown that $\beta = 0.5$ and $\lambda = 1$ are ideal for registering human pinnae. The parameter w has been set to 0, because only ground-truth point clouds were used as template which did not contain any noise and all the noise was assumed to be in the target.

3.1.2 Preserving global and local structures

Preserving global and local structures (PRGLS) extends CPD by a classifier that accounts for curvature in geometrical structures [168]. This classifier is implemented as fast point-feature histograms (FPFH), which yield information about local structures in both template and target, and thus make the point correspondence estimation more robust against outliers [169]. The result of FPFH is then incorporated into the GMM as an additional membership probability.

There are six point features included in FPFHs: A local coordinate system of a point and three angular variations to a neighbouring point [170]. The angular variations describe the rotation of the local coordinate system to the neighbouring point. These six point features are calculated for each point in both point clouds and neighbouring points within a specific number of radii (ten per default). The number of radii can not be changed without editing the programme code.

Parameters that can be set by the user were done so as follows: Parameters that overlapped with those of CPD were chosen to be the same, e.g., $\beta = 0.5$ and $\lambda = 1$, the amount of noise $w = 0$, the maximum number of iterations was set to 100, and the tolerance was set to 10^{-6}. Because calculating the FPFHs for a single iteration was taking more than 30 minutes even on a computational cluster and because the template moved only marginally between two iterations, the FPFH calculation was set to be happening at every 20^{th} iteration.

3.1.3 Bayesian coherent point drift

In CPD, a velocity function following the MCT is chosen as the regularisation term for the registration. Bayesian formulation of coherent point drift (BCPD) uses a different approach for following MCT, i.e., by introducing a prior distribution using Bayesian inference. One advantage of BCPD is that it combines the rigid and non-rigid CPD, indeed, BCPD is a generalised form of CPD [171].

Parameters that can be set by the user were done so as follows: ω was set to 0, it describes an outlier probability. β was set to 1.2, it describes a Gaussian function's width which controls the range in which deformation vectors are smoothed. λ was set to 100, it controls the expected length of deformation vectors (smaller number corresponds to longer vectors). γ was set to 0.1, it controls a weight on the accuracy of the initial alignment. K and J were set to 100 and 300, respectively. They define two constraints for solving integrals discretely. c was set to 10^{-6}, it describes the convergence tolerance as the one in CPD. n was set to 300, it describes the maximum number of iterations.

Although BCPD is not implemented in Matlab, the definition of parameters and function call are. This enabled me to apply BCPD despite the algorithm being implemented in C++.

3.1.4 Geodesic-based Bayesian coherent point drift

From the same author as BCPD, there was another NRR algorithm implementation called geodesic-based Bayesian formulation of coherent point drift (GBCPD). This algorithm – unlike previous algorithms – does not depend on the Euclidean distance between points but takes the geodesic distance into account, i.e., the surface of the shape, respectively the connection between points. This property sounded promising regarding two problems that arose in investigations: First, holes in the target could be interpolated instead of replicated with the deformed template, because the points in the template are connected. Second, when calculating a mesh from the point cloud, the relationship between points is already clear as it did not change during the registration process and did not have to be recalculated.

Parameters that can be defined were mostly the same as used in BCPD, and also set to the same values. Additionally, the flag `-ux` was included in order to specify that shapes are roughly registered globally. tau was set to 0, it controls the balance between geodesic and Gaussian kernels. Finally, the last argument controls downsampling and interpolation acceleration outside the Bayesian framework and was flagged as `-DB,8000,0.02`.

3.2 Error assessment in high-quality photogrammetry

In this section, the effect of a high-quality PR approach on the pinna geometry was investigated in order to understand what problems NRR algorithms have to face in improving PR. If an NRR algorithm is unable to register a target with artefacts at the minimum, it is

likely unfit for the task of improving low-quality PR in which artefacts occur in varying combinations and to varying extent. In this investigation, three different high-quality PR instances of the same listener served as benchmark PRs, and the ground-truth mesh of this listener was available in order to classify and quantify the geometric artefacts.

3.2.1 Methods

Photogrammetrical reconstruction

In order to obtain the three PR point clouds, three sets of photos were taken as follows. First, the ear canal was blocked via colored foam earplugs (Classic II, E-A-R). Then, the earplugs were marked with a black marker dot in the middle of it to provide a unique landmark. To reduce the noise in the captured geometry caused by the listener's vellus hair, the vellus hair was shaved using a nose trimmer (NE3450, Remington). A white matte spray (3D Laser Scanning Spray, Helling) was used on the pinna to increase the contrast and to avoid subsurface scattering [114]. The pinna was evenly lit with the help of diffuse light sources; however, in PR set 3, a flash was used. An analog zoom was applied such that the depth of field was minimised. The camera's sensitivity was set to ISO 1600 and the aperture and shutter speed were chosen to be as small as possible [123]. Then, photos were taken along the surface of a vertical hemisphere with the pinna in the centre using a compact camera (DSC-WX200, SONY). These photos served as input to the PR software (Metashape version 1.6.5, Agisoft LLC). This software aligned the photos and extracted features [125], creating a sparse cloud. Then, it calculated a dense cloud using a reconstruction quality set to "high" and a depth filtering set to "mild". The dense PR point clouds were then affinely registered to the GT pinna using the *Align* tool (Meshlab, version v2022.02, Visual Computing Lab of ISTI-CNR) [172], such that their global properties, i.e., translation, rotation, and scaling, were approximately the same. Figure 3.2 shows the ground-truth point cloud and three such high-quality PR point clouds of the same listener.

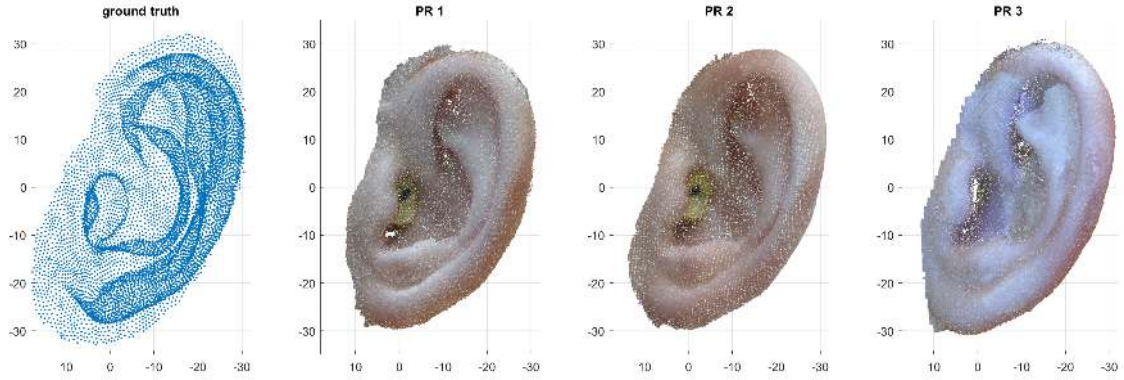


Figure 3.2: Ground-truth and three high-quality PR point clouds of the same listener (NH5).

Problem analysis

In general, the following problems with photogrammetrically reconstructed meshes were identified. First, the *global features* – location, orientation, and scaling of the head and pinnae – were unreliably estimated. Thus, the photogrammetrically reconstructed point cloud or mesh has to be aligned globally to a reference. Second, every pinna reconstruction resulted in a *different number of points* in the point cloud. Even the same method executed repeatedly does not yield even approximately a similar point cloud size. Third, it occured that the vellus hair on the pinna was reconstructed as part of the geometry, but it is an unwanted part for the purpose of HRTF calculation, because the *noise* in the mesh increases. Fourth, artefacts from the PR may result in *outliers*. And last, possibly in combination with problems three and four, the PR algorithm may *interpolated* the geometry in regions where only a small amount of features could be identified in the images. Interpolating holes is a delicate matter in human pinna geometries and depends on the number of features found and the size of the hole. As the available PR software is shape-agnostic, the decision on when a hole should stay a hole, e.g., when the Lobulus of a listener is stretched because the listener wears a tunnel piercing, was made based on software-internal penalties, likely to be distance-based.

The first problem – global alignment – was disregarded because local deformation of the pinna were the focus of this thesis, and global alignment is typically a preprocessing step in non-rigid registration [151]. The last problem – interpolation – consisted of too many unknown, and possibly dependent, variables which can be altered in a systematic way. This problem was likely to be solved only within a PR algorithm and, thus, discarded in this thesis. This left the classification covered in this investigation to be different number of points, noise, and outliers, which were quantified in the remainder of this investigation.

Artefacts quantification

The difference in point cloud size was calculated as a ratio between the number of points in the reference and the number of points in the pinna region of the sparse and dense point clouds, and the inverse ratio of the two point cloud sizes.

ℓ, a "one-sided" measure (see Eq. 2.1), was used in order to classify the points in the PR point clouds. It was calculated between the reference and each PR, and between each PR and the reference. Note that, previous studies could show a maximum ℓ of far below 1 mm [128, 129, 130]. However, these studies performed a photogrammetric reconstruction of a KU100 dummy head and compared it to a structured-light scan. This is problematic in the sense that, photogrammetric reconstructions of dummy heads tend to be more accurate due to several problems of human skin and missing parts of anatomy and behaviour (moving, blinking, etc.), as an extensive study on optical scanning methods has found [114]. In this thesis, the focus lies on photogrammetric reconstruction of real-life human pinnae.

Two thresholds for ℓ were defined in order to separate the three artefacts categories: All points showing a distance smaller or equal 1 mm were considered as *correctly registered*,

because these distances lie within the resolution of the reference mesh, and previous investigations have deemed 1 mm to be a reasonable solution for the pinna region in order to calculate plausible personalised HRTFs [78]. Points showing a distance larger than 3 mm cannot represent the pinna geometry correctly anymore, because the average thickness of skin and cartilage in the pinna region is approximately 3 mm [173]. Thus, these points were considered as *outliers* and were quantified by *outlier rates* in %. Consequently, all points showing a distance between 1 and 3 mm were considered as *noise* and were quantified by *noise rates* in %.

3.2.2 Results and Discussion

Table 3.1 shows the number of photos taken, the resulting PR point cloud sizes for the sparse and the dense cloud, the latter separately for the whole reconstruction and in the manually segmented pinna region only. The ratio between point cloud sizes was 1.5, 0.6, and 0.5 when comparing the reference to the sparse point clouds in the pinna region, and 0.008, 0.01, and 0.01 when comparing the reference to the dense point clouds in the pinna region.

Table 3.1: Number of photos and resulting size of the three PR point clouds. For comparison, the reference point cloud consisted of 8,973 points.

	PR1	PR2	PR3
Number of photos used for the PR	54	77	125
Number of points in the sparse point cloud	20,108	54,774	63,725
... in the pinna region only	6 k	15 k	18 k
Number of points in the dense point cloud	7.5 M	5.5 M	4.9 M
... in the pinna region only	1.1 M	860 k	833 k

Figure 3.3 shows the classification of points on the point clouds and by means of histograms. Between 40.9 and 66.7 % of the points were classified as correctly registered, between 2.5 and 7.8 % of the points were classified as outliers (median 4.3 %), and between 30.7 and 52.0 % of the points were classified as noise (median 43.6 %).

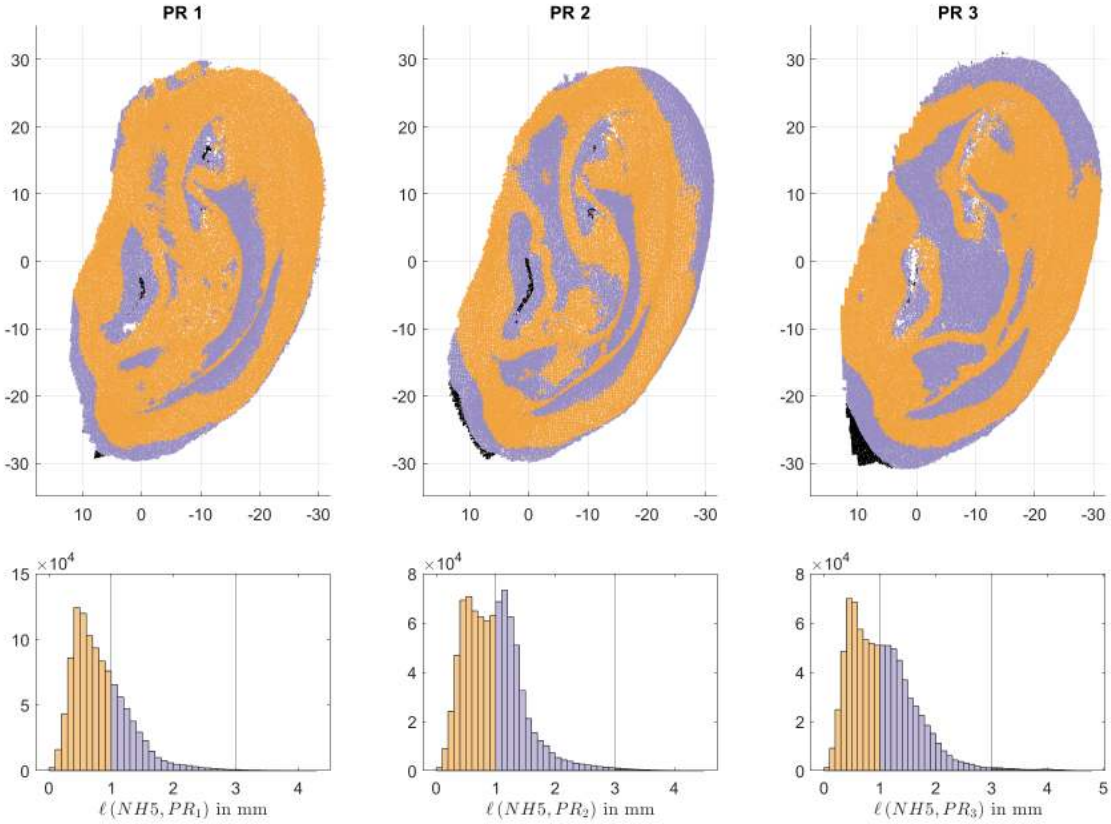


Figure 3.3: ℓ for three PR instances of the same listener, visualised on the pinna shape (top row) and as histograms (bottom row). Less than 1 mm were defined to be correctly registered (yellow), between 1 and 3 mm were deemed to be noise (purple), and points with a distance higher than 3 mm were deemed to be outliers (black).

In this investigation, three artefacts that occur in high-quality PRs were identified: point cloud size difference, noise, and outliers. Without the usage of professional equipment, the size and quality of pinnae PR point clouds varied tremendously, thus, outlier and noise rates are likely to be higher than those found in this investigation. It is suggested that wider ranges than found for noise and outliers and several intermediate ratios than found for point cloud size difference should be applied systematically. The results of this investigation were used in order to systematically apply distortions to pinna geometries and investigate the robustness of the NRR algorithms against these artefacts.

3.3 Effect of systematic geometry distortions

In Sec. 3.2, three artefact categories have been identified. The next step was to apply these artefacts with systematically varying severity to ground-truth target point clouds and to investigate the robustness of the four NRR algorithms. The template and target point clouds were from the same listener for each registration in order to observe the robustness of the algorithms separated from personalisation.

3.3.1 Methods

Evaluation measures

The quality of the registration was evaluated by means of three geometric errors calculated between the registered template $\mathcal{T}(\mathbf{Y})$ and the respective target $\mathbf{X}$. First, the Hausdorff distance h [174, 175] was calculated:

$$h(\mathbf{X}, \mathcal{T}(\mathbf{Y})) = \max\left\{ \sup_{\mathbf{x}_i \in \mathbf{X}} \ell(\mathbf{x}_i, \mathcal{T}(\mathbf{Y})), \sup_{\mathbf{y}_i \in \mathcal{T}(\mathbf{y}_j)} \ell(\mathcal{T}(\mathbf{y}_i), \mathbf{X}) \right\}. \tag{3.1}$$

Second, the balanced average Hausdorff distance $\hat{h}$ [176] was calculated:

$$\hat{h}(\mathbf{X}, \mathcal{T}(\mathbf{Y})) = \frac{1}{2N}\left(\sum_{i=1}^{N} \ell(\mathbf{x}_i, \mathcal{T}(\mathbf{Y})) + \sum_{j=1}^{M} \ell(\mathcal{T}(\mathbf{y}_j), \mathbf{X}) \right), \tag{3.2}$$

where $1/2N$ is used to account for a hidden error in ranking. Note that, in contrast to Eq. 2.1, Eqs. 3.1 and 3.2 are two-sided metrics. Several other extensions of the Hausdorff distance exist [177, 178], but $\hat{h}$ was deemed the one appropriate for this investigation.

Third, the Jaccard index J [179] was calculated:

$$J(\mathbf{X}, \mathcal{T}(\mathbf{Y})) = \frac{|\mathbf{X} \cap \mathcal{T}(\mathbf{Y})|}{|\mathbf{X} \cup \mathcal{T}(\mathbf{Y})|}, \quad 0 \leq J \leq 1 \tag{3.3}$$

by voxelising both pinnae and evaluating intersection over union of said voxel grids. The voxel size was defined as 2.0 mm to be sensitive for a distance of 1.0 mm between points, i.e., roughly the distance between points considered to result in an accurate pinna point cloud. Note that the pinna point clouds used in this thesis described surfaces, not complete volumina.

Tested conditions

The effects of the artefacts based on the results in Sec. 3.2 were investigated by systematically distorting pinna geometries and applying the four NRR algorithms to them.

The point cloud *size* of the target and template was varied. These were two separate conditions, because the results of the previous investigation showed that the reference point cloud was not always the smaller point cloud (see number of points for sparse PR_1 in pinna region), and because there is an ambiguity in point correspondence for differently sized point clouds. When the target is smaller than the template point cloud, there is a chance that different points in the template are registered to the same single point in the target. And when the template is smaller than the target point cloud, there is a chance that the registered template will include non-uniformly spatially distributed points. In the *target* size condition, the number of points in the target was reduced and the template size was constant. In the *template* size condition, the number of points in the template was

reduced and the target size was constant. The reduction was achieved by removing 1, 2, 5, 10, 20, 40, 60, 80, 90, and 95 % of the points in the corresponding point cloud in a randomized fashion based on a uniform distribution. In this investigation, the point cloud size difference is given as relative ratio, which can be interpreted as how much % of the points were removed in either the reference or the target and it was separated in rounded step sizes. Additionally, a reference condition of removing 0 % was included. Note that the largely reduced point clouds do not represented the pinna geometry sufficiently accurate for plausible HRTF calculation, but the algorithms were to be tested beyond the threshold of the resolution considered sufficiently high.

To investigate the effect of *outliers*, 5, 10, and 20 % of points were added to the target point cloud. Note that by doing so, the target size increased with the amount of outliers. The outliers were drawn from a 3D Gaussian distribution with a mean representing the arithmetic centre of the point cloud and a standard deviation of half the dimensions of the point cloud's bounding box. Again, a reference condition with 0 % outliers was included.

To investigate the effect of *noise*, 20, 40, 60, 80, and 100 % of points in the target point cloud were randomly displaced. Note that by doing so, the target size did not change. The displacement was modeled by a 3D Gaussian distribution with zero mean and a standard deviation of 10 % of the point cloud's bounding box. A reference condition of 0 % noise was included.

Added points in the outliers condition and the selected points in the noise condition were distributed evenly over the target pinna point cloud. Although some regions – especially occluded ones – might experience a higher probability of noise and outliers present, these regions are not in the same position or of the same size for all PRs (see Fig. 3.3). Note that, for a proof of concept, the NRR algorithms have been applied to left ears only.

In the evaluation, the goal was the evaluation metrics to reflect the success of the registration *despite* the distortion in the target. Thus, for the conditions *different target point cloud size*, *noise*, and *outliers*, the registered point cloud $\mathcal{T}(\mathbf{Y})$ was compared with the respective GT point cloud $\mathbf{Y}$. For the conditions *different template point cloud size*, the registered point cloud $\mathcal{T}(\mathbf{Y})$ was compared with the respectively sized target point cloud $\mathbf{X}$.

Computational aspects

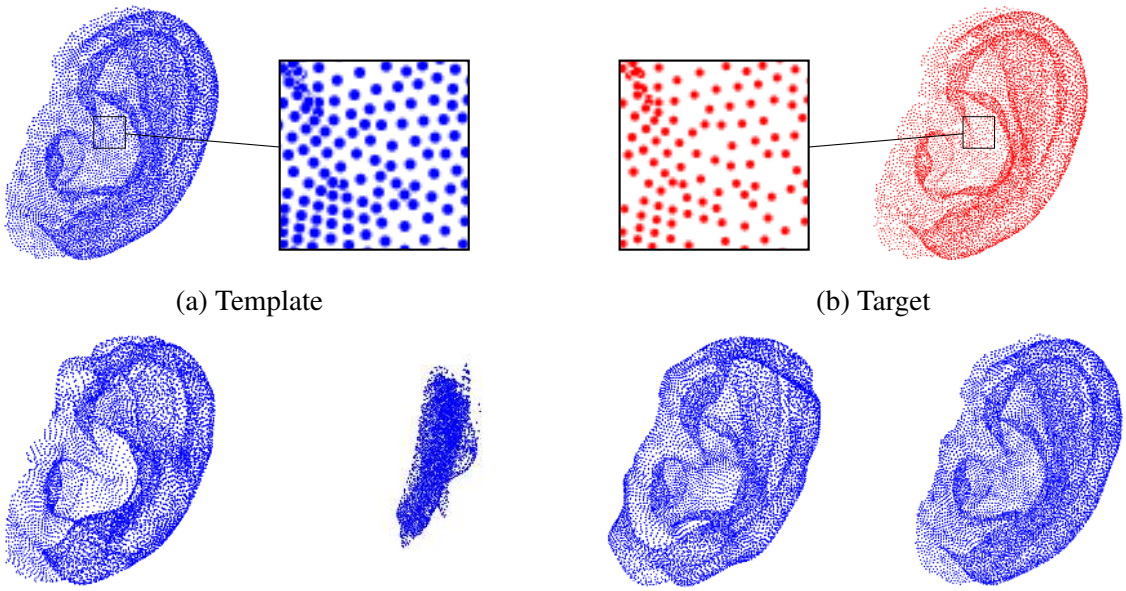
CPD, BCPD, and GBCPD used up to 2 GB peak memory, and PRGLS used up to 160 GB peak memory. CPD's runtime was around 10 to 20 minutes, for all tested conditions. PRGLS' runtime, however, varied vastly depending on the condition. For reduced target size, it was between 4 and 10 *days*, while for the remaining conditions (reduced template size, noise, and outliers), it was between 5 minutes and 3 hours. BCPD and GBCPD had a runtime between 30 seconds and 3 minutes, for all tested conditions. The reason for the large runtime differences may be the algorithms' implementations: CPD and PRGLS were implemented in Matlab mostly; BCPD and GBCPD were implemented in C, and Matlab was used only as interface to specify command line parameters and programme calls.

3.3.2 Results and Discussion

General analysis

The four applied algorithms behaved differently across conditions, and albeit independent of the pinna shapes involved, systematic differences in the performance of the algorithms have been observed and can be summarised in this section.

As an example for the general analysis, Figure 3.4 shows the point clouds before and after a registration using the four algorithms in the condition target size reduction by 10 %. In the top row, template and target pinna are shown. The zoom-in shows that the overall shape of the pinna looked the same, but the number of points was different. The bottom row shows the point clouds after applying the four algorithms to the template. From this visual inspection, PRGLS performed worst, and GBCPD performed best, with the latter yielding a good-quality point cloud from which a mesh and furthermore individual HRTFs may be calculated. CPD and BCPD did not yield well-registered results, with a lot of details lost in the helix rim and concha.



(a) Template (b) Target

(c) CPD (d) PRGLS (e) BCPD (f) GBCPD

Figure 3.4: Effect of a target size reduction by 10 %. **(a)**: Template point cloud. **(b)**: Target point cloud, i.e., template point cloud reduced by 10%. Registered templates by **(c)** CPD, **(d)** PRGLS, **(e)** BCPD, and **(f)** GBCPD.

Table 3.2 shows the corresponding geometric error metrics obtained for the same registrations. h indicates that the registration using PRGLS failed, as does $\hat{h}$. $\bar{\ell}$ does not indicate the magnitude of the shape differences as its values around 1 mm suggest similar registration results with all four algorithms. The discrepancy between the quantity value and the corresponding point-cloud quality was the reason this error measure was discarded in the further evaluation. J indicates the PRGLS registration as a failed one, CPD and BCPD as approximately of the same quality, and GBCPD as the best-performing algorithm. This

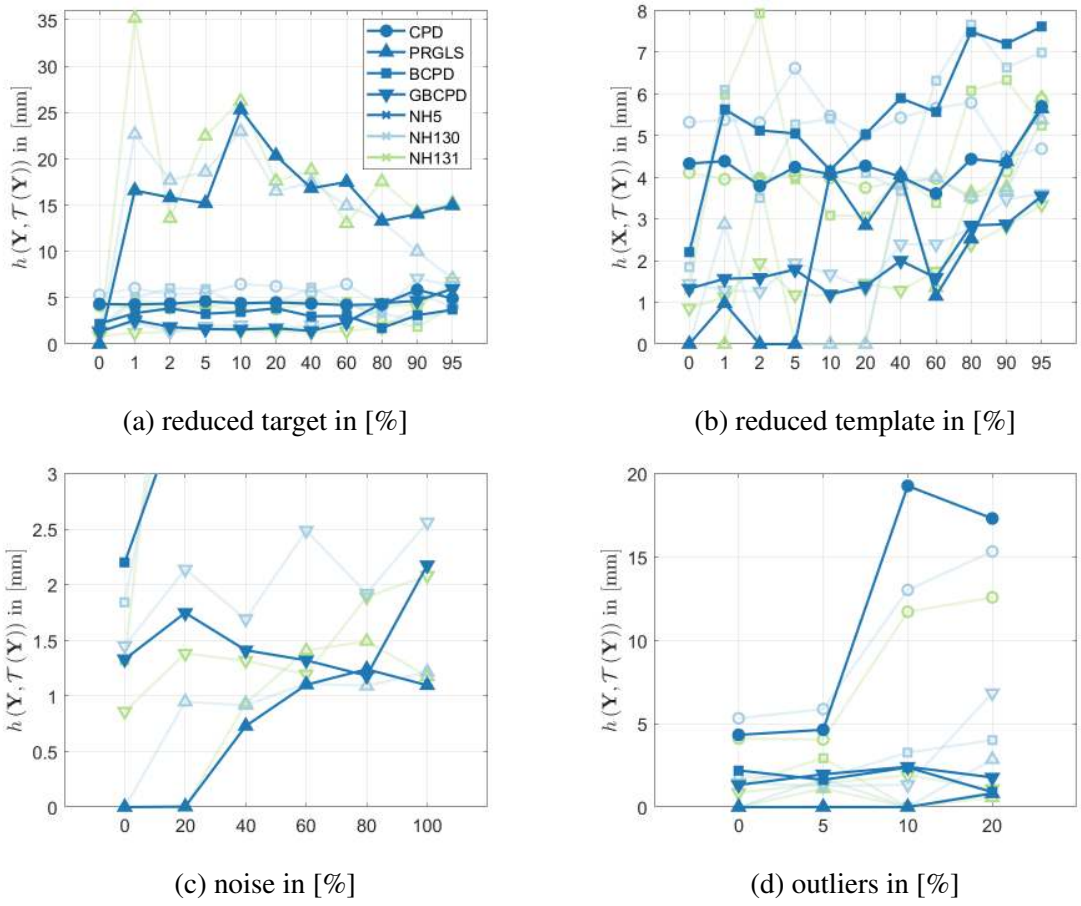
general analysis served as an example of the magnitude of the error measures used and that they are good indicators for the geometric quality assessment of a registration.

Table 3.2: Effect of a target size reduction by 10 % for registrations with the four tested algorithms. $\bar{\ell}$ represents the median of ℓ.

	CPD	PRGLS	BCPD	GBCPD
h (in mm)	4.41	25.30	3.48	1.58
$\hat{h}$ (in mm)	0.56	3.59	0.47	0.36
$\bar{\ell}$ (in mm)	0.48	1.12	0.44	0.39
J	0.72	0.08	0.79	0.87

The following figures present the results of the registrations for different conditions, algorithms, and listeners, in order of the geometric error measures introduced in Sec. 3.2. They all have the same structure: The panels represent the four investigated conditions (target size reduction, template size reduction, noise, outliers). Each panel contains the geometric error resulting from the registration for each listener (color) and each algorithm (symbol). The results for listener NH5 are highlighted as this listener covers large parts of the investigations in this chapter.

Figure 3.5 shows the Hausdorff distances h obtained for the tested conditions. h seemed to fluctuate almost randomly and present high levels, even within the same algorithm and listener, which might be because of this metric being highly sensitive to single distant points. This is a clear problem of h, because an accurate mesh can result from registration despite containing single distant points which would be discarded in a mesh construction anyway.



(a) reduced target in [%]

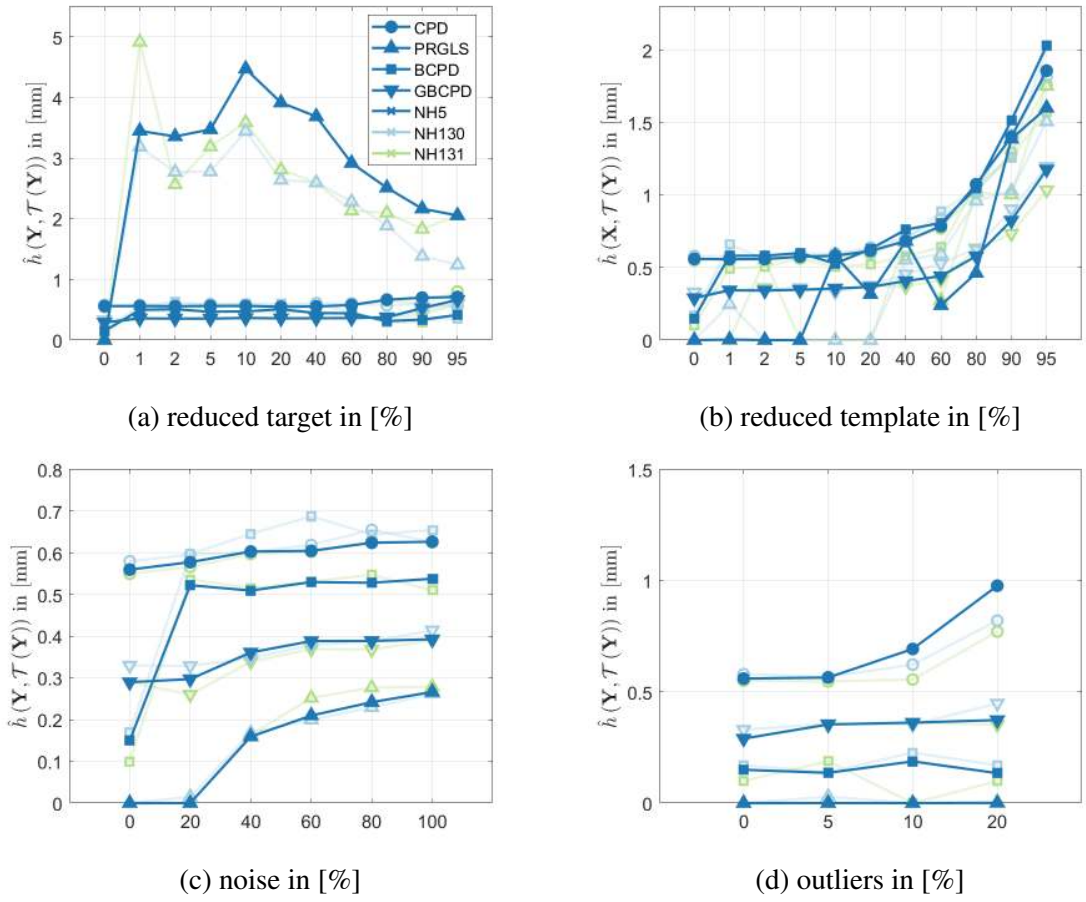
(b) reduced template in [%]

(c) noise in [%]

(d) outliers in [%]

Figure 3.5: Hausdorff distance h (in mm) for all tested conditions. CPD is depicted by a circle, PRGLS by a triangle pointing up, BCPD by a square, and GBCPD by a triangle pointing down. The results for the three pinnae are shown colour-coded, with NH5 shown as filled symbols.

Fig. 3.6 shows the balanced average Hausdorff distances $\hat{h}$. $\hat{h}$ showed smaller fluctuations across the levels of deformation. However, its interpretation raises questions. For example, all four algorithms moslty yield $\hat{h}$ below the required threshold of 1 mm, but the resulting point clouds differed highly in quality when visually compared (see Tab. 3.2 and Fig. 3.4). This issue rendered $\hat{h}$ inadequate for further analyses.



(a) reduced target in [%]

(b) reduced template in [%]

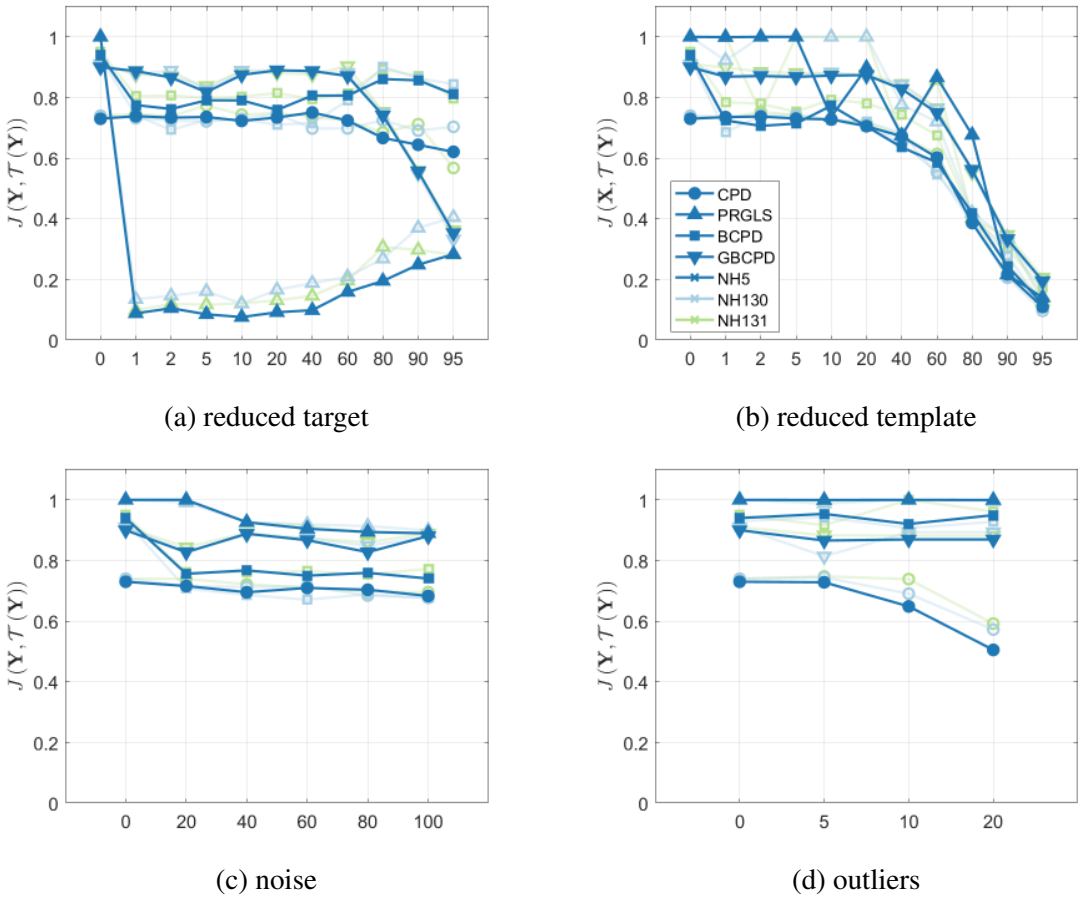
(c) noise in [%]

(d) outliers in [%]

Figure 3.6: Balanced average Hausdorff distance $\hat{h}$ in mm for all tested conditions. Conventions as in Fig. 3.5.

Fig. 3.7 shows the Jaccard indices J. This distance measure showed little fluctuations across the levels of deformations and allowed interpretation of the success of the registration. Thus, these results were selected for statistical testing and general interpretation.

A Shapiro-Wilk normality test was performed and showed that the data was not normally distributed ($p < 0.05$). Then, a Kruskal-Wallis rank sum test was applied on the Jaccard index with the factors algorithm (CPD, PRGLS, BCPD, GBCPD) and pinna (NH5, NH130, NH131). This test revealed a significant effect of the algorithm ($p < 0.05$), but not a significant effect of the pinna ($p = 0.60$). This indicates that the NRR algorithms performed differently but their performance did not depend on the pinna shape used.



(a) reduced target

(b) reduced template

(c) noise

(d) outliers

Figure 3.7: Jaccard index J for all tested conditions. Conventions as in Fig. 3.5.

In the general analysis, the registrations by means of three error measures were investigated. Additionally to the three error metrics, $\bar{\ell}$ was included, the median of the pointwise smallest distance ℓ. It was shown that this metric was useful for a pointwise classification, but its median did not indicate unambiguously whether a registration was actually successful. It was also shown that h and $\hat{h}$ were fluctuating, thus sensitive to noise and outliers, and their interpretation seemed to be inconcistent. Finally, the Jaccard index J was selected as an appropriate metric in order to evaluate the registration performance in the geometric domain.

Condition-specific analysis

The following analysis was based on the results presented in Fig. 3.7, complying with the structure of the figure: The conditions were analysed independently from each other in the order of the figure panels.

In the *target size* condition, Js obtained with PRGLS were lowest, especially when the target had much less points than the template. This may originate in incorporating the additional information of the FPFHs [169] for each point and various radii. In detail,

when calculating the point cloud correspondence, a linear equation system of dimensions *number of points in template* $\times$ *number of points in target* $\times 60$ is set up and solved using the "Hungarian" algorithm [180]. In the target size condition, this equation system has the dimensions *number of points in target* $\times$ *number of points in target* $\times 60$; its first dimension has non-zero entries from 1 to the number of points in the template and is padded with a small positive float number (double precision) to the size of the target, adding rows and thus yielding an equation system without a unique solution. This might be the reason for failed registrations, an example of which is shown in Fig. 3.4d. These results indicate that, in future applications of PRGLS, the target size needs to correspond to that of the template. GBCPD performed best, yielding Js of more than 0.8 for target size reductions of up to 60%, rendering GBCPD the favorite candidate in the target size condition.

In the *template size* condition, CPD and BCPD showed the lowest Js. PRGLS yielded excellent Js for small reductions, but performed poorly for a reduction by more than 10%. This can be explained by the size of the linear equation system within PRGLS having a size of *number of points in template* $\times$ *number of points in template* $\times 60$, its second dimension being padded to the size of the target with a small positive number, thus offering a less ambiguous solution. GBCPD yielded excellent Js for a reduction by up to 40%, rendering GBCPD the favorite candidate in this condition.

In the *noise* condition, CPD and BCPD showed Js between 0.7 and 0.8. GBCPD showed Js around 0.9, and PRGLS showed the highest Js. This renders PRGLS as the favorite candidate in the noise condition being robust against noise.

In the *outliers* condition, CPD showed the lowest Js, which may be due to CPD assuming uniform probability for point correspondence and not including any existing a-priori information about the point clouds other than their 3D coordinates. BCPD and GBCPD yielded good results, showing Js above 0.8, with BCPD performing slightly better. PRGLS showed the highest Js, which might be explained by the FPFHs being able to identifying outliers easily: For small enough radii, FPFHs were identical for all points but the outliers. The high Js suggest that PRGLS is an algorithm insensitive to outliers and render PRGLS the favorite candidate in the outliers condition.

Taken together, independent of the applied algorithm, a similar point cloud size for both template and target seemed to be advantageous. This can be achieved by resampling the point clouds to the same size before applying an NRR algorithm. Interestingly, PRGLS yielded an excellent registration for up to 20% noise, but failed when points were missing in either the template or target. This may be due to the FPFHs containing misleading information about the local coordinate systems for points with missing neighbours.

For different target or template sizes, GBCPD is the preferred candidate. However, when the target and template sizes were matched, PRGLS outperformed the other algorithms by means of robustness against noise outliers.

In summary, the registration results were evaluated using three geometric error measures: the Hausdorff distance, the balanced average Hausdorff distance, and the Jaccard index. The comparison across these measures indicates that the Jaccard index provides a clearer picture when analyzing the registration results than using the other two measures. The

statistical analysis performed on the Jaccard index showed that the NRR algorithms performed significantly different and that their performance did not depend on the pinna shape.

Overall, the investigated NRR algorithms showed high Js when registering ground truth (GT) point clouds. Based on the results of the categorization of artefacts, GT pinna point clouds were synthetically distorted in a systematic way with varying artefact severity and the performance of the four NRR algorithms – CPD, PRGLS, BCPD, and GBCPD – was evaluated. In addition to effects of noise and outliers, the effects of the mismatch between the sizes of the template and target point clouds were investigated as well.

PRGLS showed a good robustness to noise and outliers, however, it was sensitive to differences in point cloud sizes, indicating that PRGLS requires similar template and target sizes. CPD was sensitive to noise and outliers. I concluded that BCPD and GBCPD were the two algorithms fit for the registration task, because they yielded the highest Js when compared across all tested conditions.

3.4 Effect of the template

The previous investigation covered isolated conditions in which – albeit systematically evaluated – the template was from the *same* listener as the target, and the targets were *synthetically* distorted GTs. The investigation in this section extends the pool of templates by an average pinna and a non-individual GT pinna, and the pool of targets by GTs of non-individual listeners and high-quality PRs of NH5 (previously used in Sec. 3.2). This section is separated by templates used. Note that, for the last template, only BCPD is applied because for the first two templates, BCPD yielded highest Js out of the four NRR algorithms and was thus considered the best fit. For the first template, 24 registrations have been done: four algorithms were each applied to three GTs and three PRs. For the second template, 24 registrations have been done: four algorithms were each applied to three GTs and three PRs. And for the third template, three registrations have been done: one algorithm was applied to three PRs. In total, 51 registrations were done.

The GT point clouds used in this investigation were the same as in the previous investigation, i.e., listeners NH5, NH130, and NH131 from the ARI database. Three PR point clouds were the same as in Sec. 3.2. Newly introduced in this chapter was *AVG*, which is the mean pinna point cloud derived from a principal component analysis (PCA) of structured-light scans of 119 listeners [40]. In this investigation, the open ear canal of AVG was blocked, and the point cloud was resampled to match the resolution of 1 mm and to guarantee a regular distribution of points. The number of points of the targets were reduced to match the number of points to the respective template, because of the better performance for quasi-equally sized point clouds.

Selecting an adequate template may be essential for the NRR algorithm, as it is reasonable to assume that an a priori high similarity between template and target may lead to better registration results by means of geometric error. However, equally reasonable is the selection of an average pinna, as it is "closest to most" pinnae. In this section, three

templates were used: the individual GT (NH5), AVG, and a GT from a different listener. The effect of the template choice on the registration results was investigated using GTs from three listeners (NH5, NH130, and NH131) and three high-quality PRs of listener NH5 as targets.

For the remainder of this chapter, three types of conditions were covered: Validation, personalisation, and robustness. *Validation* describes a condition in which the template and target are the same point cloud, which is the same as the 0% reference condition in the previous section and is included in this section for comparison. *Personalisation* describes a condition in which the target pinnae are from different listeners than the template. *Robustness* describes a condition in which the target pinna is a PR point cloud containing previously investigated artefacts to varying degree of severity. Note that, these three conditions cannot be completely separated, e.g., when a template ear from one listener is registered to PR targets from a different listener. In this example, the conditions personalisation and robustness are combined.

In order to evaluate the registration performance, the Jaccard index J was calculated between the unregistered template $\mathbf{Y}$ and the target $\mathbf{X}$, then, the registrations were performed, and, finally, the Jaccard index J was calculated between the registered template $\mathcal{T}(\mathbf{Y})$ and the respective target $\mathbf{X}$. A J higher after the registration suggests that the registration improved the similarity between the point clouds, which was the desired outcome.

3.4.1 Template I: individual ground-truth

Validation

The left part of Figure 3.8 shows Js before and after the registrations in the validation. PRGLS was the only algorithm that showed $J = 1$, which suggests this algorithm did not deform the template at all. This is likely because of the membership probability rooted in the FPFH calculation, which yielded *the same* features for each point. BCPD and GBCPD showed Js of 0.94 and 0.90, respectively, and deformed the pinna slightly. CPD showed a J of 0.73 and deformed the template considerably. This poor result might be because CPD assumes the point corrspondence to be uniformly distributed and covariances to be isotropic within the GMM. Hence, this validation renders PRGLS as the only algorithm able to detect when template and target were the same point cloud.

Personalisation

The right part of Figure 3.8 shows Js before and after the registrations in the personalisation. Before registration, the comparison between NH5 and GTs of listeners NH130 and NH131 yielded Js of 0.23 and 0.19, respectively. After registration with BCPD, the Js increased to 0.86. GBCPD yielded Js of 0.79 and 0.71 for NH130 and NH131, respectively. The slightly poorer performance was surprising because by taking the relationships between points into account, GBCPD was expected to perform better. CPD yielded Js of 0.25 and 0.21, almost unchanged compared to the Js prior to registration. PRGLS

improved J a lot for NH131 ($J = 0.90$), but only a little for NH130 ($J = 0.64$). For the personalisation, BCPD is rendered as the favorite candidate to register a template from one listener to GT pinna point clouds from a different listener.

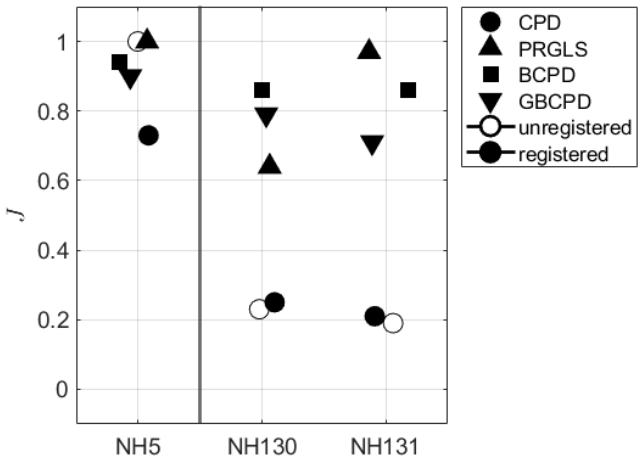


Figure 3.8: The effect of validation and personalisation with NH5 GT as template and the GTs of NH5, NH130, and NH131 as targets. J was computed between the unregistered template and the respective target before registration (open symbols), and between the registered template and the respective GT of the target after registration (closed symbols). Symbols denote the four algorithms: CPD (circle), PRGLS (triangle facing upwards), BCPD (square), and GBCPD (triangle facing downwards). The vertical line separates the validation (left) and the personalisation (right).

Robustness

Figure 3.9 shows the PR-based registrations using the NH5 template for a detailed inspection of the template, targets, and registered point clouds. Figure 3.10 shows Js before and after the three registrations in the robustness. The comparison between the unregistered template NH5 and the three PRs of NH5 showed Js of 0.29, 0.35, and 0.43. After the registration, PRGLS improved the Js to be > 0.95 for all registrations, because it deformed the template only slightly. CPD also improved the Js to 0.65, 0.56, and 0.54. However, the resulting point clouds are not a good basis for a calculation of a plausible ear manifold, because points with opposing normals drifted towards each other in thin regions, e.g., the helix. Although BCPD showed Js similar to CPD, those BCPD-obtained point clouds are a better basis for the calculation of a plausible ear manifold. Additionally, these resulting point clouds visibly showed a higher detail in smaller areas, e.g., the fossa triangularis and the cymba conchae (see Fig. 3.9). GBCPD improved the Js only marginally (Js of 0.31, 0.34, and 0.47, respectively), although the resulting point clouds seem to be plausible and a plausible manifold might be calculated easily as well. The small Js may be explained by "ripples" on the helix or errors in the global registration of the pinna. For the robustness, PRGLS is rendered as the favorite candidate to register a template from

one listener to PR targets of the same listener.

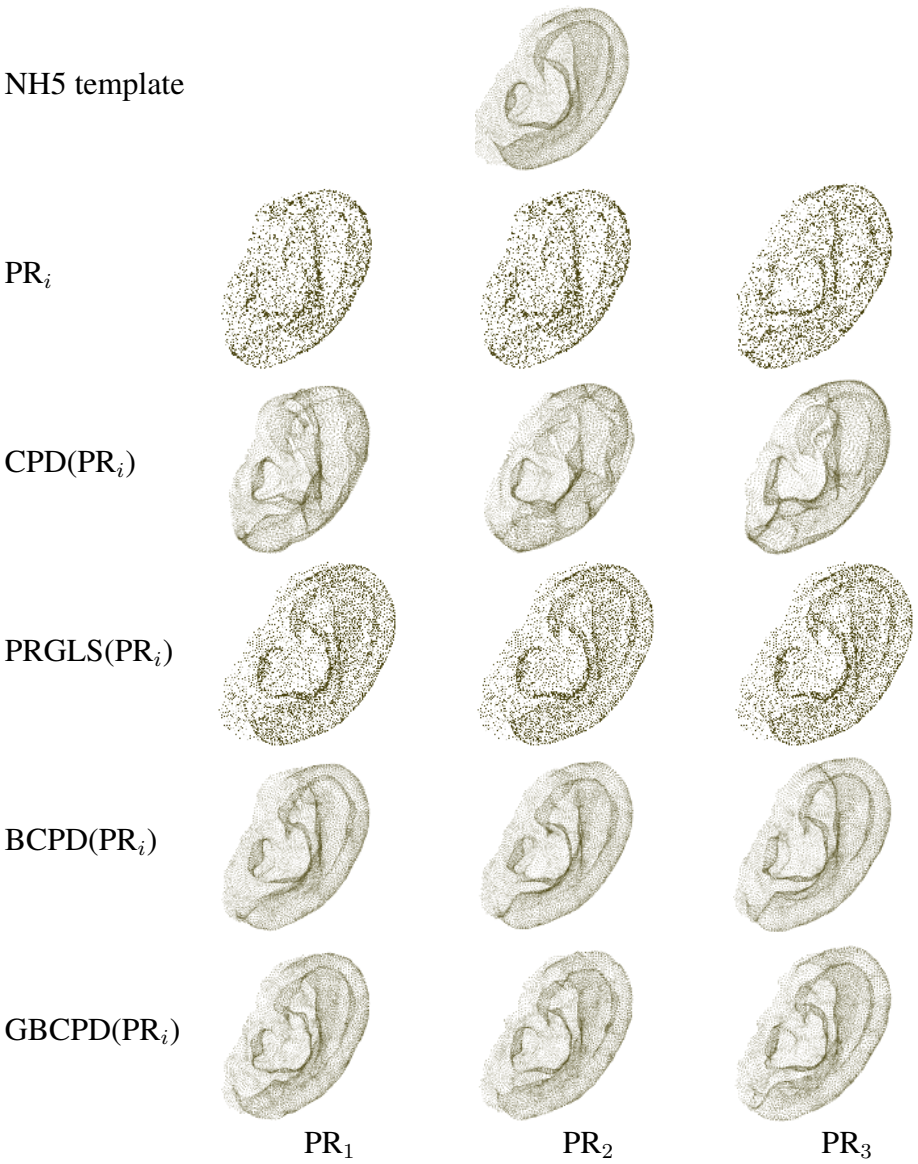


Figure 3.9: Registration results using NH5 as template. First row: Template ear. Second row: Target ears, i.e., PR scans of NH5. Other rows: Registration results.

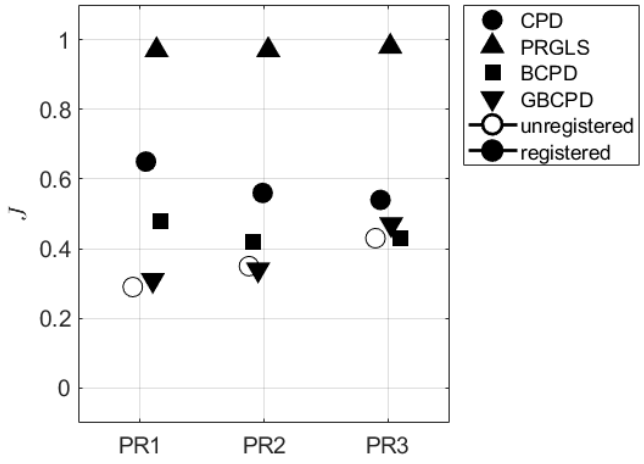


Figure 3.10: The effect of robustness with NH5 GT as template and three PRs of NH5 as targets. Conventions as in Fig. 3.8.

3.4.2 Template II: average pinna

In this section, the effects of personalisation and the combination of personalisation and robustness using an average pinna shape as template were investigated.

Personalisation

Figure 3.11 shows Js before and after the three registrations in the personalisation. The comparison between the unregistered template and the three GT point clouds of NH5, NH130, and NH131 showed Js of 0.12, 0.09, and 0.15, respectively. This is lower than the Js obtained for the personalisation using NH5 as template (Sec. 3.4.1), indicating that AVG shares less personal shape details with NH130 and NH131 than NH5 does.

BCPD improved the Js substantially to 0.80, 0.80, and 0.83, respectively. CPD improved the Js to 0.71, 0.71, and 0.76, respectively. It was observed that points with opposing normals drifted towards one another, providing no useful basis for the calculation of a plausible ear manifold from the registered template. GBCPD improved the Js to 0.63, 0.63, and 0.59, respectively. These Js indicate that the geodesic distance underlying GBCPD might not be an appropriate criterion for pinna registration when using AVG as template. PRGLS showed Js at a poor level, which indicates that PRGLS did not modify the template much. Hence, BCPD seems to be the preferred candidate when it comes to the personalisation of AVG.

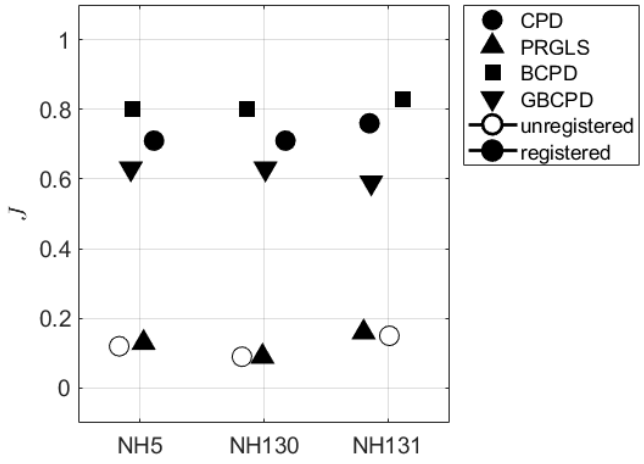


Figure 3.11: The effect of personalisation with AVG as template and the GTs of NH5, NH130, and NH131 as targets. Conventions as in Fig. 3.8.

Personalisation and robustness

Figure 3.12 shows the PR-based registrations using the AVG template for a detailed inspection of the template, targets, and registered point clouds. Figure 3.13 shows Js before and after the three registrations in the combination of personalisation and robustness. The comparison between AVG and the three PR point clouds of NH5 showed Js of 0.12, 0.20, and 0.13, respectively. These Js were less than the ones using NH5 as template (Sec. 3.4.1), indicating that AVG shares less personal shape details with the PRs than the GT of NH5 does, most prominently the pinna rotation angle.

CPD, BCPD, and GBCPD all yielded similar improvements by means of Js (around 0.4), but the resulting geometries appeared to be rather different. BCPD and GBCPD yielded point clouds which might be a good basis for the calculation of a plausible ear manifold (see Fig. 3.12). These two algorithms seemed to have registered, e.g., the concha accurately, but not the fossa triangularis. This emphasizes the problem of setting the parameters β (the width of a spatial Gaussian filter) and λ (the weighting of the regularisation of point movement), as the pinna contains regions with strongly varying curvatures. GBCPD exhibited artefacts in the helix that may be the cause of the low J. CPD, having the inherent effect of points drifting to opposing sides, yielded a plane surface instead of a manifold with a front and back side being close to each other as it can be observed, e.g., around the helix, in the CPD registration to PR_1. Additionally, the height and width of the pinna and the helical fold of PR_3 have been registered well, but the inner parts (concha, fossa triangularis, crus inferius and superius anthelicis) were poorly registered. PRGLS showed Js similar to no registration, which indicates that PRGLS deformed the template only slightly.

In summary, CPD and BCPD yielded the best Js regarding personalisation and robustness using AVG as template. Despite the similar Js, CPD registrations showed an unwanted effect of "flattening" areas in the pinna that are crucial for the calculation of personalised

HRTFs, because points of opposing sides drifted towards each other. This rendered BCPD as the preferred candidate in the combination of personalisation and robustness using AVG as template and PR-based point clouds as targets.

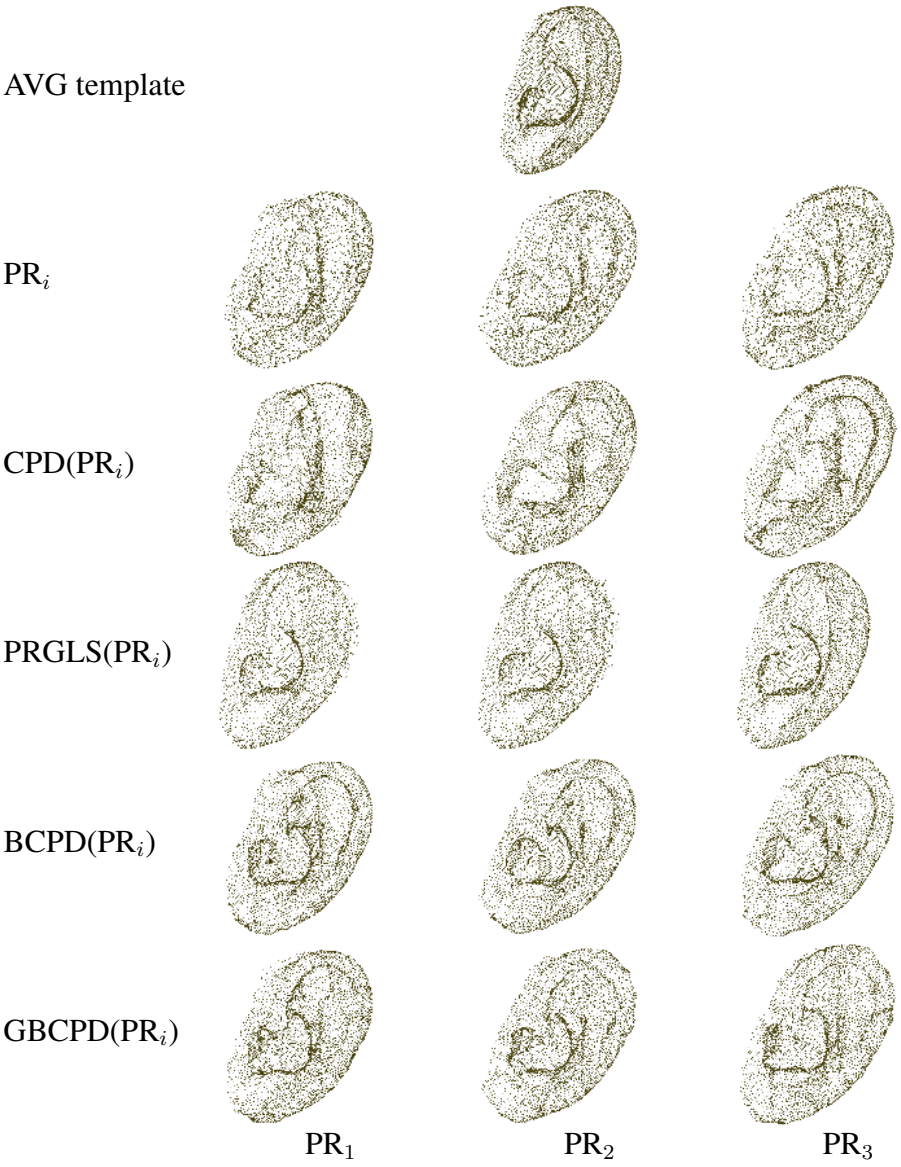


Figure 3.12: Registration results with AVG as template. Conventions the same as in Fig. 3.9.

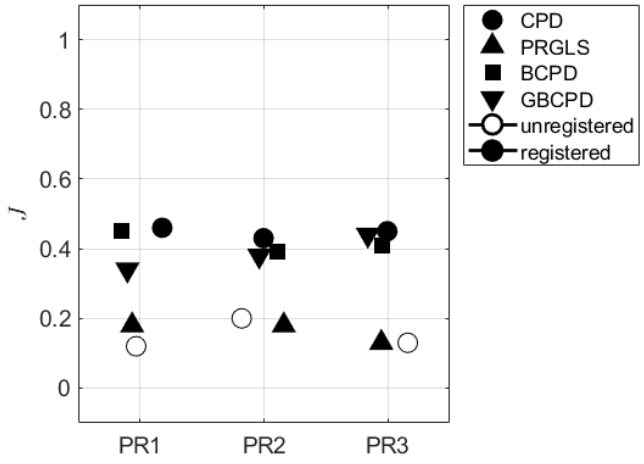


Figure 3.13: The effect of personalisation and robustness with AVG as template and three PRs of NH5 as targets. Convention as in Fig. 3.8.

3.4.3 Template III: non-individual ground-truth

An additional condition was tested, that is, using the individual GT point cloud from *a different listener*, NH131, as template, which was registered to the three PRs of NH5. Only BCPD was used because it seemed to be the best-performing algorithm so far. The mesh of listener NH131 was used because it showed the lowest J compared to the GT of NH5, thus, seems to share the least personal details across the available GT templates.

First, the PRs were resampled to the same number of points as that of NH131, and then BCPD was applied. Figure 3.14 shows the involved point clouds: The template (NH131), the three PR targets, and the results from the registrations. All registered point clouds show the personal details of NH131 without any obvious artefacts. Additionally, they seem to provide a good basis for the calculation of plausible pinna manifolds. Figure 3.15 shows the corresponding Js before and after registration. For comparison, it also shows Js obtained from registering the templates NH5 (see Sec. 3.4.1) and AVG (Sec. 3.4.2). Before the registration, the similarity between the NH131 and the PR-based clouds of NH5 was low, with Js at 0.17, 0.18, and 0.16, respectively. After the registration, all Js increased to 0.44, 0.38, and 0.41, respectively. The similar Js across the three registrations indicate that BCPD was able to adapt the NH131 template to the personal details of the NH5 targets to some extend. When compared to the registration with other two templates, the registration results were similar, indicating that registration results using BCPD were independent of the used template.

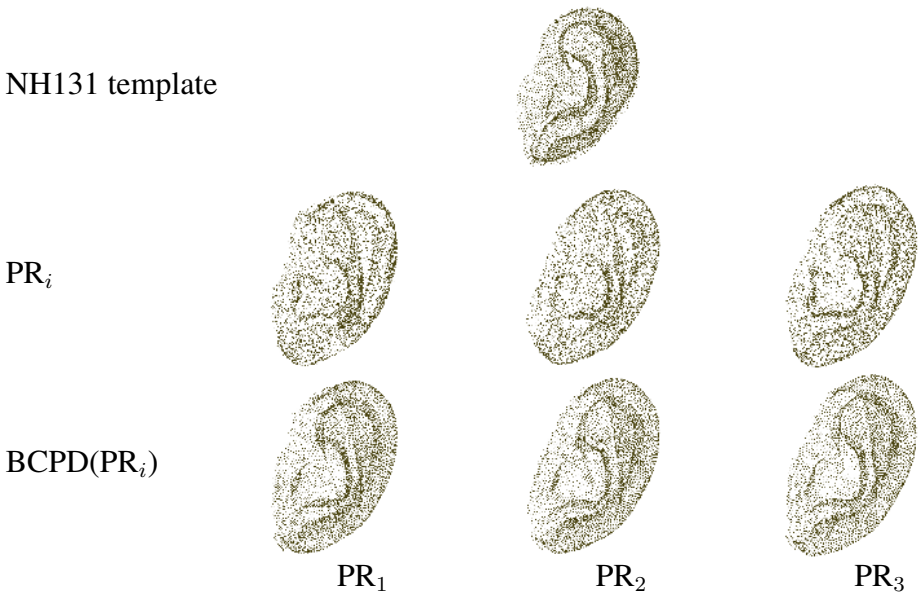


Figure 3.14: Registration results with NH131 as template. Convention as in Fig. 3.9.

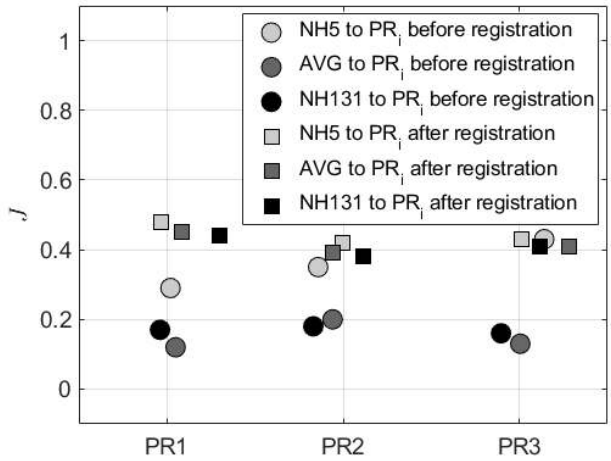


Figure 3.15: The effect of robustness with BCPD applied on various point clouds as template and three PRs of NH5 as targets. J are shown before (circle) and after (square) registration. Templates: NH5 (light grey), AVG (dark grey), NH131 (black).

Generally, the investigated NRR algorithms overall showed lower Js when registering PR point clouds than when registering GT point clouds. This was not surprising, since the PR-based geometries required the algorithm to be robust against a combination of PR-inherent artefacts with varying severity.

For BCPD, the template pinna did not have a strong effect on the registration result. BCPD also yielded the highest J for the majority of the registrations, independent of the target quality, i.e., GT or PR.

3.5 Closing holes using Geodesic-based coherent point drift

In Sec. 3.2, holes in general were neglected as part of the systematic error analysis, because holes depend on the preprocessing and the reconstruction algorithm, and a systematic analysis of this would have exceeded the scope of this thesis. However, in Sec. 3.3, GBCPD was the algorithm that yielded highest Js in the highly reduced target point cloud size conditions, which could be interpreted as an ability to close holes – or, put differently, to preserve the surface of the template despite holes being present in the target. Additionally, if the GBCPD takes the relationship between vertices into account, the chance that points in the front and points in the back of the shape would switch direction – as observed in Sec. 3.3 and 3.4 – is minimised.

3.5.1 Methods

In this investigation, GBCPD [181] registered the AVG pinna [40] to the PR of listener NH789. The exact PR acquisition method was unknown, which was on purpose in order to bring the proof of concept into one real-world application.

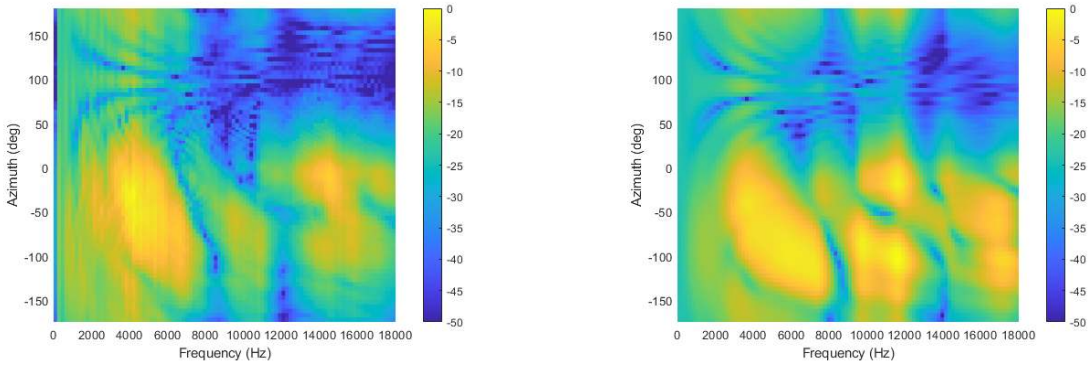
The registered (left) pinna was attached to a dummy-head mesh (Neumann KU100), and the HRTF was calculated using Mesh2HRTF [76] v1.0. For the psychoacoustic evaluation, a model for sagittal-plane sound-localisation [67] from the *Auditory Modeling Toolbox* v1.2 [10] was used, and the quadrant error rate and polar RMS error were calculated [44]. Three conditions were investigated using the auditory model: The acoustically measured HRTFs of listener NH789, the numerically calculated HRTFs of AVG registered to a PR of listener NH789, and the acoustically measured HRTFs of a dummy head (KEMAR with large ears). All HRTFs were transformed into the time domain and converted to directional impulse responses (DIR) using `SOFAhrtf2dtf` from the *SOFAtoolbox* [6] with the default parameters. The auditory model's sensitivity was calibrated such that selecting the acoustically measured individual HRTFs as template and target HRTF yielded a quadrant error rate of approximately 8%.

3.5.2 Results and Discussion

In the geometric domain, the registered template presents a continuous surface with no holes in it, which is a strong advantage over the PR and it means that individualised HRTFs can be calculated. The "rippling" effect in the geometry could have two reasons: Either, the parameters of GBCPD might not have been selected optimally, or, more likely, the template should have been a GT from an existing pinna instead of AVG, because of the stronger similarity of curvatures in template and target. Figure 3.16 shows the template, target, and registered template meshes. Note that the number and size of holes varies strongly across the shape in Fig. 3.16b.

(a) (b) (c)

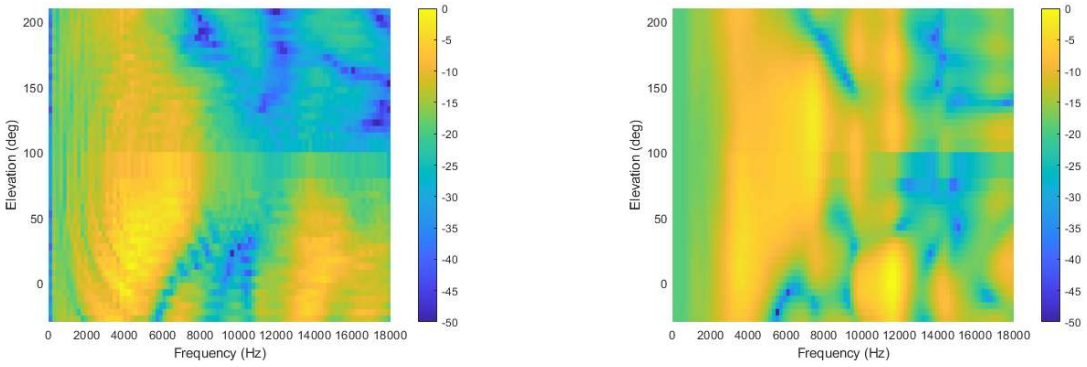
Figure 3.16: Registration of **(a)** WiDESPREaD template (AVG) to **(b)** the target NH789 acquired via photogrammetry using geodesic-based Bayesian coherent point drift (GBCPD), resulting in **(c)** the registered result.



(a) (b)

Figure 3.17: Right-ear HRTF magnitude spectra in the horizontal plane of **(a)** listener NH789 (acoustical measurement) and of the **(b)** GBCPD-registered AVG (numerical calculation).

Figures 3.17 and 3.18 show the measured and calculated right-ear HRTFs in the horizontal and median plane, respectively. The first notch of the ipsilateral region in the horizontal plane of the calculated HRTF is close to the measurement. However, for frequencies above 10 kHz, the calculated HRTF shows a high deviation from the measurement, likely due to the ripples in the registered template. For the median plane in the calculated HRTF, the first notch at around 6 kHz was close to the measurement for elevation angles of -30 to $0°$, but, again, for frequencies above 10 kHz, the calculated resonances for low elevation angles deviated from the measurement, most probably also due to the ripples in the registered template. The jittered geometry around the helix, scapha, and fossa triangularis could be the reason for the larger errors in high frequency bands (see Fig. 3.16c). The first notch of the ipsilateral region was higher in the calculated HRTF than in the measured one. This may be because of the smaller cavum conchae region in the GBCPD-registered pinna, i.e., the antihelix of the template has not been registered sufficiently accurate to the target.



(a) (b)

Figure 3.18: Right-ear HRTF magnitude spectra in the median plane of **(a)** listener NH789 (acoustical measurement) and of the **(b)** GBCPD-registered AVG (numerical calculation).

Table 3.3 summarises the results in the psychoacoustic domain using the acoustically measured individual HRTFs as internal template. The target differs in each row: the target in the first row was the acoustically measured HRTF. The target in the second row was the numerically calculated HRTF from the template registered to the PR point cloud using GBCPD. The target in the third row was the acoustically measured HRTF from a KEMAR dummy head [182].

Table 3.3: Effect of GBCPD: Quadrant error rates and polar error for listener NH789 in a simulated sound-localisation experiment. Internal template HRTF in all conditions was the acoustically measured individual one.

	quadrant error rate [%]	polar error [°]
acoustically measured (ref)	8.05	32.98
GBCPD	52.09	52.26
KEMAR	40.39	51.70

Results from Ziegelwanger et al. [78] suggested negligible perceptual differences between numerically calculated HRTFs and acoustically measured HRTFs. However, their numerical calculations were based on ground-truth meshes, which were unavailable in the current investigation. The high sound-localisation errors could be rooted in the difference between numerically calculated and acoustically measured HRTF, or in the fact that the auditory model is sensitive to small spectral differences. The mismatch of the notches and peaks between 8 and 12 kHz in the median plane may heavily influence the estimated sound-localisation performance in the model [183, 21].

In conclusion, GBCPD was indeed able to close the holes in the target, i.e., the registered template preserved the surface of the template. However, the registered template contained ripples which influenced spectral cues above 8 kHz. In a simulated sound-localisation experiment, acoustically measured HRTF of a KEMAR dummy head showed a lower quadrant error rate than the numerically calculated HRTFs of the registered template. These results suggest that using acoustically measured dummy-head HRTFs yields

a better sound-localisation performance than using numerically calculated individualised HRTFs from a pinna with a rippled surface.

3.6 Filtering PR targets by point confidence levels

Despite BCPD being chosen as the best-fitting NRR algorithm for pinna registration, GBCPD yielded the best registration results for different target sizes and was able to yield a closed surface in the previous investigation. Evidently, GBCPD may be able to close holes that have been introduced in the target point clouds using additional information from the PR. In the high-quality PR approach, the software can output additional information for each point, i.e., a "confidence level". The confidence level for each point corresponds to the number of stereo images used to estimate the point's position. The more stereo images used, the higher its confidence level. Confidence levels are not normalised in the software, so the maximum strongly varies between PRs of the same ear [184]. Figure 3.19 shows three PR instances of the same listener and corresponding point confidence levels. Note the different maxima of 28, 57, and 74, respectively and that the confidence levels in this figure are displayed as absolute values; For the remainder of the investigation, normalised confidence levels (between 0 and 100%) were used.

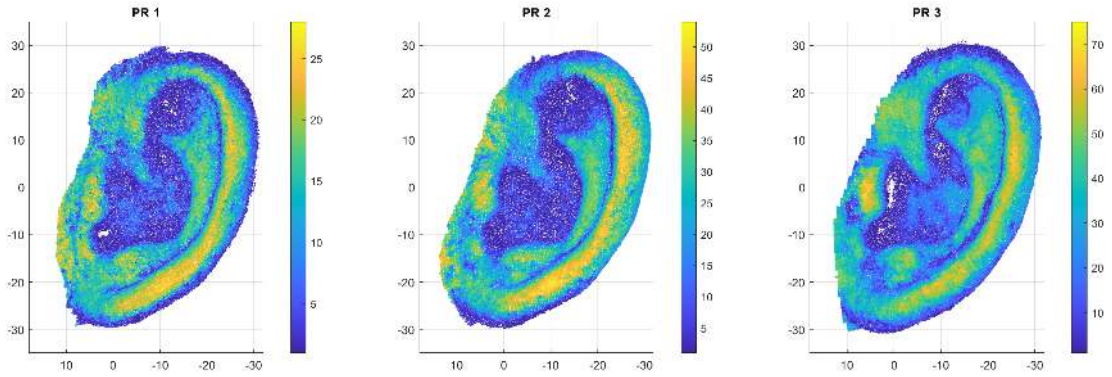


Figure 3.19: Confidence levels obtained from the photogrammetric reconstruction of a pinna. The colourbar shows the confidence level, that is, the number of depth maps of image pairs which where used to reconstruct the respective point in 3D.

3.6.1 Methods

In the used software (Metashape version 1.6.5, Agisoft LLC), "confidence levels" describe how many depth maps of image pairs were used in the reconstruction of the corresponding point in the point cloud. This additional value for each point was utilised in order to introduce a threshold to the registration. Regarding the PRs in this investigation, 54, 77, and 125 photos of the left ear of NH5 were taken around a vertical hemisphere, details of which are described in Sec. 3.2 [123]. The resulting point clouds yielded confidence

levels of up to 28, 57, and 74, respectively. The confidence levels were normalised separately for each PR set such that the maximum number of depth maps used corresponds to the respective maximum confidence of 100%.

As a registration template, the high-quality GT point cloud from NH5 was used [114], and GBCPD was used to register the template to the targets. As targets, three dense PR point clouds of the *same* listener were used which contained the information of confidence levels. In order to systematically investigate the influence of various confidence levels, the target point clouds were filtered with confidence level thresholds at 25, 40, 50, 60, and 75%, including a control condition of 0%, i.e., including all points. The level thresholds range from the complete number of points to a drastically reduced number of points.

3.6.2 Results and Discussion

Figure 3.20 shows the drastic change in point cloud size for the three PR instances of one listener. Despite PR3 showing the highest confidence level maximum, PR1 was chosen for this investigation, because especially for the higher filtering stages, PR1 had the highest amount of preserved points. Note that the lowest filtering threshold already discarded approximately 60 % of the points in the target.

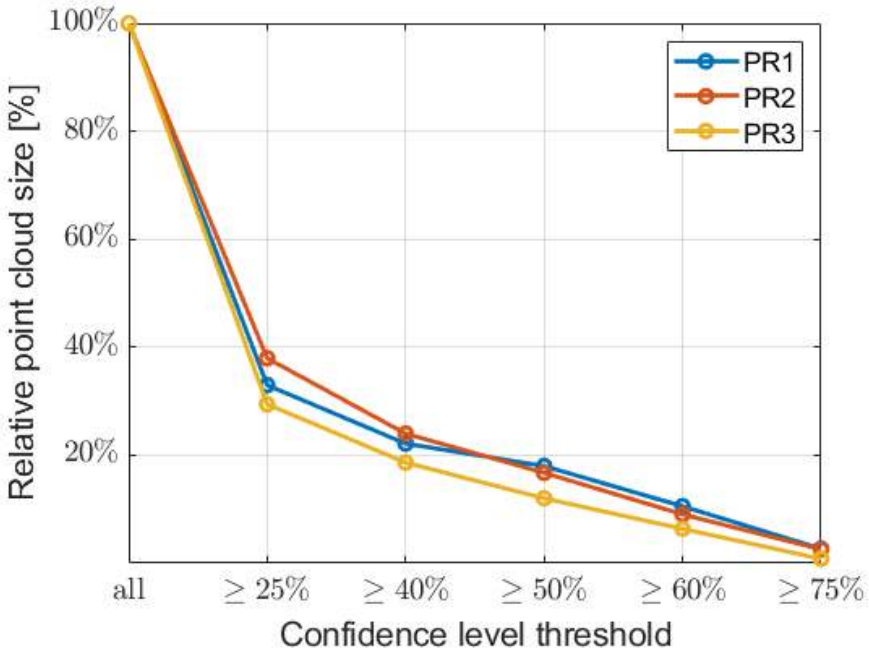


Figure 3.20: The number of points in a point cloud when filtered by confidence level.

Table 3.4 shows the point cloud size for three PRs and the rapid decrease when points only above various confidence level thresholds were considered, and Fig. 3.21 shows the resulting thresholded point clouds. Note that the point clouds including confidence levels below 40 % contain more information about the shape than point clouds including confidence levels above 40 %, which resemble almost a 2D contour. For the confidence threshold

above 75 %, the resulting point cloud does not even represent the bounding ellipse of the individual ear, however, this condition was still included because I hypothesized there was a chance that the shape of the template would be preserved and only the global size was registered anisotropically.

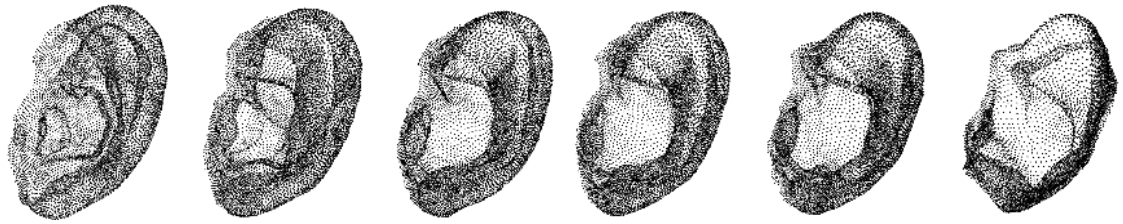
Table 3.4: Point cloud size change with applied filtering by various confidence-level thresholds.

Threshold	PR1	PR2	PR3
$\geq 0\%$ (ref)	1,075,135	860,035	833,082
$\geq 25\%$	354,484	326,452	245,212
$\geq 40\%$	238,134	206,405	155,172
$\geq 50\%$	193,340	143,898	100,077
$\geq 60\%$	113,334	77,863	52,992
$\geq 75\%$	28,752	21,925	6,736

(a) all (b) $\geq 25\%$ (c) $\geq 40\%$ (d) $\geq 50\%$ (e) $\geq 60\%$ (f) $\geq 75\%$

Figure 3.21: Target point cloud (from PR) filtered by relative confidence thresholds for above 0 (all points), 25, 40, 50, 60, and 75%, respectively.

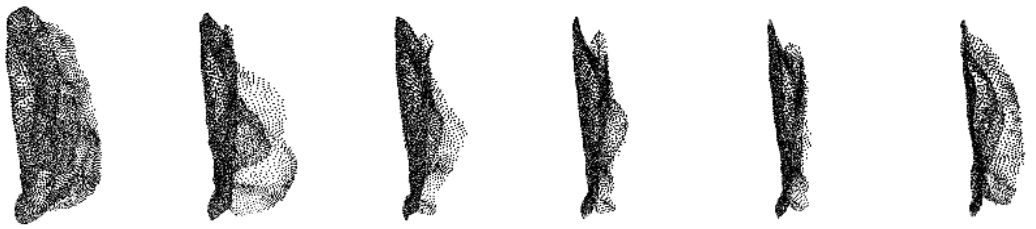
Figures 3.22 and 3.23 show registered templates to the respective confidence-level thresholded point clouds from two angles. The geometry of the registered templates seems to "flatten" for confidence levels above 40%, likely because of the rapidly decreasing depth detail in the target point clouds. Also already at a threshold of 25 %, e.g., the Crus helicis shows a slimmer width in the registered template.



(a) all (b) $\geq 25\%$ (c) $\geq 40\%$ (d) $\geq 50\%$ (e) $\geq 60\%$ (f) $\geq 75\%$

Figure 3.22: Registration of NH5 to PR1 using GBCPD for several relative confidence level thresholds from the front.

GBCPD was expected to move the template towards the target with its global pinna shape intact. This is true only for the unprocessed point cloud, i.e., if no threshold introduced



(a) all (b) $\geq 25\%$ (c) $\geq 40\%$ (d) $\geq 50\%$ (e) $\geq 60\%$ (f) $\geq 75\%$

Figure 3.23: Registration of NH5 to PR1 using GBCPD for several relative confidence level thresholds from the side.

(Fig. 3.21). Already with a confidence level threshold of 25%, i.e., leaving out points that are below the lowest 25% of the global confidence level, the registered template was not sufficiently accurate in order to calculate plausible HRTFs. GBCPD seemed to interpolate a flat surface over regions of missing points, especially for regions with large spatial variations, i.e., the crus inferius anthelicis, the crus superius anthelicis, and the helix rim.

In conclusion, discarding points with low confidence did not yield better registration results despite GBCPD being able to close holes in human pinna point clouds. GBCPD was applied to a high-quality template point cloud in order to register it to three target point clouds acquired with a high-quality PR. Even with a confidence level threshold of 25%, the quality of the registered point clouds decreased drastically and with increasing thresholds, the pinna transformed more and more into a flat surface.

3.7 Conclusions

BCPD was the best NRR algorithm for pinna registration because it yielded the highest geometrical similarity between the registered template and the target by means of the Jaccard index J, as compared across four algorithms and a variety of conditions.

In Sec. 3.3, the effect of systematic distortions of the target geometry was investigated after identifying and classifying them in Sec. 3.2. This investigation resulted in PRGLS as a candidate being most robust to noise and outliers but sensitive to differently sized point clouds, and yielded plausible pinna surfaces even when the target was reduced to 5 % of its original size. BCPD and GBCPD were the best candidates overall.

Then, in Sec. 3.4, the effect of the template was investigated while the targets were both GT and PR point clouds. This investigation resulted in BCPD showing the highest geometric similarity by means of J and plausible point clouds from which meshes could be constructed for numerical HRTF calculations. The next-best algorithm was GBCPD.

In Sec. 3.5, it was shown that GBCPD was able to maintain a closed pinna surface even if holes in the target were present. GBCPD was the applied algorithm, because it was robust to point cloud size differences.

In Sec. 3.6, the effect of filtering the PR targets by point confidence levels was investigated. It was shown that already leaving out points with confidence levels smaller than 25 % led to significantly worse registration results. In summary, GBCPD can be used for PR targets with holes, but not if holes are created purposefully by applying confidence level thresholds.

The mixed results from the non-parametric registration proposal and the open question of having a template *surface* be deformed and registered to a target motivated me to approach the registration problem from a parametric perspective, which is described in the following Chapter.

Chapter 4

Proposal II: Parametric pinna model BezierPPM

A different way to deform a pinna is to parameterise the surface and control the shape transformation via a set of parameters, leading to a significant decrease of the degrees of freedom in the registration. Describing the pinna geometry with a finite set of anthropometrically-informed parameters is by no means trivial. Investigations so far resulted in data collections of anthropometric parameters, pinna meshes, HRTFs, and various combinations thereof [12, 185, 74, 186, 118, 187, 188, 103, 189, 121, 45, 116]. Based on these collections and research emphasizing the convex regions of the pinna [16, 142, 17], several parametric models of the pinna geometry have been proposed. Some of which are based on anthropometric-agnostic approaches, such as the principal component analysis (PCA) and active shape models (ASM) [133], e.g., [40, 134, 132]. Others are based on anthropometric data, e.g., [135, 32, 136]. In contrast to pinnae parameterisation, other human anthropometric data, e.g., bones [190] or faces [164, 191, 192, 193, 194], are more extensively studied in the fields of image processing and medical image processing. Inspired by the application to faces in [195, 196], one of the first idea for the parameterisation was to develop a model that would imitate the growth process of a pinna [197, 198] based on the antenatal development of the pinna. However, because there are several different interpretations of which parts in the pharyngeal arches develop into which part in the pinna, this trajectory was discarded. The parametric solution proposed in this chapter is based on the animation of digital human characters. Typically, in animation, characters are created in a T-pose, standing upright with their arms laterally spread out. Then, the character gets "rigged" [199, 200], which means that an armature is created by fitting an imaginary skeleton to the mesh. Once the armature is connected to the mesh, the movement of a character can be controlled by the parameters of the armature. Additionally, predefined local changes can be applied to a mesh. In summary, when personalising a character in a game, the user has a few parameters that they can tweak, such as head size or the distance between the eyes. With these deformation methods stemmed from animation, I implemented a parametric pinna model (PPM) [201, 202, 203, 204]. This PPM can be applied to either register the underlying template pinna to a target pinna or to deform the underyling template pinna in order to generate plausible human pinnae. The former

application is the focus in this chapter, the latter application can be used to generate large pinna datasets, e.g., in order to train machine-learning algorithms on. In this chapter, I investigate the applicability of the PPM and establish a proof of concept on ground-truth target meshes.

4.1 Anatomical properties of the outer ear

The average thickness of a fully developed human pinna is approximately 3 mm [173]. Under the skin are cartilage and tissue, albeit not evenly distributed throughout the shape. Figure 4.1 shows a schematic drawing of the cartilage shape underlying the tissue and skin (from Fig. 2.2 in [205]). The cartilage is not connected directly between the tragus and the helix, i.e., underneath the insicura anterior, and between the antitragus and the helix. This might be relevant for different materials in numerical calculations of HRTFs which consider the entire volume, e.g., the finite element method [206]. Note that, in this thesis, the boundary element method is applied for numerical HRTF calculation, and the surface is assumed to be sound hard [97, 98], i.e., different materials are not considered, as Ziegelwanger et al. showed that this boundary condition does not seem to contribute to sound-localisation performance in the end [78].

Figure 4.1: Cartilage (blue) underlying the tissue and skin (grey) in the pinna.

The pinnae start to form in the 5^{th} week of development on the lower neck and move towards their final lateral position until the 20^{th} week of development, when the lower jaw forms. Scientific consensus has been reached that the pinna develops from the so-called six "hillocks of His" on the first and second pharyngeal arch [207, 208]. However, the exact transformation of which hillock turns into which part of the ear is still an open question [207, 209, 210, 211]. Figure 4.2 shows an excerpt of interpretations of the hillocks' transformation from a review article [207] and highlights the discrepancy within the last 140 years from several investigations [212, 213, 214, 215, 216, 209, 217]. In this article, the authors conclude that observations of the hillocks' geometric movement on the pinna surface may not be sufficient to understand the formation of the pinna. In a more recent review, Veugen et al. concluded that the missing link is to be found in the two nerves underlying the pharyngeal arches: The development of the trigeminal and facial nerve and

their respective branches suggested that, likely, the tragus is formed by the mandibular arch, and the protruding rest of the pinna is formed by the hyoid arch [218, 219, 208].

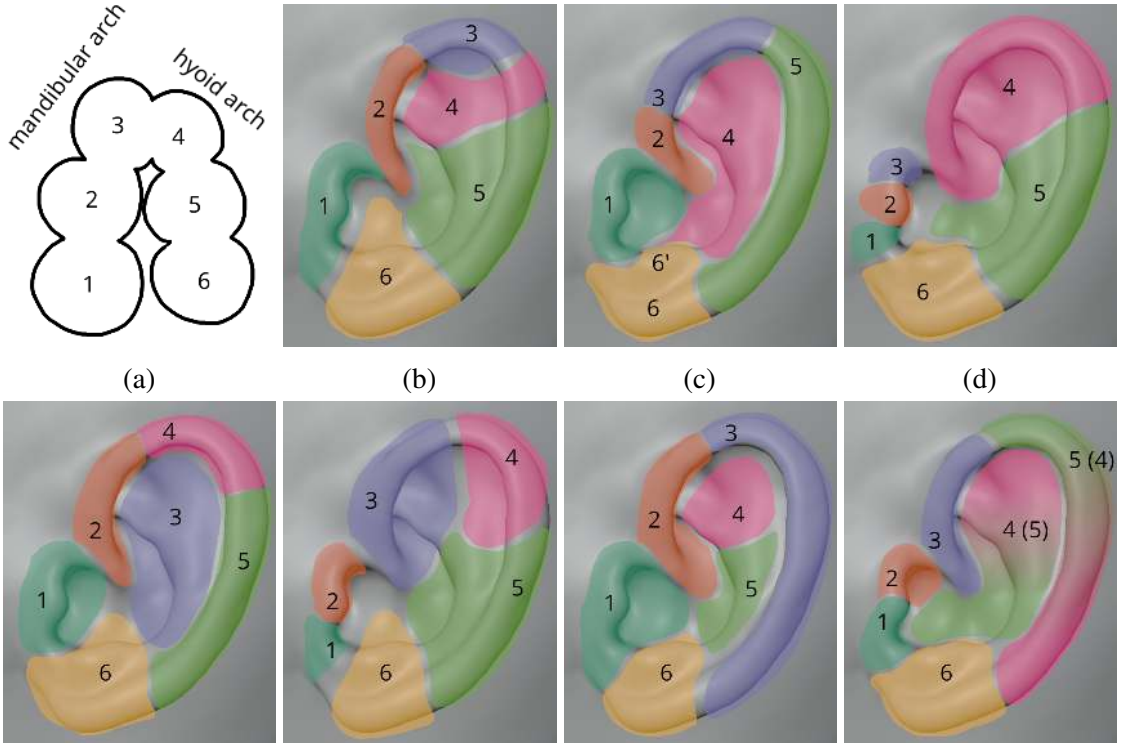


Figure 4.2: Historical overview of various interpretations of the movement of hillocks of His. **(a)** Embryonic hillocks, **(b)** His (1885), **(c)** Streeter (1922), **(d)** Wood-Jones and I-Chun (1934), **(e)** Midera (1982), **(f)** Davis (1987), **(g)** Karmody and Annino (1995), **(h)** Park and Roh (1999).

In summary, the literature search towards the antenatal development of the pinna resulted in discarding further efforts towards modelling pinna growth, because no connection could be found on a morphological level. While the approach of modelling the pinna growth was discarded, the antenatal development contributed to the idea to separate the pinnae shape into *concave* and *convex* structures.

4.2 The BezierPPM

Inspired by the antenatal development of the pinna, its parameterisation was divided into covering concave and convex parts of the shape. In a way, the convex parts were intended to deform the shape in order to roughly match an individual pinna. The concave parts were intended to change the shape in a more subtle way, because these parts are important for localisation [16, 142, 17]. The former is implemented in Blender as so-called *armature* consisting of *bones*, and the latter is implemented as so-called *shape keys*.

In order to emphasize the underlying method that mainly transforms the template ear in the PPM and in order to separate it from other PPMs, this PPM and its implementation is further denoted as *BezierPPM*. The source code for the parametric pinna model based on Beziér curves is made available in an online repository[1].

An armature was defined along the countours of the pinna. For each part that had to be considered separately, a Bendy bone and two control bones at the start and end of it were placed, respectively. Shape keys can be described as a weighted local transformation of a mesh. For example, if the depth of the cavum conchae should be increased, the weight of its shape key was increased, and if the cavum conchae should be made flatter, the weight of its shape key was decreased. Note that the number of bones and shape keys changed significantly across the version history of the model. In the methods section, only the latest version is presented, but the preceding versions of the model are discussed in this chapter.

4.2.1 Mathematical description

Let $\mathbf{X}_{PPM} \in \mathbb{R}^{3\times N}$ be an affine transformation applied to a template mesh $\mathbf{X}_0$ described as

$$\mathbf{X}_{PPM} = \mathbf{X}_a\left(\mathbf{X}_0, \mathbf{X}_s\right), \tag{4.1}$$

with $\mathbf{X}_a$ being the affine transformation performed by the armature, and $\mathbf{X}_s$ being the transformation performed by shape keys.

Despite the development process of first defining the Beziér curves before defining the shape keys, mathematically, it makes more sense to first look into the transformation performed by shape keys, as the transformation performed by the armature is a function of $\mathbf{X}_s$. Additionally, this is the way the shape transformation is implemented in Blender.

4.2.2 Shape keys

Shape keys can be described as a weighted transformation that affects each vertex in the mesh:

$$\mathbf{X}_s = \sum_{j=1\ldots14,17} s_j \frac{\mathbf{B}_j - \mathbf{X}_0}{2} + \sum_{j=15,16} s_j\left(\mathbf{B}_j - \mathbf{X}_0\right) \tag{4.2}$$

where $\mathbf{X}_s$ describes the resulting mesh after all $N = 17$ shape keys have been applied to the template mesh $\mathbf{X}_0$, $\mathbf{B}_j$ describes the deformed limit mesh of $\mathbf{X}_0$ for the j^{th} shape key, and w_j describes the weight for the j^{th} shape key. Because the weights of shape keys 1-14 and 17 are limited between $[-1, 1]$, the factor ½ is introduced. The weight boundaries of

[1] `https://github.com/Any2HRTF/PPM/`

shape keys 15 and 16 are $[0, 1]$. Note that, typically, the limit mesh $\mathbf{B}_j$ deforms only a local region of the mesh.

4.2.3 Armature

In the parametric pinna model based on Beziér curves (BezierPPM), the armature deforms the convex regions of the pinna. It consists of nine cubic and one linear Beziér curve. A cubic Beziér curve is defined as a function $\mathbf{c}_i(k) : [0, 1] \mapsto \mathbb{R}^3$:

$$\mathbf{c}_i(k) = (1-k)^3 \mathbf{p}_0 + 3(1-k)^2 k\mathbf{p}_1 + 3(1-k)k^2\mathbf{p}_2 + k^3\mathbf{p}_3, \tag{4.3}$$

with $i = 1 \ldots 9$ and $0 \leq k \leq 1$, $\mathbf{p}_0$ describes the start point of $\mathbf{c}_i(k)$, $\mathbf{p}_3$ describes the end point of $\mathbf{c}_i(k)$, and $\mathbf{p}_1$ and $\mathbf{p}_2$ describing the curvature of $\mathbf{c}_i(k)$ in an implicit way. A linear Beziér curve corresponds to a straight line between two points - this linear interpolation is used for controlling the global transformation of the pinna and defined as "parent" bone. Figure 4.3 shows the implementation of a bendy bone in Blender next to the underling Beziér curve.

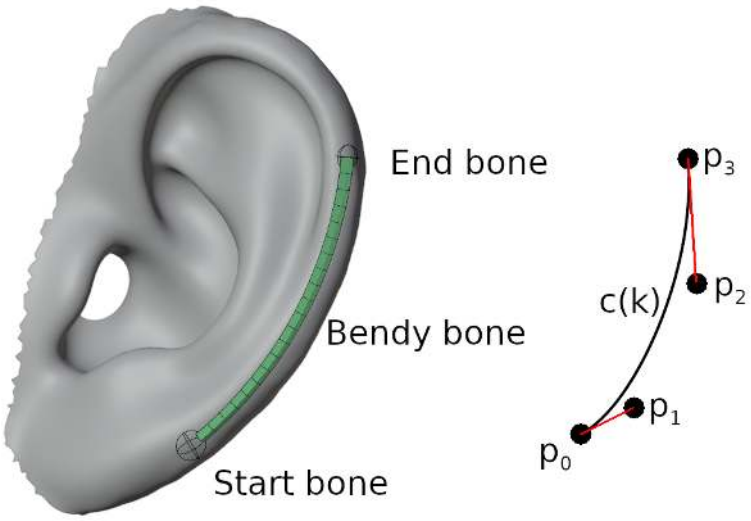


Figure 4.3: Implementation of a bendy bone with control bones in Blender (left) and underlying Beziér curve (right).

Every Beziér curve is linked to local parts of the pinna mesh and can deform the mesh, which corresponds to a linear transformation, $\mathbf{V}_i : \mathbb{R}^{3\times N} \rightarrow \mathbb{R}^{3\times N}$. This linear transformation is applied to the mesh after the shape keys (see Eq. 4.2.1). Thus, the mesh deformation from the armature can be formulated as:

$$\mathbf{X}_a = \sum_{i=1}^{9} \mathbf{V}_i (\mathbf{X}_0 + \mathbf{X}_s). \tag{4.4}$$

Each transformation $\mathbf{V}_i$ consists of weights that control the influence of the displacement of $\mathbf{c}_i$ on each vertex. Similar to the shape keys, each Beziér curve controls vertices in the proximity of its location, i.e., vertices closer to the Beziér curves are affected stronger than vertices further away. Vertices that do not belong to a proximal connected anatomical region, e.g., the Antitragus is not connected to the Crus superius anthelicis, were assigned a weight of 0. These weights are assigned manually by a "weight painting" brush, because of the curvature of many structures and to enable smooth transformations (see also Fig. 8 in [199, 220]). Figure 4.4 shows the Beziér curve $\mathbf{c}_3$ and its transformation heatmap (red denotes a weight of 1, blue a weight of 0).

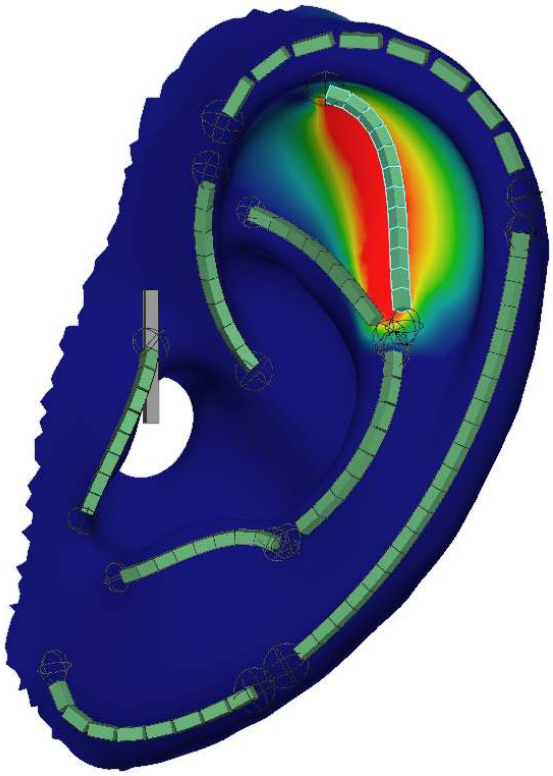


Figure 4.4: Beziér curve $\mathbf{c}_3$ (Helix middle) and its underlying transformation of the region and proximal points as heatmap.

Note that in order to limit the parameter space somehow, the armature's elements are restricted to certain transformations. The parent bone may be translated, rotated, and anisotropically scaled (9 degrees of freedom). Start and end bones (the control points of a Beziér curve) may be translated and rotated, but not scaled at all (6 degrees of freedom each). Bendy bones may be isotropically scaled only (1 degree of freedom each). Summarising the degrees of freedom, the armature is controlled by 126 parameters. Adding the 17 shape keys to it, the BezierPPM in total has 143 parameters that control the deformation when using it manually.

4.2.4 Template

Selecting an adequate template for the registration is important. It is reasonable to assume that the choice of the template pinna influences the parameter space the BezierPPM has to cover. Another unknown variable is the parameter range. Regarding the parameter space the BezierPPM has to span, there is the obvious question of where the parameter limitations of the model – and limitations in reality – are.

In order to symetrically control the shape key weights around zero, an average pinna – further denoted as AVG – was chosen as template underlying the BezierPPM. AVG was taken from 119 scans of individual pinnae acquired with a structured-light scanner [40], and processed in the following way. The pinna was cut out such that the mesh only contained parts of the pinna, without parts of the head. Then, the global properties of AVG, i.e., position, orientation, and size, were set such that they represent average values of height, rotation angle, and flare angle from the CIPIC database [12]. In order to avoid unwanted distortions, the ear canal edge was smoothed using a Boolean filter with a cylinder on the interaural axis. And, lastly, the AVG mesh was resampled using a mesh grading tool from the Mesh2HRTF pipeline [2] [101].

4.2.5 Implementation

The bones were chosen after identifying the structures that change the overall pinna shape, analogous to pencil strokes when drawing contours of a pinna. The shape keys were defined analogous to the remaining areas between contours. Figure 4.5 shows the armature connected to AVG, i.e., the final BezierPPM armature in its resting pose. Tables 4.1 and 4.2 show the Bezér curves and shape keys for the final model, respectively.

Table 4.1: Final Beziér curves for the BezierPPM.

i	Name
1	Lobulus
2	Helix low
3	Helix middle
4	Helix up
5	Antitragus
6	Antihelix
7	Crus superius anthelicis
8	Crus inferius anthelicis
9	Tragus
10	Parent

[2] `https://sourceforge.net/p/mesh2hrtf-tools/`

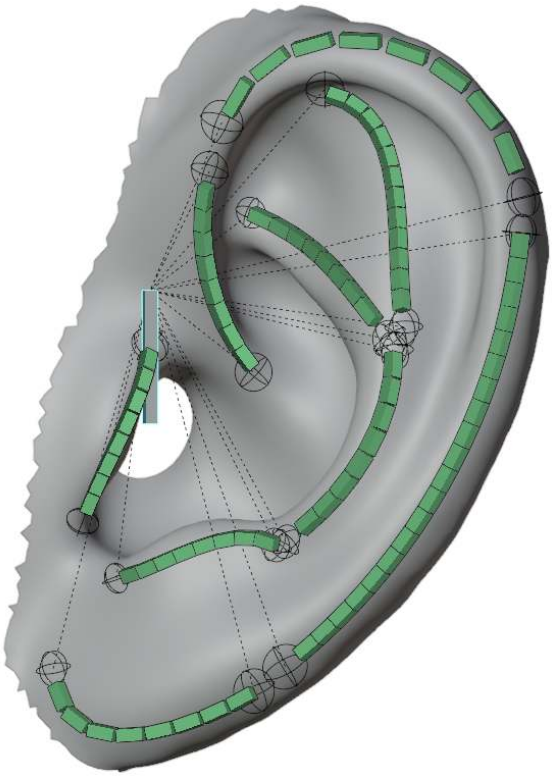


Figure 4.5: Final armature for the BezierPPM.

4.2.6 Registration

In order to register the template to a target pinna using the BezierPPM, the target mesh has to be loaded into the same Blender workspace as the model. In a first step, the global location, orientation and size of the parametric pinna model based on Beziér curves are changed using the parent bone such that the template aligns with the target as close as possible, with a focus on the pinna flare angle. Next, starting from the first Beziér curve, every Beziér curve is aligned to match the convex shape of the target. Then, the shape keys are changed such that the smallest median distance is minimised. In order to refine the alignment, and because the armature and the shape keys influence each other, transforming the armature and adjusting shape key weights is done in an alternating fashion, representing an iterative process. The registration is finished once the 1 mm threshold for the median of ℓ is reached, and if ℓ is lowest in the concave regions. Figure 4.6 shows the template (Fig. 4.6a, the final stage of the registration (Fig. 4.6b and the target (Fig. 4.6c).

Table 4.2: Final shape keys for the BezierPPM.

j	Name	Weight range
1	Antitragus-Crease	$-1 \leq w_1 \leq 1$
2	Cavum conchae-Depth	$-1 \leq w_2 \leq 1$
3	Cymba conchae-Depth	$-1 \leq w_3 \leq 1$
4	Crus helicis-Prominence	$-1 \leq w_4 \leq 1$
5	Upper helix-Depth	$-1 \leq w_5 \leq 1$
6	Middle helix-Depth	$-1 \leq w_6 \leq 1$
7	Lower helix-Depth	$-1 \leq w_7 \leq 1$
8	Scapha-Depth	$-1 \leq w_8 \leq 1$
9	Fossa triangularis-Depth	$-1 \leq w_9 \leq 1$
10	Crus inferius anthelicis-Lower crease	$-1 \leq w_{10} \leq 1$
11	Crus inferius anthelicis-Upper crease	$-1 \leq w_{11} \leq 1$
12	Crus superius anthelicis-Lower crease	$-1 \leq w_{12} \leq 1$
13	Crus superius anthelicis-Upper crease	$-1 \leq w_{13} \leq 1$
14	Tragus-Upper dent	$-1 \leq w_{14} \leq 1$
15	Crus helicis-Upper dent	$0 \leq w_{15} \leq 1$
16	Crus helicis-Lower dent	$0 \leq w_{16} \leq 1$
17	Ear canal-Diameter	$-1 \leq w_{17} \leq 1$

(a) NH5 (template) (b) NH5 registered to NH131 (c) NH131 (target)

Figure 4.6: Pinna mesh of NH5 **(a)** before and **(b)** after registering it to **(c)** NH131.

4.3 Evaluation on ground-truth meshes

In order to evaluate the BezierPPM, it was registered to six target meshes.

4.3.1 Methods

In the geometric domain, the smallest pointwise distance ℓ from Eq. 2.1 and its median were used, i.e., the smallest Euclidean distance between every point of the target pinna and the deformed template pinna. As commonly used so far, a threshold of 1 mm for the median distance [78] was deemed adequate as the criterion to end the registration process. In order to monitor this threshold, the toolbox *PointCloudCompare* was used, which is implemented directly in Blender and calculates the median distance, among other measures[3].

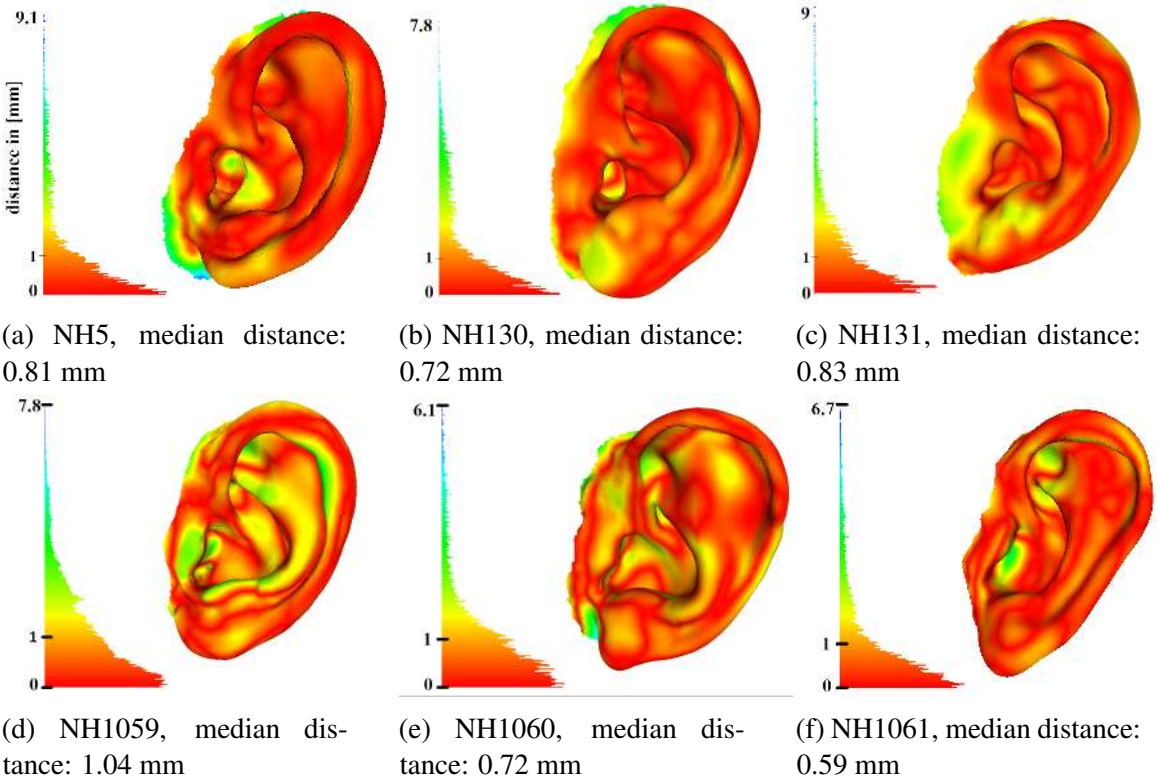
HRTFs from the ground-truth meshes were numerically calculated and used as reference HRTFs. After the alignment using the parametric pinna model based on Beziér curves, the left pinna mesh was stitched back to the ground truth mesh and HRTFs were calculated again. By doing so, it was guaranteed that the only change between the meshes, and subsequently HRTFs, was due to the influence of the left pinna. HRTFs were calculated using the the open-source software Mesh2HRTF[4]. All HRTFs have been converted to DTFs in order to be used with the auditory model. This was done by using `SOFAhrtf2dtf` from the SOFAToolbox v2.1 with default settings.

An auditory model for sagittal-plane sound localisation [67] was used from the AMT [5] [10], and simulated sound-localisation errors quadrant error rate and local polar RMS error were calculated. For the initial and the updated model, I used the AMT version 1.2, and for the final model, I used the AMT version 1.5. Lateral errors were not evaluated, as the change in HRTFs influences monaural spectral cues which have been shown to contribute largely to sound-localisation in sagittal planes and only negligibly to sound-localisation in the horizontal plane [86, 67]. The sensitivities of the auditory model have been set to yield a quadrant error rate of approximately 8 % for the reference DTFs, and the internal template always was the DTF calculated from the reference mesh.

4.3.2 Results

The six target meshes were acquired using two imaging methods, three from a CT scan of a mold [79] and three from a structured-light scan [116]. The former three meshes were already available in the ARI database during the development of the BezierPPM [4], the latter three were added in the evaluation only. For all six listeners, HRTFs were calculated of the reference meshes and reference meshes with the left ear swapped by the BezierPPM-registered ear mesh. Figure 4.7 shows the smallest pointwise distance distribution on the surface of the reference meshes for the six listeners.

[3] `https://github.com/Any2HRTF/PPM/tree/PointCloudCompare/PointCloudCompare`
[4] `https://github.com/Any2HRTF/Mesh2HRTF`
[5] `https://amtoolbox.org/`



(a) NH5, median distance: 0.81 mm

(b) NH130, median distance: 0.72 mm

(c) NH131, median distance: 0.83 mm

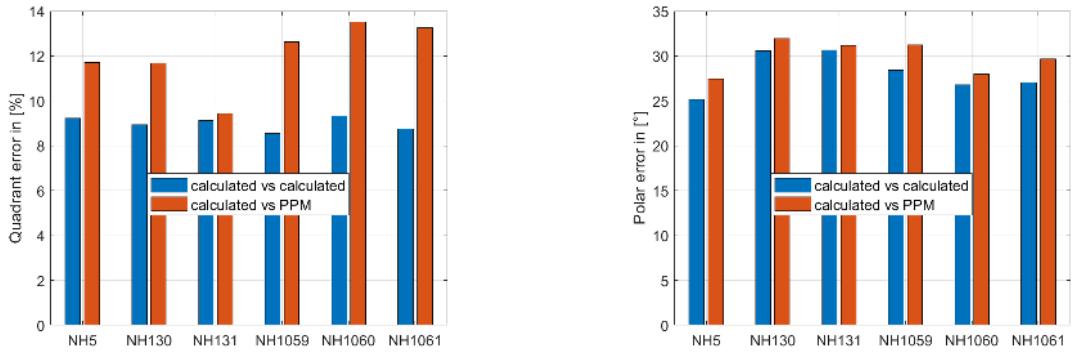
(d) NH1059, median distance: 1.04 mm

(e) NH1060, median distance: 0.72 mm

(f) NH1061, median distance: 0.59 mm

Figure 4.7: Pointwise smallest distances calculated between the BezierPPM-registered mesh and the corresponding ground-truth mesh. Legends to the left of each pinna describe distance distribution in mm, ranging from red (low distance) over green to blue (high distance).

Figure 4.8 shows the quadrant error rate and local polar RMS error. In the parametric pinna model based on Beziér curves-registered condition, the quadrant error rate was higher for the listeners NH1059, NH1060, and NH1061 than for the listeners NH5, NH130, and NH131. Regarding the local polar error, all six listeners performed similar as in the respective reference condition.



(a) (b)

Figure 4.8: Simulated sound-localisation errors **(a)** quadrant error rate and **(b)** polar error, both calculated using an auditory model from the AMT.

4.3.3 Discussion

For all but one registration, a median ℓ of less than 1 mm was achieved. The one remaining registration showed a median ℓ of 1.04 mm, i.e., close to the target threshold of 1 mm. It is evident that the BezierPPM was designed based on the pinnae from the ARI database (top row in Fig. 4.7), as the largest geometrical errors on the pinna occur on the lobulus. Once three pinnae from a different database were registered, the BezierPPM showed higher geometrical errors in concave regions such as the fossa triangularis or the cymba conchae. This suggests that the design of the BezierPPM was limited to available meshes, and the parameter space the model spans can be extended when incorporating a large variety of shapes.

In the psychoacoustic domain, the local errors did not seem to be affected by the geometrical error in the fossa triangularis and cymba conchae. However, the quadrant error rates were in fact higher in the newly added pinnae (bottom row in Fig. 4.7). The higher quadrant error rates can have multiple reasons: First, the specific shape key defining the depth of the fossa triangularis may not cover a rotation of the region. Second, the Beziér curve connected to the crus superius anthelicis might affect the vertices close to the fossa triangularis. And third, given the strong differences between the three new pinna shapes and the – extensively used and heavily influencing the design of the PPM – previously used pinna shapes, the fossa triangularis and the cymba conchae of the three new meshes seem to be just outside the border of the parameter space the PPM spans.

To this end, there are two problems to identify when using the parametric pinna model based on Beziér curves. First, the average ear template is based on a principal component analysis (PCA) of structured-light scans of 119 people, who were randomly selected in a specific area of the world, likely leading to a bias of the “average” in the population. Second, the parameter space of the BezierPPM is unknown. Certainly, the weight painting can be improved by exploring the dependencies between Beziér curves and shape keys that cover the same parts. Finally, in order to support smooth boundaries in the deformation process, cutting the oval edge of the pinna – where it can be attached to a head –

using a Boolean filter with an ellipse would help in the stitching process.

The goal of the parametric pinna model based on Beziér curves proposal was to maximise the overlap of plausbile human pinna shapes with the shapes that can be modelled with the parametric pinna model based on Beziér curves. One advantage the parametric pinna model based on Beziér curves already in its current form has is, despite it being based on an average pinna, the parameter space the parametric pinna model based on Beziér curves spans is greater than, e.g., the parameter space a PCA spans [204]. It is not known whether the chosen parameters would be normally distributed across the human population. However, one possible application of the BezierPPM is not only to register a template to a target, but also to deform the template in a way that a plausible pinna is created [221, 191]. This would prove useful, e.g., to generate large datasets for machine learning algorithms to use, albeit one could argue that a synthetically creasted dataset may not be ideal in order to represent real-world data [222]. This would also - coming back to animation - enable players to customize ears when personalising a character in a game.

4.3.4 Lessons learned during the development

Several changes in the design of the parametric pinna model based on Beziér curves were made during its development, and for the sake of completeness the are mentioned in this discussion.

Template pinna

In the first implementation of the parametric pinna model based on Beziér curves [201], the underlying template mesh was a full head including both ears from a subject of the ARI database [114].

However, the parametric pinna model based on Beziér curves should only the left ear, and there were sharp edges on the boundary between the ear and the head when deforming the armature. Because the used mesh was from a real person, it already showed individual detail which was likely to not be average for the human population.

These disadvantages were the reason a pinna geometry detached from the head was used in order to avoid rough edges on the boundary. This is the reason the average mesh of 119 individual structured-light scans was used for further implementations of the parametric pinna model based on Beziér curves [40]. Using an average pinna helped in defining the shape keys around the average such that a wider parameter space could be covered. Since the first iteration of the parametric pinna model based on Beziér curves [202], the average pinna of the WiDESPREaD database was used as template pinna (AVG). Figure 4.9 shows the three different pinnae selected as template for the parametric pinna model based on Beziér curves.

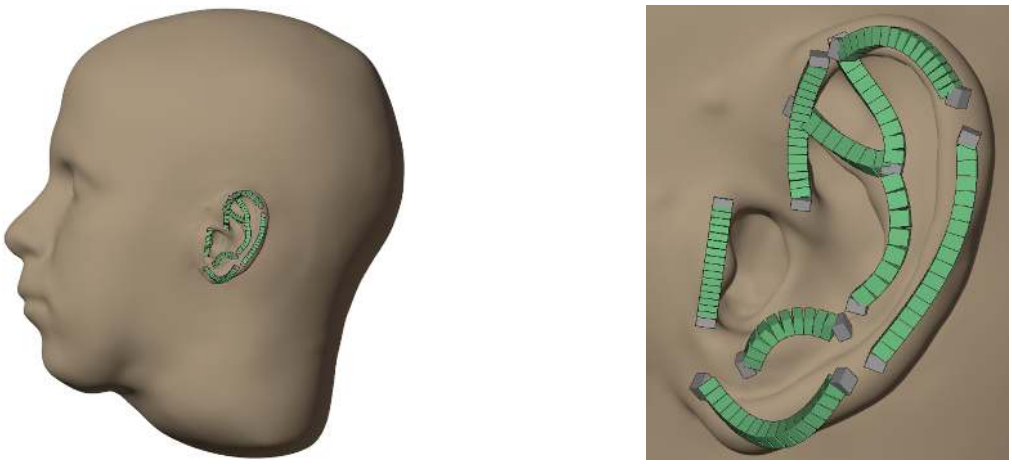
(a) (b) (c)

Figure 4.9: **(a)** For the first implementation of the BezierPPM, a full head mesh including both pinnae was used, despite the deformations being applied to the left ear only. **(b)** Updated BezierPPM. **(c)** Final BezierPPM.

Initial model

In the initial implementation of the parametric pinna model based on Beziér curves, the armature consisted of nine bones and 32 distinctive shape keys, and the template was an entire head mesh albeit the BezierPPM influenced vertices only in the pinna region. In order to explore the implementation, the model was applied to the left ear only, as applying it to a right ear would be the same process but mirrored. In the appendix, Tab. A.1 lists the nine bones that formed the armature, Tab. A.2 gives an overview of the shape keys. In that table, the weight range w_j is given for each shape key. The weights for some shape keys are within $[0, 1]$, as the general formulation intended (see Sec. 4.2.1 for the mathematical description). However, for some other shape keys it was more useful being able to deform the mesh in "both directions", i.e., increase or decrease the depth of the Cavum conchae in the range of $[-1, 1]$. Because of the limitations of the template mesh, i.e., the mesh being intersected with itself, $[-1, 1]$ was not always possible, and thus the outmost possible value was defined as the limit. Note that, for the Fossa triangularis depth, a range of even $[-2, 1]$ was possible to define.

Figure 4.10 shows what the armature looked like connected to the ear, in this case a high-resolution head mesh including the pinnae with an average edge length of 1 mm [79].



(a) (b)

Figure 4.10: First proposal of the parametric pinna model based on Beziér curves armature. **(a)** Attached to a full head mesh. **(b)** Pinna region in focus.

Figure 4.11 shows the distribution of the smallest pointwise distance on the surface of the target showing a median value of 0.5 mm.

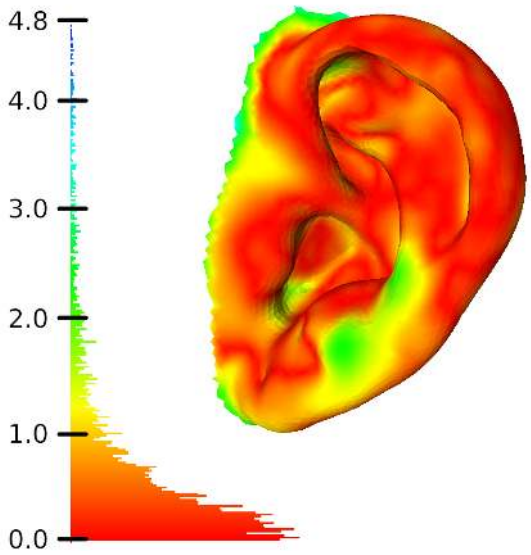


Figure 4.11: Smallest pointwise distance distribution, showing a median of 0.5 mm.

Figure 4.12 shows the localisation errors "quadrant error rate" (in %) and "local polar RMS error" (in °) for several stages of the registration process. On the far left, the localisation performance of the target (NH131) is shown as the baseline. On the far right, the localisation performance of the template (NH5) is shown for reference. Then, localisation performances from second from right towards the second from left: NH131 given NH5's ears; NH131 given NH5's ears registered manually given two 2D reference photos (from the front and lateral); NH131 given NH5's left ear registered to the mesh of NH131 using a preliminary version of the initial BezierPPM; NH131 given NH5's left ear registered to

the mesh of NH131 using the initial BezierPPM.

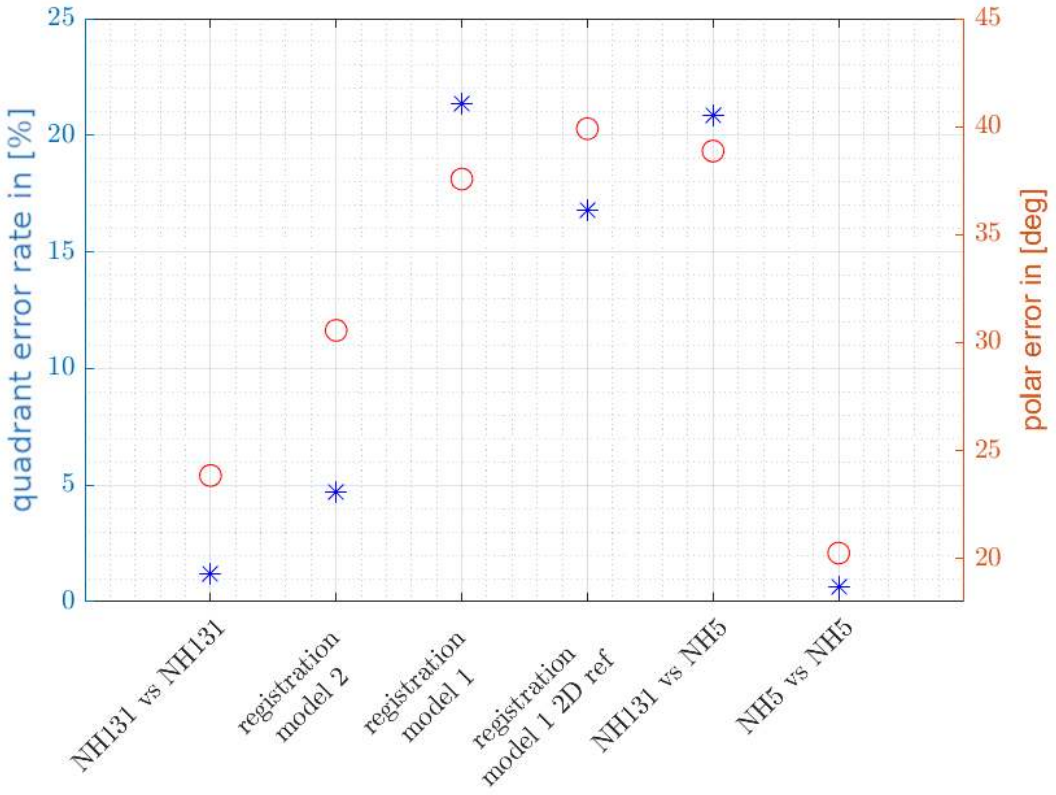


Figure 4.12: Quadrant error rate in % and local polar RMS error in ° for baseline (far left) and reference (far right), and several stages of registration (from second from right to second from left).

Updated model

Several limitations were encountered during the evaluation of the first initial version of the parametric pinna model based on Beziér curves. First, the border around the pinna did not show a smooth transition to the rest of the head, which resulted in extensive smoothing after registration. Figure 4.13 shows a deformed template using the BezierPPM resulting in discontinuous borders. This was solved by using a cutout pinna, without a head attached, and attach it to a head after the registration.

Second, the template pinna for the initially proposed BezierPPM was an individual pinna, already showing individual details that were above or below average values. For example, the depth of the helix tunnel (shape keys 5-7) or the creases around the fossa triangularis (shape keys 10-14) could only be changed in one direction, i.e., in the range of $[0, 1]$, because of the already distinct individual detail in the mesh.

Third, it proved to be impractical for the three Beziér curves Antihelix, Crus inferius anthelicis, and Crus superius anthelicis, to share a control point. Because the movement

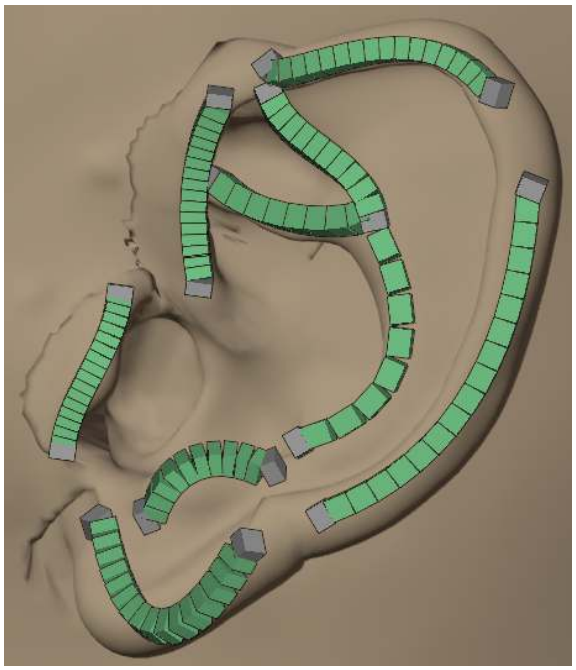


Figure 4.13: One limitation of the initial BezierPPM: discontinuity on the boundary to the head.

of the control point influenced all three bones which most of the time when registering was not the intention and interfered with the registration. This was solved by introducing three control points instead of one, one for each of the mentioned Beziér curves. The three control points were initialised at the same position but nevertheless assigned to each of the three Beziér curves, respectively.

Fourth, a majority of the shape keys overlapped with the output of the Beziér curves, so I decided to reduce the number of shape keys to a minimum. Specifically, shape keys 20 to 22 were replaced by a control bone that is able to globally transform the pinna.

In summary, Tab. A.3 and Tab. A.4 list the updated Beziér curves and shape keys of the updated BezierPPM; and both tables can be found in Appendix A. Figure 4.14 shows the updated version of the parametric pinna model based on Beziér curves's armature, including the newly introduced "Parent" bone highlighted in blue.

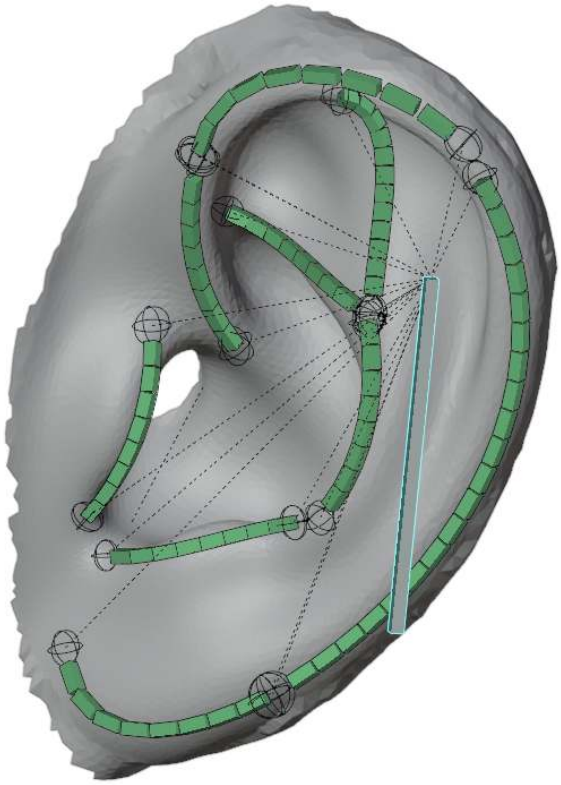


Figure 4.14: Updated armature for the BezierPPM. The grey cuboid with a blue outline represents the newly introduced parent bone.

The updated model was evaluated in the geometric and psychoacoustic domain. The results in the geometric domain showed a median ℓ of under the threshold of 1 mm, i.e., 0.81 mm for NH5, 0.72 mm for NH130, and 0.83 mm for NH131. Table 4.3 shows the results in the psychoacoustic domain by means of quadrant error rate and polar error.

Table 4.3: Sound-localisation errors calculated from the simulated sound-localisation experiment using an auditory model for sagittal-plane sound localisation [67]. Internal reference for "acoustic" was the acoustically measured HRTF, for "calculated" and "PPM", the internal reference was the numerically calculated HRTF from the reference mesh.

	quadrant error rate [%]	polar error [°]
NH5 (acoustic)	2.70	25.60
NH5 (calculated)	7.42	30.54
NH5 (PPM)	4.89	26.30
NH130 (acoustic)	2.50	23.27
NH130 (calculated)	4.75	28.15
NH130 (PPM)	5.29	27.16
NH131 (acoustic)	4.77	27.87
NH131 (calculated)	12.14	35.51
NH131 (PPM)	11.48	34.04

When evaluating the updated model, another four issues were encountered: First, the border of the cutout pinna was still too wide allowing for "folding" of the border when deforming the armature. This was solved by cutting out the pinna even further. Second, the global parameters of the template, i.e., position, rotation, and scaling, were not representing average values as its primary use was the calculation of pinna-related transfer functions (PRTFs). However, the BezierPPM should represent an average pinna as close as possible, which led to globally transforming the pinna in Blender according to the average measures d_5 (pinna height), θ_1 (pinna rotation angle), and θ_2 (pinna flare angle) from the CIPIC database. Third, because the edge of the entrance of the ear canal led to unwanted distortions, it was straightened using a Boolean filter in Blender with a cylinder whose centre coincided with the origin of the coordinate system. Fourth, the sampling of the WiDESPREaD pinna corresponded to the points of the PCA which was not uniformly distributed across the shape. Not only for the application of the numerical calculation of HRTFs it is favorable to have a quasi-uniformly sampled mesh, but also when training machine learning algorithms on using the parametric pinna model based on Beziér curves, a quasi-uniformly sampled mesh is preferrable. The resampling was done using the mesh grading tool from the Mesh2HRTF pipeline [6] [101].

4.3.5 Limits of the parameter space

Some ears that can be modelled may not exist in reality, and the model may be not able to model some ears that exist in reality. A close and reasonable assumption would be to divide the parameter space of plausible ear shapes into pathological and nonpathological ears. However, this assumption is disputable, because not all pathological ears necessarily show strong shape deformities. A few observations I have made shall be mentioned here and are visualised in Fig. 4.15. The Cauliflower ear, or "wrestler's ear", is a deformation of a pinna caused by fluid accumulation in the cartilage after a blunt trauma [223] not covered by the PPM, but since all of the regions important for localisation are nonexistent, it is questionable whether the PPM should be able to model this pathological diagnosis at all. Similarly, the preauricular sinus, a small dell at the start of the Helix tube above the Tragus [224] is not covered by the PPM, but also likely does not influence localisation performance at all. Tuberculum auriculare, also known as Darwin's tubercle is a pathological deformation of the ear which affects between 10 and 58% of the population depending on ethnicity [225, 226]. The PPM covers this deformation partially, but may need an extension in the weight painting on the transition between Beziér curves on the helix. A deformation similar to the Darwin's tubercle is the Stahl ear (in pop culture typically referred to as "Vulcan ear", "elf ear", or "Spock ear"), which the PPM covers as well only partially and may be covered with the same changes to the PPM [227, 228, 229]. On one listener, I observed two protrustions on the Crus helicis, which the first iteration of the PPM covered. However, as I observed a lot more pinnae more closely, these protrusions were not significantly prominent in other individuals, and I did not find any literature on pathological defitinitions of these deformations.

[6] `https://sourceforge.net/p/mesh2hrtf-tools/`

(a) (b) (c) (d) (e)

Figure 4.15: Deformities of the pinna. **(a)** Cauliflower ear, taken from Fig. 1 in [223], **(b)** Preauricular sine, taken from Fig. 1 in [224], **(c)** Darwin's tubercle, taken from Fig. 2b in [225], **(d)** Stahl's ear, taken from Fig. 1 in [229], and **(e)** two protrusions along the Crus helicis.

4.4 Conclusions

In this chapter, the development of a parametric pinna model (PPM) was described and the model was evaluated geometrically and in a simulated sound-localisation experiment. In the model, the pinna surface was divided into parts modeled by shape keys and parts modelled using Beziér curves, the latter coining the name of the model, *BezierPPM*.

The parametric pinna model based on Beziér curves was able to describe a variety pinna geometries using Beziér curves and local deformation (shape keys). The parametric pinna model based on Beziér curves consists of an armature of Beziér curves, and a predefined set of shape keys whose values can be set between -1 and 1. With these, in total 143, parameters, the parametric pinna model based on Beziér curves was able to accurately register ground-truth meshes from six listeners acquired with a structured-light scanner or computer tomography. The median smallest pointwise difference for all six registrations was roughly 1 mm. Sound-localisation errors calculated from a simulated sound-localisation experiment showed similar performance as with the reference pinnae.

Because the parameter space of human pinnae is unknown, it was concluded that the best template pinna for the parameterisation was an average ear taken from 119 structured-light scans. This selection is in contrast to the template used in the NRR proposal, which showed the best results when a non-individual but existing pinna was used as template.

As an additional aspect, common deformities of pinnae were discussed that would be considered pathological from a medical point of view. However, some pathological deformities are not relevant to sound-localisation, e.g., the preauricular sine. Others may be relevant to sound-localisation, e.g., Stahl's ear, and could be covered by extending the parameter space of the parametric pinna model based on Beziér curves. It could be shown that the weight painting of shape keys – which is defined manually – is the main contributor to the accuracy of concave structures. For example, the registrations to the three meshes NH5, NH130, and NH131 that were used already during the development of the parametric pinna model based on Beziér curves showed a smaller median smallest pointwise distance than the three meshes from another lab. The limits of the proposed model

may be unclear but the evaluation on ground-truth meshes suggested that the model is able to result in plausible HRTFs for six listeners with distinctive target pinnae.

Chapter 5

Behavioural evaluation of the proposals

In the psychoacoustic evaluation of investigations so far in this thesis, an auditory model for sagittal-plane sound localisation was used from various versions of the auditory modeling toolbox [67, 10]. Previously, I was able to show that the two proposals were able to register human pinnae with small geometric error and negligible differences in simulated sound-localisation performance. The environment of the simulation was highly controlleable and allowed close examination of systematic changes, which resulted in a proof of concept for the two proposals. However, simulation is one thing, a behavioural experiment with listeners is an entirely different story. In this chapter, the two registration proposals were applied to high-speed PRs and evaluated in a behavioural experiment. Several investigations were conducted that will cumulatively strengthen the arguments towards answering the research question of whether the two proposals can help in HRTF personalisation. In a first step, it was examined whether templates that show a higher similarity to the target have a higher chance to be registered closer to the target, because the two proposals use different templates (NRR uses NH5 and PPM uses AVG). Then, it was examined whether the PR scans show a higher or a lower similarity to the reference meshes than the templates. Next, it was essential to investigate whether the proposals improved the PR meshes in terms of the similarity to the reference. Finally, it was investigated whether the output of the two proposals yielded a higher similarity to the PR, i.e., the low-quality targets the respective templates were actually registered to, or to the references, i.e., the "hidden" and high-quality targets. In the acoustical domain, a crucial feature was the modification of spectral cues in two scenarios. First, the influence that the PR has on an HRTF compared to the respective reference. And, second, the influence that the two proposals have on the HRTF compared to the PR and the reference. In the psychoacoustic domain, the following four conditions were evaluated in a behavioural sound-localisation experiment: HRTFs numerically calculated from the reference, the PR, and the two proposals (NRR and PPM) being registered to the PR. Various errors between target and response angles were calculated.

5.1 Methods

In this chapter, the relationship between two conditions was classified in the geometrical and acoustical domain. In the geometric domain, the Jaccard index J (Eq. 3.3) was calculated in order to quantify the similarity of two point clouds. In the acoustical domain, the Spatial Audio Metrics (SAM) Toolbox [230][1] was used to calculate the log-spectral difference (Eq. 2.2) was calculated and averaged over frequency and position.

Results in this chapter abide by the following notation: One of the proposals applied to a target was noted as *proposal(target)*, e.g., NRR applied to PR was noted as *NRR(PR)*. If the error measure stated a comparison, e.g., between reference and the NRR algorithm applied to the PR, the condition was noted as *REF, NRR(PR)*.

5.1.1 Mesh acquisition

Four different meshes were obtained for each listener: A structured-light scan, a photogrammetric reconstruction, and the two proposals applied to the PR.

The *structured-light scan* mesh was taken as a reference mesh from the SONICOM database [116] and postprocessed in the following way: The watertight version of the mesh, saved in the stereolithography (.STL) file format, was imported into the Blender workspace (v3.6, Blender Foundation) [148]. Vertices that were not linked to the geometry were identified and deleted. The mesh was aligned to the coordinate system such that the center of the head aligned with the origin, the centers of the ear canals were aligned with the y-axis, and the tip of the nose was aligned with the x-axis. After the alignment, the mesh was inspected for artifacts introduced by the scanning process, which were then removed using the "simplify" brush. Next, the mesh surface was smoothed (using the "smooth" brush) such that acute angles were flattened. In order to remove the wig cap and compressed hair from the mesh, a Bézier curve was added to follow the head boundary along the median plane and closed a loop above the head. Then, Blender's knife project tool was applied using the "cut through" option – which corresponds to applying a Boolean modifier – and the introduced gap in the head was filled with faces and the new surface was smoothed. Other sculpting brushes, e.g., "flatten" and "inflate", were used in order to make the mesh surface as smooth as possible without losing important detail in the pinna region. Finally, a Boolean modifier was applied to the mesh with a plane such that all parts below the neck were discarded. Due to the submillimeter resolution of the scanner, the cleaned mesh was considered adequate to be used as reference, further denoted as *REF*.

Photogrammetry can also be executed without the use of special equipment [128]. Albeit previous studies [114, 231] concluded that taking a video results in worse scans than using photos as input because of blurring between frames, I chose a video as source in order to keep the process as easy and fast to execute as possible [128]. The *photogrammetric reconstruction* (PR) was performed using the high-speed method described as follows:

[1] `https://github.com/Katarina-Poole/Spatial-Audio-Metrics`

The (seated) listener was filmed in the following trajectory: Start in front of the listener's face, move the camera horizontally towards their left ear, follow a vertical circle around the right ear, continue the horizontal trajectory towards the back of their head and their right ear, repeat the vertical circle around the right ear, and finish at the starting point in front of the listener's face. The emphasis on the pinnae was done in order to have more data on the pinna such that the PR would yield a higher resolution in these areas [232]. The video was recorded in a lossy format (.MP4) using the built-in (20 Megapixels) camera application of a smartphone (5T, OnePlus), and lasted about 30-40 seconds. The video was exported as 100 frames in a lossy format (.JPG). To reconstruct a 3D mesh from the photos, here the Educational version of Recap Photo was used (v24.0.0.231, Autodesk). Figure 5.1 shows the source material and the result of the PR compared to the reference mesh.

(a) (b) (c)

Figure 5.1: **(a)** Frame of the source video for the PR, **(b)** PR mesh, **(c)** reference mesh.

After the reconstruction was finished, a watertight mesh was exported as .STL, globally (position, orientation and size) aligned to the reference, and rudimentary cleaned up in Blender using the software's built-in add-on "3D-Print Toolbox" and the "smooth" brush used for the postprocessing of the reference scans. Note that the postprocessing here was the bare minimum in order to prepare the mesh such that HRTFs can be numerically calculated.

The first registration proposal, *non-rigid registration*, utilized the Bayesian formulation of coherent point drift (BCPD) to register a template pinna to the target pinna such that the distance between the two point clouds was minimized [171]. The left and right ears of NH5 from the ARI database were used as template pinnae, and the listeners' left and right ears cut out from the reference as well as the PR were used as target pinnae, respectively. Note that this algorithm is applied to point clouds and not meshes, and the output is a point cloud as well. In order to being able to calculate HRTFs, a mesh had to be calculated from the point cloud. Thus, after the registration, the registered point cloud was imported to Meshlab (v2023.12, Consiglio Nationale delle Ricerche) [172]. First, point normals were estimated using Meshlab's internal tool with the default parameters, taking into account

the 10 nearest neighbouring points. Then, a mesh surface was reconstructed using the Screened Poisson surface reconstruction with the default parameters [233]. If the element normals pointed towards the inside of the geometry, their orientation was inverted such that normals point outwards. Finally, the resulting mesh was exported as .PLY file.

The second registration proposal, the *BezierPPM*, was used to deform a template pinna such that the distance between the surface of the template and the target was minimized (v3.0 from [204]). In this method, the average ear mesh taken from the WiDESPREaD database [40] was used as template pinna (further denoted AVG). As the BezierPPM is available as a left ear only, the left ears from the target meshes could be registered as they were. The right ears of the target meshes were mirrored around the median plane before registration, and the registered BezierPPM mesh was mirrored back to represent a registered right ear. The BezierPPM was aligned manually such that the Blender plugin PointCloudCompare showed a median pointwise Euclidean distance of less than 1 mm. After the registration, the deformed mesh was exported as .PLY file.

The NRR algorithm and the BezierPPM were applied to the PR ears, and the registered ear meshes were stitched to the same head as the target, because this investigation focused on the pinnae. For both registration methods, Blender was used to stitch the registered ear meshes to the head of the PR mesh. All three meshes were imported into a clean Blender workspace. The ears from the target head were removed, and one registered ear mesh was joined with the PR head mesh. Then, the boundary of the joined ear and the boundary of the respective hole in the target head were selected and connected. The process was repeated for the other ear as well. Finally, the transition between ears and head was smoothed again in order to ensure continuity between elements – again, with minimum postprocessing effort. Figure 5.2 shows the four different meshes for one listener. Note that, in order to compare the PR and its improvements to the reference, only the pinnae were swapped – the head for these three conditions was the same, i.e., the PR head.

All head meshes were graded using curvature-adaptive mesh grading in order to speed up the numerical HRTF calculation [101]. All HRTFs were numerically calculated using Mesh2HRTF v1.1.1 [113], using the SONICOM grid and a frequency resolution of 129 bins equidistantly distributed between 0 and 24 kHz, such that the corresponding impulse responses had a sampling frequency of 48 kHz.

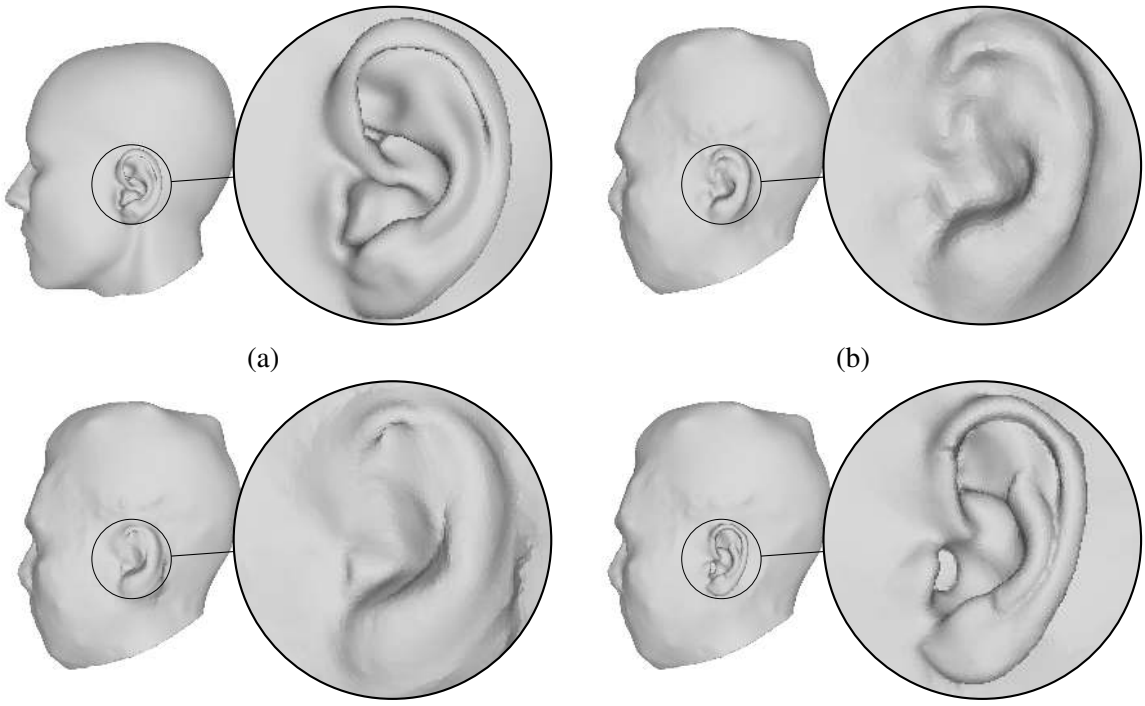
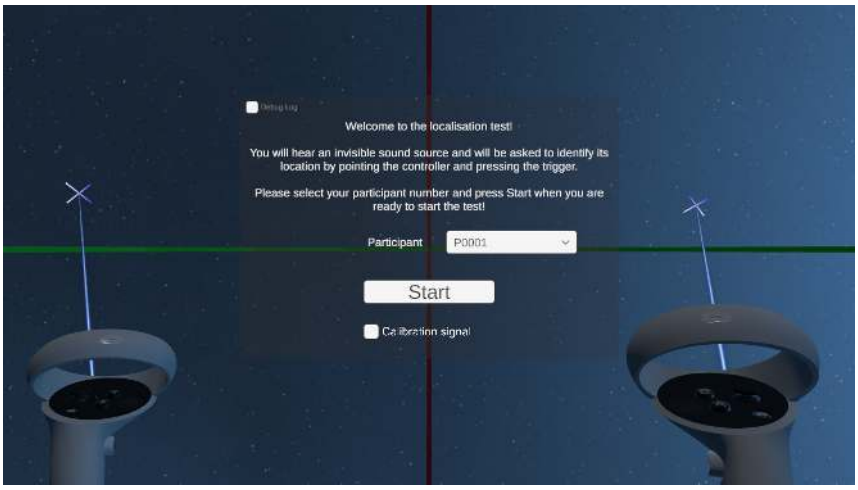


Figure 5.2: Obtained meshes shown for one listener. **(a)** Structured-light scan from SONICOM database (REF), **(b)** photogrammetric reconstruction (PR), **(c)** registration result using the NRR approach, **(d)** registration result using the PPM approach.

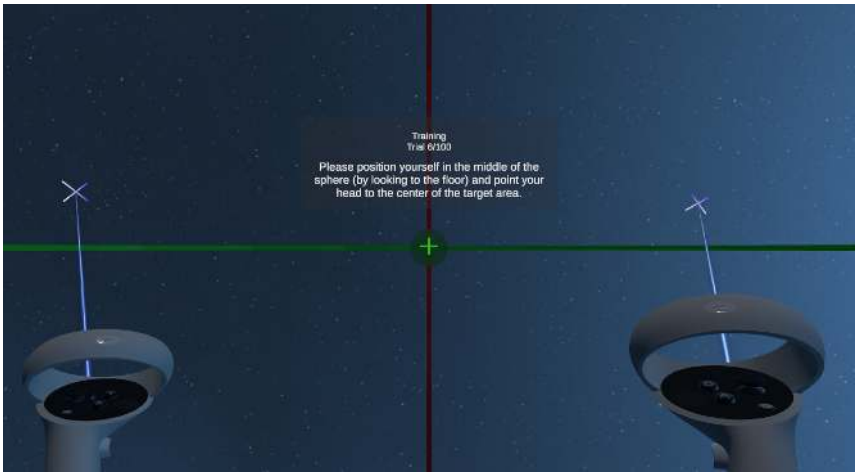
5.1.2 Behavioural experiment

Signals

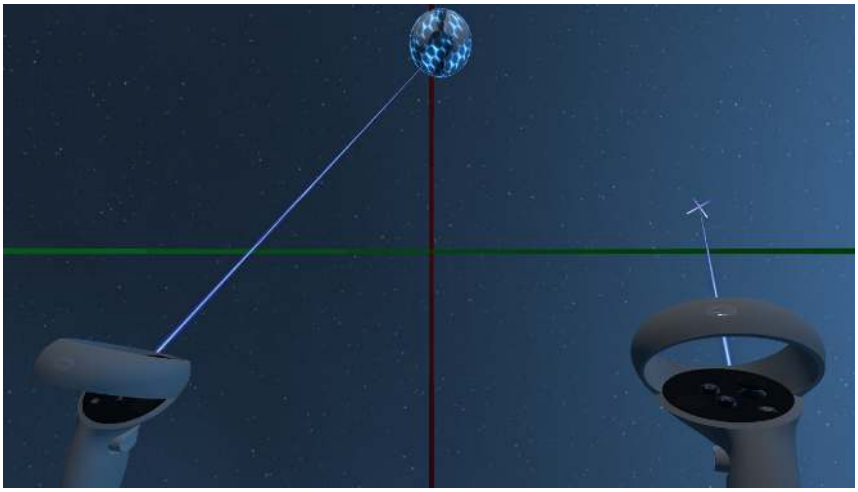
The stimulus presented in all conditions consisted of three consecutive noise bursts, created as follows: a 300-ms white noise signal was windowed with three consecutive Hann windows, each with a length of 100 ms. The experiment code was taken from [234], and HRTFs were rendered using the 3D Tune-In Toolkit in Unity (v2020.3.42f1, Unity Technologies) [7]. Figure 5.3 shows three screenhots of the user interface as displayed in the VR goggles.



(a) Startup canvas. Listeners had to select their participant ID using the controllers.



(b) Listeners had to align the green crosshair with the green circle in the front so that the stimulus would play.



(c) In the familiarisation phase, the real position of the sound source was revealed and visualised as blue sphere.

Figure 5.3: Screenshots of the sound-localisation experiment in VR.

Participants and apparatus

Nine listeners (three female) without any self-reported hearing loss from the SONICOM database participated. The room was the same in which the SONICOM dataset was acquired, and the whole experiment was deployed on a Meta Quest 2 over direct headphone playback with a pair of Sennheiser HD599. The volume of the Quest 2 was calibrated using an EARS (MiniDSP, Hong Kong) headphone calibrator with attached pinnae, such that a sine played at 1 kHz filtered with the HRIR from the front ($[azi, ele] = [0, 0]°$) would yield 65 dB(A) in the measurement programme "Room EQ Wizard" (v5.20.13, John Mulcahy). Figure 5.4 shows a listener participating in the experiment and the room in part.

Figure 5.4: Listener participating in the experiment.

The sound-source grid was constructed at 42 positions distributed in four rings around a sphere with a radius of 1.2 m in order to match the distance of the SONICOM grid. The four horizontal rings were at elevations of -30, 0, 30, and 60°, consisting of 12, 16, 12, and 2, respectively, equi-distantly spaced directions starting at an azimuth of 0°.

Procedure

Prior to the experiment, a pre-test was conducted in which four listeners participated in. This pre-test made sure that the setup was working properly and that the test was sensitive to differences between individual and non-individual HRTFs. In order to familiarise the listener with the virtual user interface and controls, the pre-test started with 100 trials using KEMAR HRTFs acoustically measured with small ears. For the first 30 trials, the sound source was visualised only, in order to familiarise the participant with the VR equipment. A trial ended when the participant indicated a response by pressing a button on

the controller and realigned their head to the centre in the front (as illustrated in Fig. 5.3b). For the subsequent 70 trials, a static sound was played, during which the listener was not allowed to move their head. After the response of the perceived sound-source location was given by the listener, visual feedback was given on the actual position of the sound source while the sound source emitted continuous Gaussian noise. During the continuous noise playback, dynamic head movements were allowed until the listener pressed the button again. After the listener realigned their head to the centre of the coordinate system, the next trial started. The rest of the pre-test was divided into blocks of 100 trials, after which participants were given the opportunity to take a short break. In the remaining part of the pre-test, no feedback was given on the true sound-source position, and for each trial the stimulus was played once, during which the listener was not allowed to move their head. There were two conditions in the pre-test: individual (acoustically measured) HRTF, and the (acoustically measured) HRTF of a KEMAR dummy head with large ears. The order of the blocks was randomised. There were 300 trials per condition, yielding 600 trials in total. The total test duration was approximately 30 minutes.

In the main experiment, there were four conditions: PR, NRR registered to the PR, and the PPM registered to the PR, and the REF. The experiment design was similar to the pre-test: The experiment started with a familiarisation phase using the HRTF of a KEMAR head with small ears. The first 10 trials were silent using only visual feedback of the sound-source position. For the remaining 90 of in total 100 trials, the stimulus was played, the position of the sound source was revealed after the listener submitted their response, and the sound source emitted continuous Gaussian noise until the next response submission. After the listener realigned their head to the centre of the coordinate system, the next trial started. The rest of the experiment was divided into blocks of 100 trials, after which participants were given the opportunity to take a short break. After the familiarisation phase, no feedback was given on the true sound-source position, and for each trial the stimulus was played once, during which the listener was not allowed to move their head. The order of blocks was randomised. There were 300 trials per condition, yielding 1,200 trials in total. The total experiment duration was approximately 1.5 to 2 hours.

From the mismatch between target and response angles, the following sound-localisation errors were calculated: Lateral bias and error, quadrant error rate, and polar bias and error (excluding the responses classified as quadrant errors). These errors were calculated using the AMT v1.5 [10]. Contrary to previous investigations, lateral bias and error were included in the evaluation of the behavioural experiment, because it was evaluated whether using the PR head would influence binaural cues.

5.2 Results

5.2.1 Geometrical domain

Similar to the CT scan reference case, the point cloud sizes were distinguishably different between the reference and the PR. Table 5.1 shows the difference in point cloud size between the reference and the PR for each listener.

Table 5.1: Differences in point cloud size for each listener's between the reference scan and PR.

pID	reference	PR
P0001	9,251	3,690
P0006	8,404	1,840
P0031	8,871	3,728
P0107	8,073	3,803
P0121	8,518	4,300
P0172	7,413	2,440
P0181	8,136	3,455
P0204	8,312	2,341
P0205	9,474	3,513

In this approach, the whole head mesh was reconstructed, which allowed for a comparison of head measurements to the reference. These measurements were performed within the Blender workspace. A Boolean filter was used between the head mesh and a planar mesh in the horizontal plane, yielding the cross section of the head. In order to calculate the circumference of the head, the edge lengths of the cross section were summed up. Blender's "measure" tool was used to calculate the furthest distance between left and right (head width), and front and back (head depth). Because hair easily gets captured as part of the shape by the PR, it was likely that the head measures were larger in the PR meshes than in the reference despite the ear contours being aligned. Figure 5.5 shows the differences in head measures in cm between the reference and the PR for all nine listeners. Positive values indicate that the reference was higher, negative values indicate that the PR was higher. One reason P0031 shows such a large difference in the PR may be that the listener had a prominent haircut, which was not altered by wearing a headband.

Table 5.2: Differences of head measurements between the reference and the PR by means of average μ and standard deviation $\pm\sigma$.

	μ (cm)	$\pm\sigma$ (cm)
circumference	-3.26	3.29
depth	-0.67	1.07
interaural distance	-1.27	0.51
width	-2.03	0.9

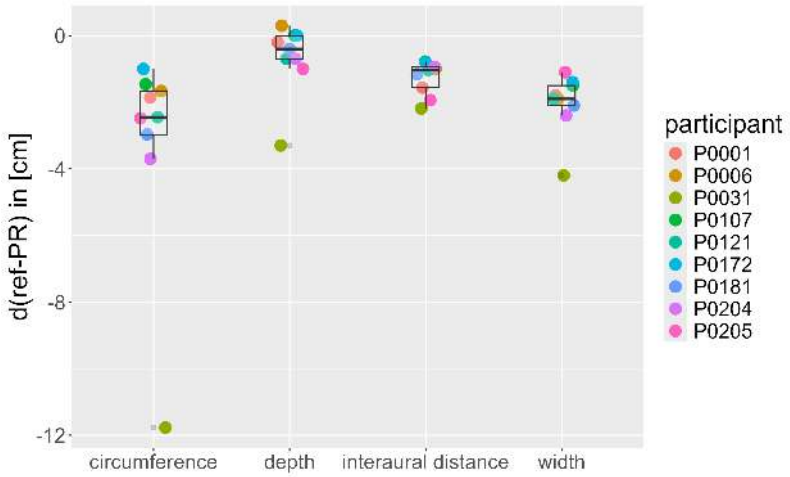


Figure 5.5: General impact of the PR on head measures (circumference, depth, interaural distance, and width).

Figure 5.6 shows Js between the reference meshes and the respective templates for NRR (NH5 template) and PPM (AVG template), between the reference and the PR, between the PR and the proposals applied to the PR, and between the reference and the proposals applied to the PR. The median values of J between the reference meshes and the templates for NRR were 0.26 and for PPM 0.29, between the reference and the PR 0.18, between the PR and the proposals applied to the PR 0.65 for NRR and 0.30 for PPM, and between the reference and the proposals applied to the PR 0.22 for NRR and 0.24 for PPM. A Shapiro test rejected the Null hypothesis and showed that the data was not normally distributed ($p < 0.05$) and a Kruskal-Wallis rank sum test rejected the Null hypothesis and showed that the data was likely not drawn from the same distribution ($p < 0.05$). Thus, pairwise comparisons were conducted using a Wilcoxon rank sum exact test. These comparisons showed significant differences between *PR, NRR(PR)* to all other conditions ($p < 0.05$), significant differences between *PR, PPM(PR)* and *REF, NRR(PR)*, *REF, PPM(PR)*, and *REF, PR* ($p < 0.05$). Table 5.3 shows the p values for the pairwise comparisons, significances are highlighted in bold.

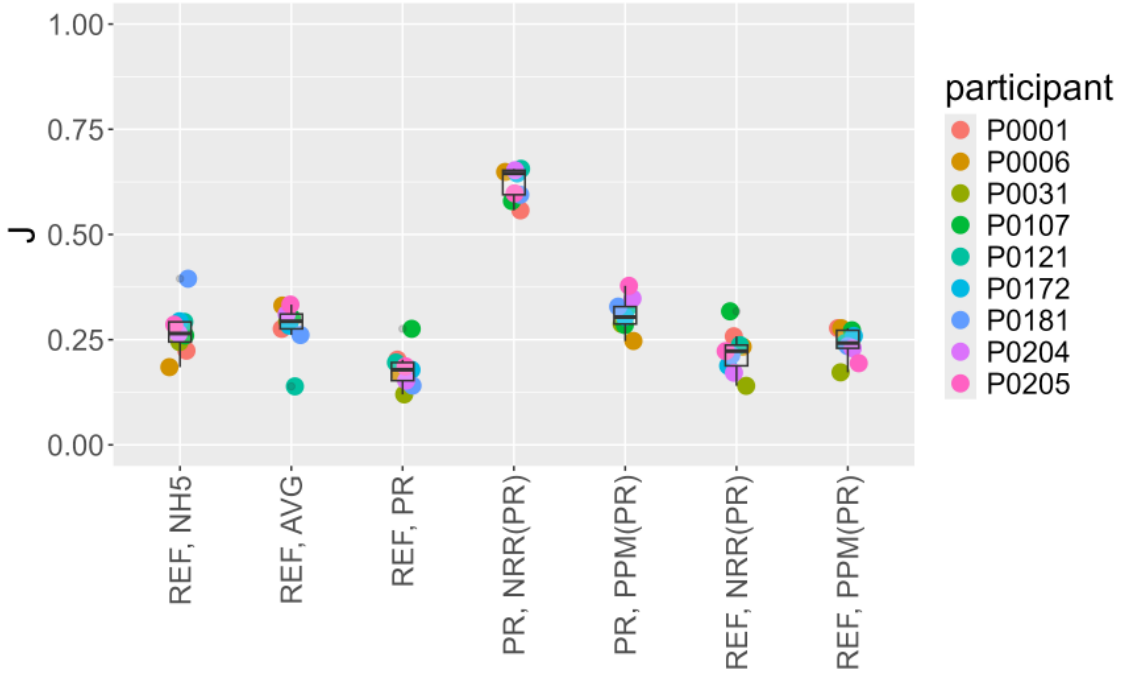


Figure 5.6: Jaccard index J between the reference meshes and the respective templates for NRR (NH5) and PPM (AVG), between the reference and the PR, between the PR and the proposals applied to the PR, and between the reference and the proposals applied to the PR.

Table 5.3: Pairwise comparisons between Js using Wilcoxon rank sum exact test.

	PR, NRR(PR)	*PR, PPM(PR)*	*REF, NH5*	*REF, NRR(PR)*	*REF, PPM(PR)*	*REF, PR*
PR, PPM(PR)	**0.00086**	-	-	-	-	-
REF, NH5	**0.00086**	0.37515	-	-	-	-
REF, NRR(PR)	**0.00086**	**0.01604**	0.35216	-	-	-
REF, PPM(PR)	**0.00086**	**0.00691**	0.76100	0.76100	-	-
REF, PR	**0.00086**	**0.00123**	**0.03307**	0.46956	0.14192	-
REF, AVG	**0.00086**	0.76100	0.76100	0.19548	0.16882	0.06199

5.2.2 Acoustical domain

Figure 5.7 shows the HRTFs of one listener in the median plane for the reference, the PR, and the two proposals being registered to the PR. Note the compressed dynamic range of the HRTF in the PR and NRR(PR) conditions.

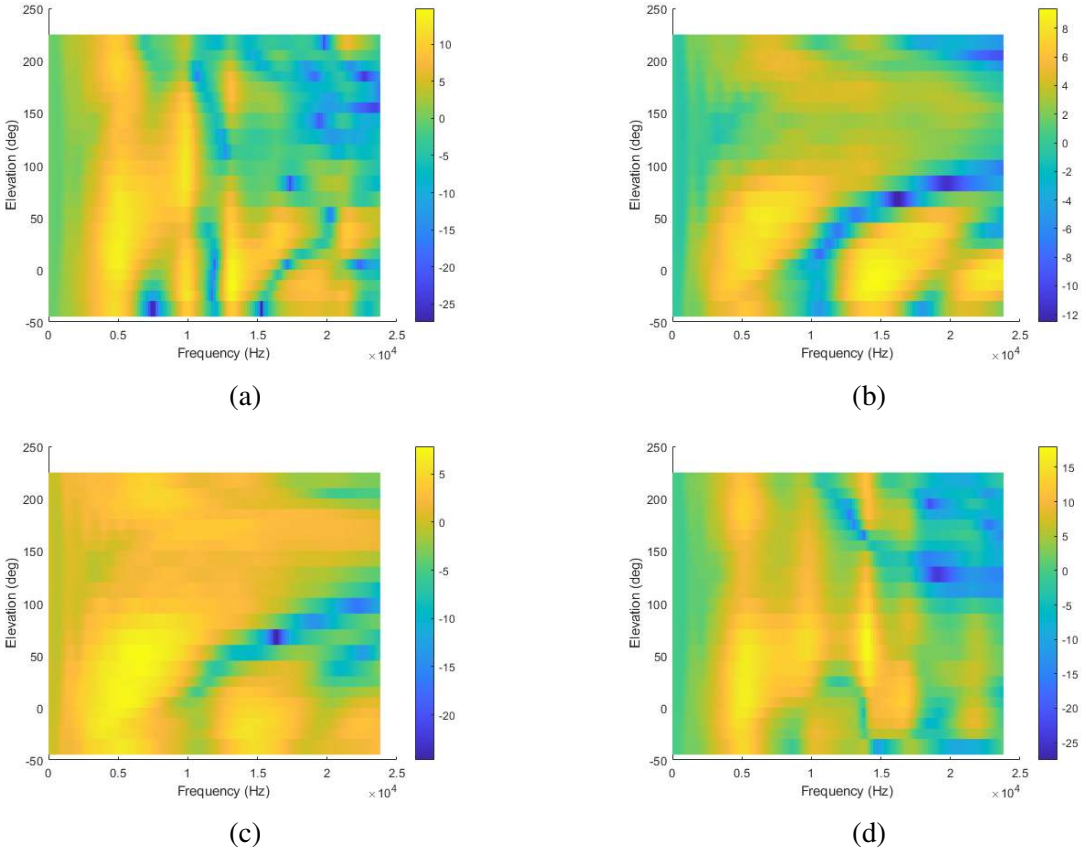


Figure 5.7: P0172's HRTFs in the median plane, calculated from (a) REF, (b) PR, (c) NRR registered to the PR, (d) PPM registered to the PR.

Figure 5.8 shows the log-spectral difference between the PR and the proposals applied to the PR, between the reference and the proposals applied to the PR, and between the reference and the PR. The log-spectral difference between the reference and the used templates (NH5, AVG) could not be calculated, because AVG was only the pinna without a head attached to it. This rendered a numerical calculation of an HRTF impossible. The median log-spectral difference between the PR and the proposals applied to the PR was 2.0 dB for NRR and 3.8 dB for PPM, and between the reference and the proposals applied to the PR 5.7 dB for NRR and 5.4 dB for PPM, and between the reference and the PR 5.9 dB. A Shapiro test rejected the Null hypothesis and showed that the data was not normally distributed ($p < 0.05$), and a Kruskal-Wallis test rejected the Null hypothesis and showed that the data was likely not drawn from the same distribution ($p < 0.05$). Thus, pairwise comparisons were performed using the Wilcoxon rank sum exact test, yielding significant differences between *REF, NRR(PR)* and *PR,NRR(PR)* and *PR, PPM(PR)*, between *REF, PPM(PR)* and *PR,NRR(PR)* and *PR, PPM(PR)*, and between *REF, PR* and *PR,NRR(PR)*, *PR, PPM(PR)*, and between *PR, NRR(PR)* and *PR, PPM(PR)*. Significances are not shown in the figure due to readability. Table 5.4 shows the p values for the pairwise comparisons, significances are highlighted in bold.

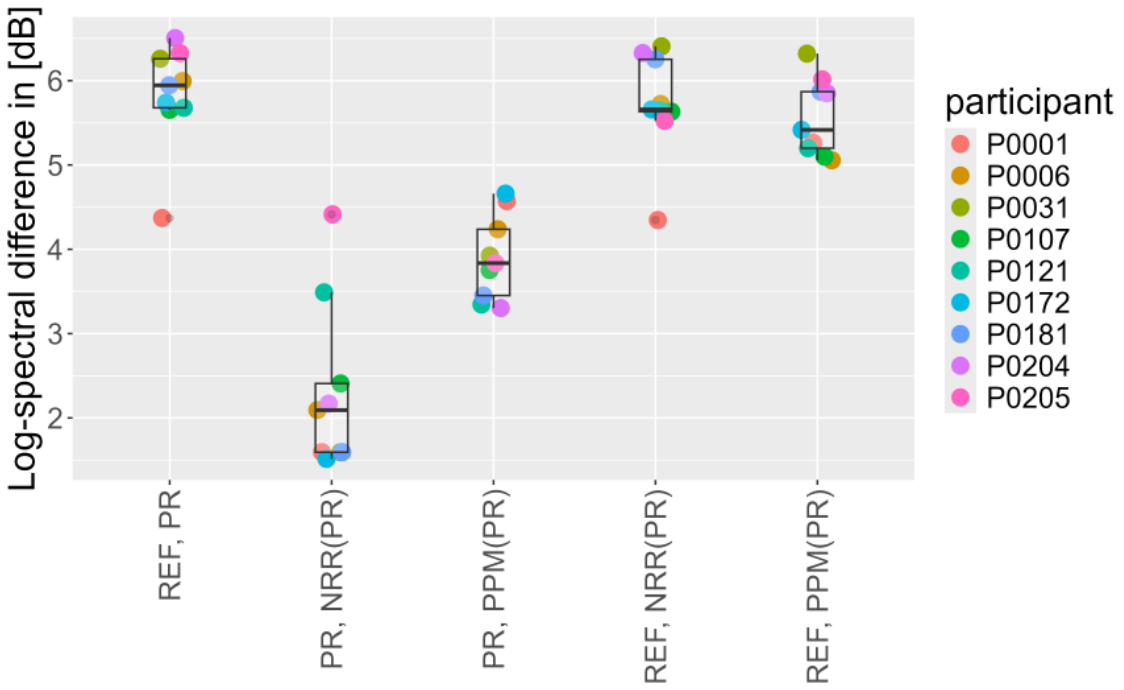


Figure 5.8: Log-spectral difference (averaged over frequency and position) between registrations to the PR and the reference, registrations to the PR and the PR, and between reference and PR.

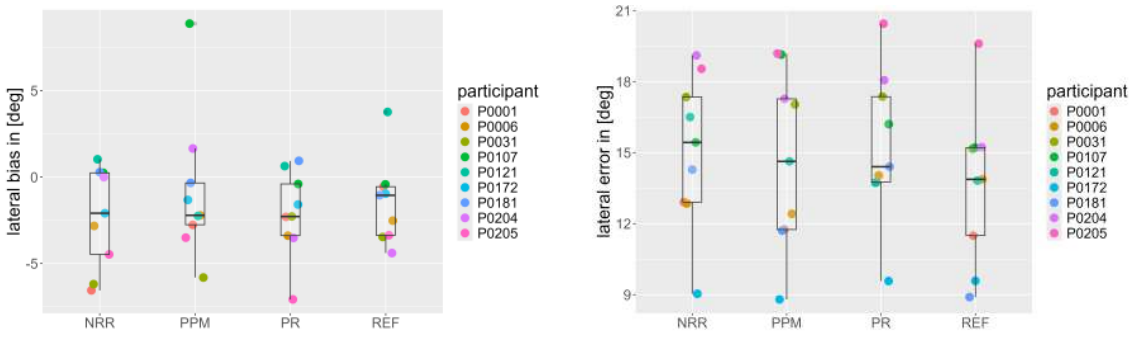
Table 5.4: Pairwise comparisons between log-spectral differences using Wilcoxon rank sum exact test.

	PR, NRR(PR)	*PR, PPM(PR)*	*REF, NRR(PR)*	*REF, PPM(PR)*
PR, PPM(PR)	**0.02254**	-	-	-
REF, NRR(PR)	**0.00066**	**0.00099**	-	-
REF, PPM(PR)	**0.00041**	**0.00041**	0.77302	-
REF, PR	**0.00066**	**0.00099**	0.77302	0.66726

5.2.3 Behavioural domain

Figure 5.9 shows the lateral accuracy and precision error for the sound-localization experiment in the median plane. For the lateral accuracy error, a Shapiro test rejected the Null hypothesis and showed that the data was not normally distributed ($p < 0.05$). A Kruskal-Wallis test accepted the Null hypothesis that the data was drawn from the same distribution ($p > 0.05$), thus, no significant differences between groups were found. For

the lateral precision error, a Shapiro test accepted the Null hypothesis ($p > 0.05$) followed by a Bartlett test that accepted the Null hypothesis as well ($p > 0.05$). Thus, an analysis of variance (ANOVA) was performed, showing no significant differences between groups ($p > 0.05$).



(a) (b)

Figure 5.9: **(a)** Lateral accuracy and **(b)** precision error of the sound-localization experiment.

Figure 5.10 shows the quadrant error rate from a simulated sound-localization experiment in the median plane. A Shapiro test accepted the Null hypothesis ($p > 0.05$), followed by a Bartlett test that also accepted the Null hypothesis ($p > 0.05$). Thus, an ANOVA was performed, and yielded significant differences between groups ($p < 0.05$). A Tukey multiple comparisons of means revealed significant differences between the reference and NRR, and between the reference and PR. Even when the outlier in the PPM condition was discarded, no significant differences between the reference and the PPM condition could be shown.

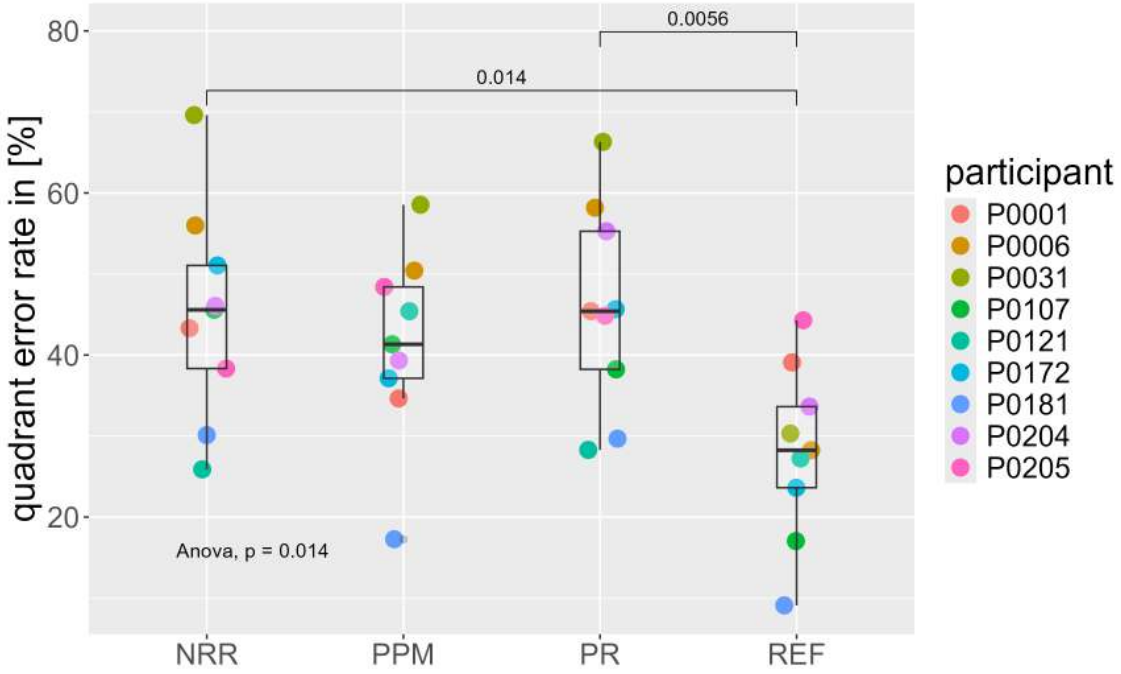
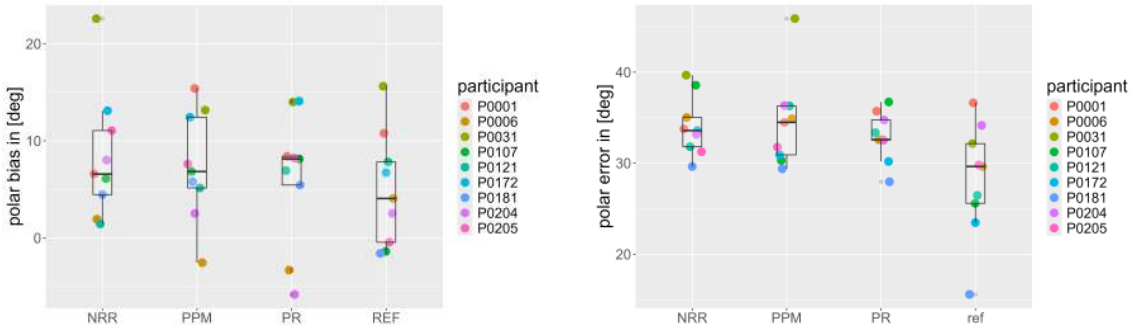


Figure 5.10: Quadrant error rate in % for NRR- and BezierPPM-registrations to the PR, and the reference.

Figure 5.11 shows the polar bias and error from the sound-localisation experiment in the median plane. For the polar accuracy, a Shapiro test accepted the Null hypothesis and showed that the data was normally distributed ($p > 0.05$), followed by a Bartlett test that also accepted the Null hypothesis and showed that equal variances ($p > 0.05$). An ANOVA yielded no significant differences between groups ($p > 0.05$). For the polar precision error, a Shapiro test accepted the Nully hypothesis ($p > 0.05$), as well as the Bartlett test ($p > 0.05$). A subsequent ANOVA revealed significant differences between groups ($p < 0.05$), followed by a Tukey multiple comparison of means that showed significant differences between the reference and NRR and between the reference and PPM.



(a) (b)

Figure 5.11: Polar **(a)** bias and **(b)** error for NRR- and BezierPPM-registrations to the PR, the PR, and the reference.

5.3 Discussion

In the geometrical domain, J was highest for *PR, NRR(PR)*, i.e., between the NRR registration to the PR and the PR. This indicates that the NRR approach was able to register a target closer, which may in general be desirable for registrations to high-quality targets, but may not necessarily be desirable for the high-speed PR scans. Resulting in a mesh with a high J to the PR suggests that the acoustical results, and likely the behavioural results as well, will not be significantly different than those of the PR. Both registration methods - NRR and PPM showed a higher J when registered to the PR than between the reference and the PR. This indicates that both registration approaches improved the similarity to the reference despite only having the sparse shape information of the PR. It is worth noting here that the shape keys in the PPM registrations were set to their low end in order for the concave structures to be as close as possible to the PR, i.e., flatter than to the reference. One thing to consider is whether shape keys would be necessary to change at all - because having a bit of concave detail albeit non-individual may be better than having little to no concave shape at all.

In the acoustical domain, the log-spectral difference was highest for comparisons to the reference, without significant differences between the three conditions, i.e., between *REF, NRR(PR)*, *REF, PPM(PR)*, and *REF, PR*. The log-spectral difference in the conditions *PR, NRR(PR)* and *PR, PPM(PR)* was less than 5 dB, in case of the former resulting in a median log-spectral difference of 2.0 dB. The fact that the log-spectral difference between PR and NRR is less than 3 dB for the majority of listeners and because there were no significant differences found between *REF, NRR(PR)* and *REF, PR* suggests the sound-localisation performance of NRR will be similar to the PR condition. This is in line with the findings in the geometrical domain, in which J was high in the case of *PR, NRR(PR)*, which strengthens the argument of PR and NRR(PR) resulting in similar sound-localisation performance. Unlike the NRR-related log-spectral differences, the PPM-related log-spectral differences do not suggest similar sound-localisation performance to the PR.

In the behavioural domain, the sound-localisation errors observed in general are rather

high, especially in the Sagittal planes, but comparable to naive and untrained listenerst [235, 4]. The response angles were corrected by the polar accuracy, because the PR head alignment may have introduced a polar bias, however, this changed the results marginally.

The lateral accuracy, i.e., the lateral bias, shows a range typical for naive and untrained listeners [4]. This error on average is approximately close to or slightly below $0°$, some participants show a lateral accuracy error of more than $\pm 5°$. This could have multiple reasons, e.g., a mismatch between the reference mesh alignment to the coordinate origin and the head orientation of the participant in real life, or the blocking of the virtual ear canal introducing a mismatch in the interaural time difference (ITD), or the unfamiliarity of participants with sound-localisation experiments in general, or the usage of VR equipment, e.g., pointing by using the VR controllers. Similar aspects can be said for the lateral precision error. Additionally, participants performed far below the chance rate of 54 %. For both lateral errors, no significant differences could be found between conditions. Regarding the lateral precision errors, the results seem quite high compared to other static sound-localisation experiments with (longer) sound-localisation training but comparable to naive and untrained listeners' results [4].

The quadrant error rates are high for a sound-localisation experiment (chance rate at 64%), but comparable to previous experiments lacking a training phase [235, 4]. The familiarisation phase with the dummy head showed a quadrant error rate of $28.6 \pm 7.7\%$. The fact that the familiarisation phase resulted in similar quadrant error rates between the own and a dummy-head HRTF suggests that the familiarisation phase may have been too short with 100 trials, as the quadrant error rate after training should be at around 8 % [31, 4]. The REF condition shows the lowest quadrant error rates and significant differences between the PR and the NRR condition. The quadrant error rate in the REF condition does not show significant differences to the PPM condition, however, neither does the PPM condition to NRR or PR. The fact that the latter difference is not significant prevents me from drawing conclusions on the performance of the PPM proposal. However, if a training phase to the own (calculated) HRTFs would be introduced to the experiment, I would expect the quadrant error rate in the REF condition to decrease drastically, and in the PPM condition to decrease more than in the PR or NRR condition as these two conditions contain far less spectral cues.

The polar accuracy did not show significant differences between conditions. For the majority of participants, the polar accuracy is nonzero in all conditions, and for some even close to the chance rate of $14°$. Because not all listeners show the lowest polar accuracy for the reference condition, the alignment of the PR head mesh is unlikely the reason for the polar bias. One explanation might be that the participants have been familiarised with dummy-head HRTFs instead of their own during the first 100 trials, another explanation could be that the loudspeaker rig they were sitting in was introducing a visual upwards bias albeit the loudspeaker rig was not used in the experiment and not mentioned to the participant. In the polar precision, significant differences were found between the reference and the two proposals. No significant differences between the reference and the PR, but also no significant differences were found between the PR and the two proposals. This suggests that the two proposals contain local spectral information that is different from the

reference, but it is peculiar that the polar precision for the PR condition did not show significant differences to the reference condition. Many of the participants performed close to the chance rate for the polar precision of $36°$. This suggests that there was a ceiling effect and that extensive training on the own (calculated) HRTFs would have helped in magnifying the apparently subtle differences between conditions.

In general, the overall duration of the experiment of up to 2 hours was short for a sound-localisation experiment investigating such a small difference in sound-localisation errors. The experiment duration restriction, i.e., to keep it as short as possible, was the reason that a control condition was missing, as this would have increased the duration by approximately 30 minutes. One additional general aspect may be that the amount of hearing loss was *self-reported*, because conducting an audiogramme prior to the experiment for each listener would be considered a medical procedure, which implied significant effort to officially conduct. A possible workaround would have been to calculate the quadrant error rate after the training – with the individual HRTFs – and if it's above a certain threshold, the participant would been deemed unfit for the sound-localisation task. Then again, the experiment duration is likely to be increased drastically, as such a training would have taken 4 times longer, i.e., 400 trials [4].

I argue that two aspects could have improved the sound-localisation errors overall: First, extensive training of 400 trials on the listener's individual HRTF, because what the familiarisation phase achieved was familiarisation with the VR user interface and controllers, and with the dummy-head HRTF pair that the stimuli were convolved with. Second, because the perception of timbre is completely disregarded in the experiment, DTFs may have helped in distinguishing directions, because they contain only direction-dependent spectral cues. As the former improvement would likely increase the overall duration of the experiment, the experiment could be split in two - each one focusing on a different proposal.

It would have bene interesting to compare the sound-localisation errors to the condition with a non-individual or a dummy-head HRTF, which would serve as a control condition. Such a control condition was not included in the experiment, but the auditory model was utilised here post-hoc, as the investigations in the next section describe.

5.3.1 Behavioural results compared to an auditory model

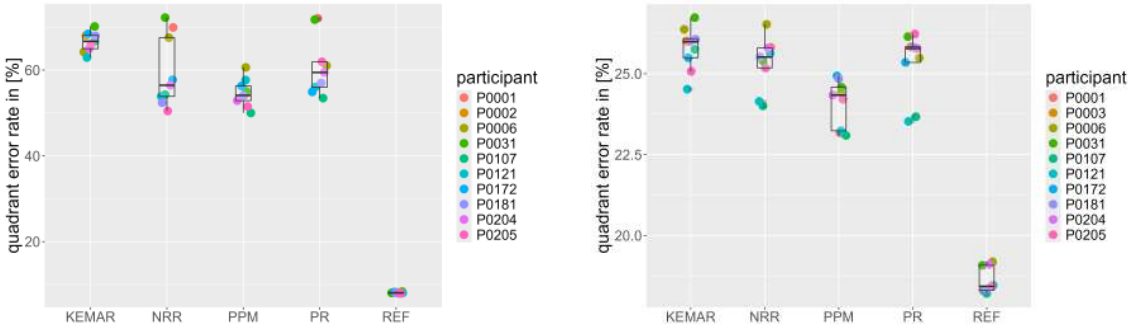
The numerically calculated reference DTF was used as the internal template for the auditory model, and the conditions from the behavioural experiment were simulated: PR, NRR, and PPM. Additionally, a dummy-head DTF was included as control condition. In addtion, the simulation of a scenario was investigated in which the proposals were registered to the references.

Sound-localisation performance for registrations to PR

The auditory model was calibrated to yield a quadrant error rate of 8 % when taking the reference DTF as internal template and target. The only parameter that was varied in

the calibration was the sensitivity of the model, i.e., how "good" of a sound localiser the listener would be. For this investigation, only the quadrant error rate and the polar error were calculated using `localizationerror` from the AMT v1.5 [10].

Figure 5.12 shows the quadrant error rate and the polar error for a simulation of the behavioural experiment using the auditory model including one additional condition: a dummy-head HRTF from a KEMAR dummy head with large ears. Regarding the quadrant error rate, a Shapiro test rejected the Null hypothesis and showed non-normality in the data ($p < 0.05$), and a subsequent Kruskal-Wallis rank sum test rejected the Null hypothesis and showed that the data was likely not drawn from the same distribution ($p < 0.05$). Applying a Wilcoxon multiple comparison of means showed significant differences between the PPM and the KEMAR, and between the reference and all other conditions. For the polar error, a Shapiro and Kruskal-Wallis test rejected the Null hypothesis ($p < 0.05$), and a Wilcoxon multiple comparison of means showed significant differences between the PPM and the KEMAR, and between the reference and all other conditions. Tables 5.5 and 5.6 show the p values for the pairwise comparisons, significances are highlighted in bold.



(a) (b)

Figure 5.12: **(a)** Quadrant error rate and **(b)** polar error for modelled sound-localisation experiment using the auditory model. NRR and PPM were registered to the PR.

Table 5.5: Pairwise comparisons of quadrant error rates using Wilcoxon rank sum exact test, significant differences are highlighted in bold ($p < 0.05$).

	KEMAR	NRR	PPM	PR
NRR	0.34048	-	-	-
PPM	**0.00041**	0.59465	-	-
PR	0.20123	0.59465	0.15734	-
REF	**0.00369**	**0.00369**	**0.00369**	**0.00369**

Table 5.6: Pairwise comparisons of polar precision errors using Wilcoxon rank sum exact test, significant differences are highlighted in bold ($p < 0.05$).

	KEMAR	NRR	PPM	PR
NRR	0.66726	-	-	-
PPM	**0.00173**	0.05306	-	-
PR	0.77302	0.79617	0.05306	-
REF	**0.00041**	**0.00041**	**0.00041**	**0.00041**

The quadrant error rate for conditions other than the reference are at or above chance level. The polar precision error, however, was estimated to be better for each condition than in the behavioural experiment, albeit the REF condition being significantly different than all other conditions. For both error metrics, significant differences other than to the reference have been found between the KEMAR and the PPM condition, which suggests that the PPM HRTFs contain direction-dependent spectral cues that are significantly different from KEMAR HRTFs.

This investigation suggests that the PPM proposal would be able to improve the PR in such a way that significant differences in global and local sound-localisation errors are achieved when compared to a non-individual, in this case a dummy-head, HRTF.

Sound-localisation performance for registrations to reference

The two proposals have been registered to REF meshes during their development process already and the concept of applying the proposals was proven in the respective chapters. However, in these proof-of-concepts investigations, I always swapped the left ear only, which the auditory model apparently was able to compensate. Additionally, the interesting case is whether the two proposals still work when registering to low-quality meshes. Thus, I investigated the sound-localisation performance in the case of the two proposals being registered to the references. The following investigation covers the scenario in which the proposals would have been registered to the left and right reference pinnae, and the head mesh would have remained the same for all conditions. Here, the auditory model was used to compare four conditions, always using the numerically calculated HRTF from REF as internal template: REF, KEMAR, NRR(PR), and PPM(PR). As in previous investigations, differences in the geometrical and acoustical domain are shown as well as sound-localisation errors from a simulated behavioural experiment.

Figure 5.13 shows J between the reference and KEMAR, and between the reference and the registrations of the two proposals to the reference. A Shapiro test rejected the Null hypothesis ($p < 0.05$) and showed non-normality in the data, and a subsequent Kruskal-Wallis test also rejected the Null hypothesis and showed that the data is unlikely drawn from the same distribution ($p < 0.05$). Finally, a Wilcoxon multiple comparison of means showed significant differences between all three conditions ($p < 0.05$).

Figure 5.14 show four different meshes: template, reference, and registration results to the reference of the two proposals for one listener. The NRR was indeed able to register

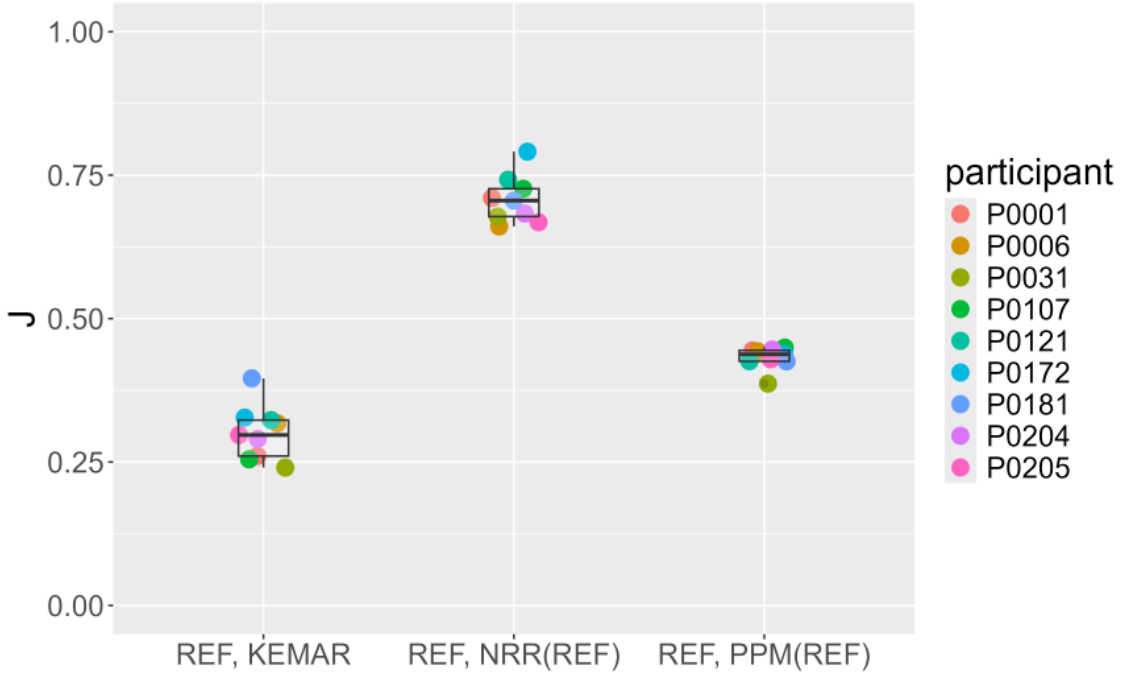


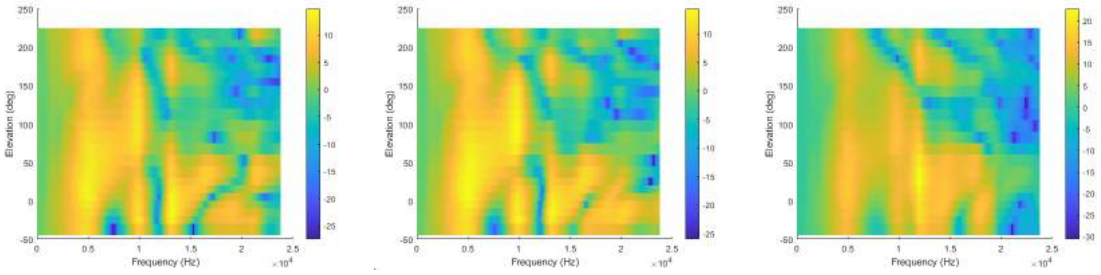
Figure 5.13: Jaccard index J between the references and the proposals applied to the references.

the target closely again, however, there are several aspects to highlight. For regions with narrow folds, ripples result from the algorithm being "indecisive" of to which region the points belong to, e.g., across the lower part of the helix. Some regions were only registered globally to the target, e.g., the tragus resembled the template much more than the target. And, finally, the error in the reference of the cymba conchae being closed, i.e., the crus helicis connected to the crus inferius anthelicis, seemed to have vanished in the registered result. The PPM managed to register not too well to the target from a geometrical point of view, because, e.g., the crus helicis shows a bend which is not covered in the parameter space of the PPM. The crus inferius anthelicis seemed to be thinner than in the reference, and the crus superius anthelicis seemed to have a different tilt.

(a) (b) (c)

Figure 5.14: Obtained meshes shown for one listener. **(a)** Structured-light scan from SONICOM database (REF), **(b)** registration result using the NRR approach, **(c)** registration result using the PPM approach.

Figure 5.15 shows the HRTFs of one listener in the median plane for the reference, the PR, and the two proposals being registered to the reference. Note the expanded dynamic range of the HRTF in the PPM condition. Note that, in the reference mesh, the cymba conchae and helix rim are mistakenly connected, but in the registrations using both proposals, this connection was not there. This connection results in a different surface for the cymba conchae region in the pinna, and together with the results from Chapter 2, the different shape of the cymba conchae might be the main reason for log-spectral differences.



(a) (b) (c)

Figure 5.15: P0172's HRTFs in the median plane, calculated from (a) reference, (b) NRR registered to reference, (c) BezierPPM registered to reference.

Figure 5.16 shows calculated HRTFs of the different meshes of a listener in three positions: the top one shows the frontal position at $[azi, ele] = [0, 0]°$, the middle one shows an elevated frontal position at $[azi, ele] = [0, 45]°$, and the bottom one shows an elevated rear position at $[azi, ele] = [180, 45]°$. In all six subfigures, the reference and the proposals registered to the reference, and the PR and the proposals registered to the PR are included, in order to compare the two registration approaches and the two targets (high-speed versus high-quality). In the frontal position, the reference shows the

first pinna-related notch at 12 kHz. The PR spectrum starts to deviate from the reference at approximately 3 kHz and shows a dynamic range of less than 15 dB, without any steep peaks or notches. The BezierPPM-registration to the PR shows a prominent peak at 5.250 kHz, slightly above the reference and the registrations to the reference, and the first notch appears at 13.687 kHz. Similar to the first peak and notch, the rest of the spectral curve contains prominent peaks and notches slightly higher than the reference. The NRR-registration to the PR shows a dynamic range of approximately 10 dB, even less than that of the PR. The BezierPPM-registration to the reference shows a prominent peak at 5.062 kHz, and a fluctuating magnitude for higher frequencies. The NRR-registration to the reference yields a similar spectral trend compared to the reference, except that the second notch at 12.375 kHz is less steeper. Similar trends can be observed for the other two positions in the median plane.

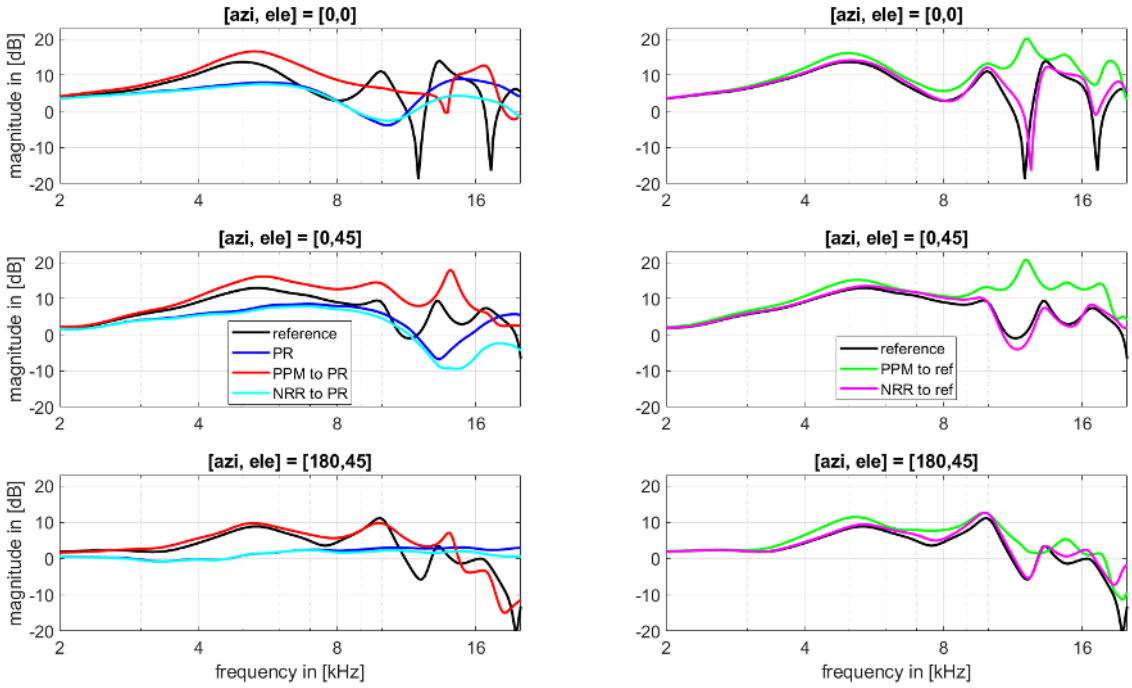


Figure 5.16: Different HRTF magnitude spectra at three different positions. Top: frontal position ($[azi, ele] = [0, 0]^\circ$), middle: elevated front ($[azi, ele] = [0, 45]^\circ$), bottom: elevated back ($[azi, ele] = [180, 45]^\circ$).

Figure 5.17 shows the log-spectral differences between the two proposals registered to the reference and the reference. A Shapiro test accepted the Null hypothesis and showed that the data was normally distributed ($p > 0.05$), and a subsequent Bartlett test confirmed homoscedasticity of the data ($p > 0.05$). Finally, a Tukey multiple comparison of means showed significant differences between all the conditions ($p < 0.05$). These findings suggest that the NRR should yield sound-localisation results closer to the reference condition than the PPM, and that both proposals should result in lower sound-localisation errors than in the KEMAR condition.

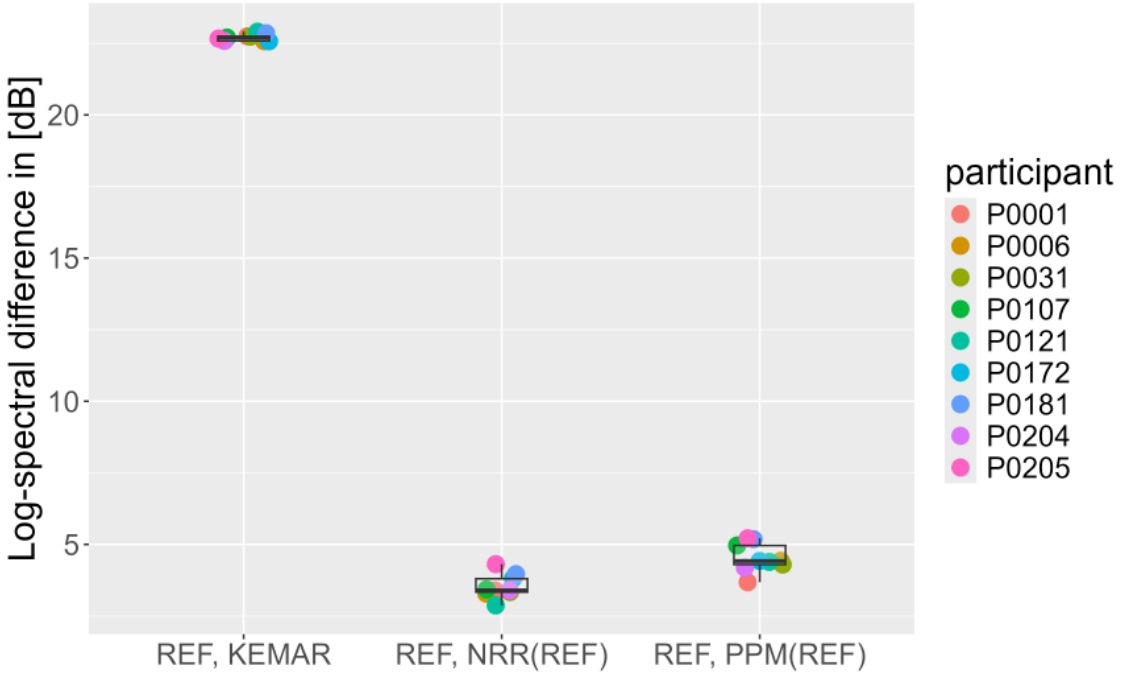


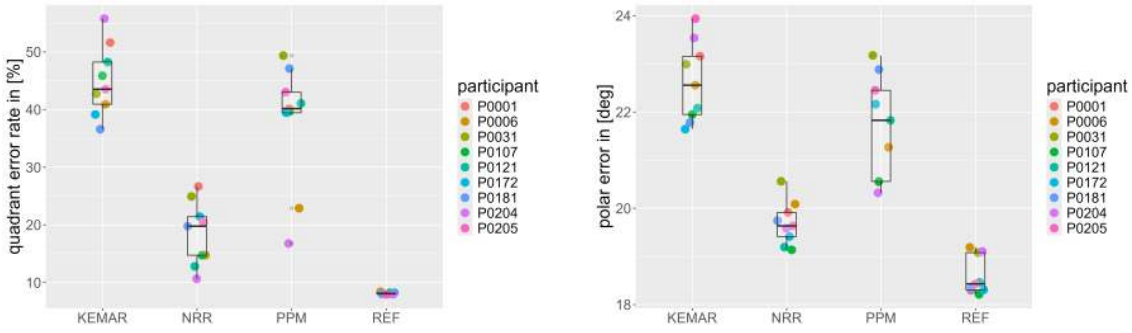
Figure 5.17: Log-spectral difference (averaged over frequency and position) between registrations to the reference and the reference.

Figure 5.18 shows the quadrant error rate and polar error for the simulated sound-localisation performance using the auditory model for sagittal-plane sound localisation [67]. Regarding the quadrant error rate, a Shapiro test rejected the Null hypothesis and showed non-normality in the data ($p < 0.05$), and a subsequent Kruskal-Wallis test revealed significant differences between groups ($p < 0.05$). Finally, a Wilcoxon multiple comparison of means showed significant differences between KEMAR and NRR, between NRR and PPM, and between the reference and all other conditions. Regarding the polar error, a Shapiro test rejected the Null hypothesis and showed non-normality of the data ($p < 0.05$), and a subsequent Kruskal-Wallis test revealed significant differences between groups ($p < 0.05$). Finally, a Wilcoxon multiple comparison of means showed significant differences between the KEMAR and NRR, between NRR and PPM, and between the reference and all other conditions. Table 5.7 shows the p values for the pairwise comparisons, significances are highlighted in bold.

Table 5.7: Pairwise comparisons of quadrant error rates (left) and polar precision errors (right) using Wilcoxon rank sum exact test, significant differences are highlighted in bold ($p < 0.05$).

	KEMAR	NRR	PPM
NRR	**0.00025**	-	-
PPM	0.22242	**0.00370**	-
REF	**0.00205**	**0.00205**	**0.00205**

	KEMAR	NRR	PPM
NRR	**0.00025**	-	-
PPM	0.11349	**0.00058**	-
REF	**0.00025**	**0.00025**	**0.00025**



(a) (b)

Figure 5.18: **(a)** Quadrant error rate and **(b)** polar error for modelled sound-localisation experiment using the auditory model. NRR and PPM were registered to the GT.

In this investigation, the median log-spectral difference between the NRR and the reference was less than 4 dB but resulted in quadrant error rates of 20 %. The log-spectral difference in the PPM condition was significantly lower than in the KEMAR condition but both conditions showed no significant differences and close-to chance results for the quadrant error rate and the polar precision error. These findings contradict previous findings of the proposals in chapters 3 and 4, in which the auditory model showed similar sound-localisation results to reference conditions. These results indicate that the auditory model is able to compensate the spectral differences if *one* ear was swapped with a NRR or PPM ear, but not if both ears are swapped.

5.4 Conclusions

In general, it can be concluded that the NRR proposal is unlikely to be useful for state-of-the-art high-speed PRs as it registers the high-quality template to the low-quality target closely. The PPM proposal seems to be the more promising approach, because the model's own parameter space limits prevent the concave regions to be overly flattened. Some spectral cues – albeit not exactly the individual ones – were reconstructed due to these limitations in the PPM. Although a slight trend for a decrease in the polar errors can be reported for the PPM condition in the behavioural experiment, I conclude that – without excessive training on the individual (numerically calculated) HRTF – the two proposals NRR and PPM do not affect listeners' performances in a sound-localisation experiment significantly when compared to the PR. Because a high-speed PR is showing almost no individual detail, it can be concluded that more work is needed on improving the quality of high-speed PR of human pinnae before the two proposals can be applied.

In the geometrical domain, the templates (NH5 and AVG) showed little similarity to the reference pinna meshes of the participants, and no significant differences between listeners was found. Based on these findings, I conclude that the selection of the template pinna to the target pinna based on similarity prior to registration is not necessary for the two proposals.

PR meshes of individuals showed little similarity to the reference – even less than, e.g., a non-individual pinna (NH5). I conclude that a non-individual but high-quality ear mesh can be used if no individual high-quality ear mesh is available instead of using a low-quality PR mesh obtained with a high-speed approach.

The NRR algorithm was able to register the targets from reference scans closer than the manual deformation approach with the PPM. This was shown because J between the PR and the non-rigid registration to the PR is significantly higher than J between the reference and the template used (NH5), and because J between the PR and the PPM registered to the PR is not significantly different from J between the reference and the template used (AVG). The findings indicate that the NRR algorithm is able to register the target closely (whether the target was the PR or the reference). Although the parameter space of the PPM is limited, I argue that the shape keys controlling concave regions could have been left in the default position in order to inflict some curvature on the otherwise flattened regions in the PR.

Neither NRR nor PPM results are significantly closer to the reference than the PR by means of J. This indicates that the proposals did not significantly change the PR meshes.

In the acoustical domain, spectral cues, especially in the median plane, were drastically affected by the PR. There were differences between the cues in the front and back of a listener, but the complete individual cues due to the pinna shape were completely eliminated.

Similar to the PR, the NRR was able to model the targets very close and, thus, yielded similar log-spectral differences and a high loss in spectral detail. The PPM, however, was able to again introduce direction-dependent as well as direction-independent cues to the magnitude spectra in the sagittal planes, and some form of individual detail was (re-)constructed.

In the behavioural domain, the PR of the head likely had no influence on the sound-localisation errors in the lateral dimension, because the lateral accuracy yielded no significant differences between conditions.

The PR and NRR condition showed a significantly different quadrant error rate compared to the reference. Albeit the PPM condition showed no significant differences to the reference (even after discarding outliers), no significant differences to the reference condition was found. The polar bias and polar error showed no significant differences between the reference and the PR. However, significant differences were found between the reference and the two proposals for the polar error. This suggests that, in this behavioural experiment, only global polar errors such as the quadrant error rate were affected by the PR, and that the two proposals even increased the local polar errors compared to the PR.

The alignment of the PR head mesh was unlikely the reason for a polar bias, because not all listeners show the lowest polar bias for the reference condition, and no significant differences for the polar bias across conditions in the behavioural experiment.

Chapter 6

Summary and Conclusions

This thesis investigated the effect of shape deformation methods applied to human pinnae in order to numerically calculate personalised head-related transfer functions (HRTFs). It aimed at providing a step towards a widely and easily available method to obtain personalised HRTF. Numerical calculation of HRTFs is a viable way for the future of HRTF personalisation, because it is easy to use and the code is open source. Compared to an acoustical HRTF measurement, the advantages of numerical calculation are clear: First, the listener does not have to travel to specific facilities which are able to acoustically measure HRTFs using elaborate and expensive professional equipment and an acoustically treated (semi-)anechoic chamber. Second, the listener does not have to stay still during the HRTF acquisition as they have to in the measurement, and they are not fatigued by the procedure. One of the obstacles to calculated personalised HRTFs, however, is the required high-quality geometry of the listener's pinnae. Advances in the photogrammetrical reconstruction (PR) as an easy-to-use solution have been made but were unsuccessful so far because of the poor quality yielding from the typical high-speed PR procedure applied outside of controlled laboratory conditions.

In this thesis, two methods were proposed to improve the quality of PR-based pinna geometries: First, several NRR algorithms were applied to register a non-individual high-quality pinna to various individual pinnae. The investigations related to NRR algorithms resulted in Bayesian formulation of coherent point drift (BCPD) as the preferred candidate. Second, I proposed a parametric pinna model based on Beziér curves (BezierPPM) and I manually registered it to various individual target pinnae. Both deformation proposals were applied to high-speed PR and evaluated in terms of sound-localisation performance obtained in a behavioural experiment and simulations.

Both proposals resulted in improved PR quality by means of increasing the similarity between the deformed shape and the reference. However, the conducted sound-localisation experiment yielded ambiguous results, likely because participants were not fully trained on the task. The conditions from the behavioural experiment were extended by a dummy-head condition and compared to a simulated sound-localisation performance using an auditory model for median-plane sound localisation. The results of the simulation suggest that the auditory model is more sensitive to small changes in HRTFs than a real listener

which rendered a fair comparison difficult.

I conclude that the deformation proposals are able to improve the geometric similarity when applied to high-quality photogrammetrical reconstructions (PRs). When applied to high-speed and, thus, often low-quality PRs, the BezierPPM yielded promising results when tested in a behavioural sound-localisation experiment. Thus, I conclude that this parametric type of geometry improvement has good chances for HRTF personalisation in the future. However, the parameter space of plausible human pinna shapes is yet to be explored, and the BezierPPM could be improved in overlapping regions controlled by multiple parameters. While the NRR algorithms were able to register a template pinna close to a target pinna, the results were not suitable for low-quality targets because they did not achieve to transfer individual geometry features while preserving the high quality of the template at the same time. Hence, either high-speed PR algorithms have to yield better results before NRR algorithms can be applied, or NRR and PR need to be combined into one registration approach.

Appendix A

BezPPM parameter version history

Table A.1: Initially proposed Beziér curves for the PPM.

i	Name
1	Lobulus
2	Helix1
3	Helix2
4	Helix3
5	Tragus
6	Antitragus
7	Antihelix
8	Crus inferius anthelicis
9	Crus superius anthelicis

Table A.2: Initially proposed shape keys for the PPM.

j	Name	Weight range w_j
1	Anthelix curvature	$-1 \leq w_1 \leq 1$
2	Anthelix inner thickness	$-1 \leq w_2 \leq 0$
3	Antitragus tilt	$-1 \leq w_3 \leq 1$
4	Antitragus yaw	$-1 \leq w_4 \leq 1$
5	Concha depth	$-0.4 \leq w_5 \leq 1$
6	Crus helicis appendix	$0 \leq w_6 \leq 1$
7	Crus helicis thickness	$0 \leq w_7 \leq 1$
8	Helix curvature	$-1 \leq w_8 \leq 1$
9	Helix (pinna circumference) thickness	$-0.6 \leq w_9 \leq 1$
10	Helix1 thickness	$-1 \leq w_{10} \leq 1$
11	Helix2 thickness	$-1 \leq w_{10} \leq 1$
12	Helix3 thickness	$-1 \leq w_{10} \leq 1$
13	Insicura intertragica depth	$-1 \leq w_{11} \leq 1$
14	Insicura intertragica width	$-1 \leq w_{12} \leq 1$
15	Insicura anterior	$-1 \leq w_{13} \leq 1$
16	Lobulus attachment	$0 \leq w_{14} \leq 1$
17	Lobulus dent	$-1 \leq w_{15} \leq 1$
18	Lobulus length	$-0.5 \leq w_{16} \leq 1$
19	Lobulus yaw	$-1 \leq w_{17} \leq 1$
20	Pinna length	$-1 \leq w_{18} \leq 1$
21	Pinna size	$-1 \leq w_{18} \leq 1$
22	Pinna width	$-1 \leq w_{18} \leq 1$
23	Scapha depth	$-0.5 \leq w_{18} \leq 1$
24	Tragus upper appendix	$-1 \leq w_{18} \leq 1$
25	Tragus lower appendix	$-1 \leq w_{18} \leq 1$
26	Tragus roll	$-1 \leq w_{18} \leq 1$
27	Tragus yaw	$-1 \leq w_{18} \leq 1$
28	Fossa triangularis depth	$-2 \leq w_{18} \leq 1$
29	Fossa triangularis width	$0 \leq w_{18} \leq 1$
30	Cymba conchae vs. Crus superius anthelicis depth	$0 \leq w_{18} \leq 1$
31	Cymba conchae depth	$0 \leq w_{18} \leq 1$
32	Cymba conchae vs. Crus helicis depth	$0 \leq w_{18} \leq 1$

Table A.3: Updated Beziér curves for the PPM.

i	Name
1	Lobulus
2	Helix low
3	Helix middle
4	Helix up
5	Antitragus
6	Antihelix
7	Crus superius anthelicis
8	Crus inferius anthelicis
9	Tragus
10	Parent

Table A.4: Updated proposed shape keys for the PPM.

j	Name	Weight range
1	Antitragus inside crease	$-1 \leq w_1 \leq 1$
2	Cavum Conchae depth	$-1 \leq w_2 \leq 1$
3	Cymba Conchae depth	$0 \leq w_3 \leq 1$
4	Crus Helicis prominence	$-1 \leq w_4 \leq 1$
5	Upper Helix depth	$0 \leq w_5 \leq 1$
6	Middle Helix depth	$0 \leq w_6 \leq 1$
7	Lower Helix depth	$0 \leq w_7 \leq 1$
8	Lobulo attachment	$0 \leq w_8 \leq 1$
9	Scapha depth	$0 \leq w_9 \leq 1$
10	Fossa Triangularis depth	$-1 \leq w_{10} \leq 1$
11	Crus Inferius Anthelicis lower crease	$0 \leq w_{11} \leq 1$
12	Crus Inferius Anthelicis upper crease	$0 \leq w_{12} \leq 1$
13	Crus Superius Anthelicis lower crease	$0 \leq w_{13} \leq 1$
14	Crus Superius Anthelicis upper crease	$0 \leq w_{14} \leq 1$
15	Tragus Upper dent	$-1 \leq w_{15} \leq 1$
16	Crus Helicis upper dent	$0 \leq w_{16} \leq 1$
17	Crus Helicis lower dent	$0 \leq w_{17} \leq 1$
18	Ear Canal diameter	$-1 \leq w_{18} \leq 1$

Chapter 7

Bibliography

[1] J. Blauert, *Spatial hearing. The Psychophysics of Human Sound Localization*. Cambridge, MA: The MIT Press, revised edition ed., 1997.

[2] P. Majdak, R. Baumgartner, and C. Jenny, "Formation of Three-Dimensional Auditory Space," in *The Technology of Binaural Understanding* (J. Blauert and J. Braasch, eds.), Modern Acoustics and Signal Processing, pp. 115–149, Cham: Springer International Publishing, 2020.

[3] M. Morimoto and H. Aokata, "Localization cues of sound sources in the upper hemisphere," *Journal of the Acoustical Society of Japan (E)*, vol. 5, no. 3, pp. 165–173, 1984.

[4] P. Majdak, M. J. Goupell, and B. Laback, "3-d localization of virtual sound sources: Effects of visual environment, pointing method, and training," *Attention, perception, & psychophysics*, vol. 72, no. 2, pp. 454–469, 2010.

[5] P. Majdak, T. Carpentier, R. Nicol, A. Roginska, Y. Suzuki, K. Watanabe, H. Wierstorf, H. Ziegelwanger, and M. Noisternig, "Spatially Oriented Format for Acoustics: A Data Exchange Format Representing Head-Related Transfer Functions," in *Proceedings of the 134th Convention of the Audio Engineering Society (AES)*, (Roma, Italy), p. Convention Paper 8880, may 2013.

[6] P. Majdak, F. Zotter, F. Brinkmann, J. De Muynke, M. Mihocic, and M. Noisternig, "Spatially oriented format for acoustics 2.1: Introduction and recent advances," *J. Audio Eng. Soc*, vol. 70, no. 7/8, pp. 565–584, 2022.

[7] M. Cuevas-Rodríguez, L. Picinali, D. González-Toledo, C. Garre, E. de la Rubia-Cuestas, L. Molina-Tanco, and A. Reyes-Lecuona, "3d tune-in toolkit: An open-source library for real-time binaural spatialisation," *PloS one*, vol. 14, no. 3, p. e0211899, 2019.

[8] D. González-Toledo, L. Molina-Tanco, M. Cuevas-Rodríguez, P. Majdak, and A. Reyes-Lecuona, "The binaural rendering toolbox. a virtual laboratory for reproducible research in psychoacoustics," in *Forum Acusticum*, 2023.

[9] P. Søndergaard and P. Majdak, "The Auditory Modeling Toolbox," in *The Technology of Binaural Listening* (J. Blauert, ed.), pp. 33–56, Berlin-Heidelberg, Germany: Springer, 2013.

[10] P. Majdak, C. Hollomey, and R. Baumgartner, "Amt 1. x: A toolbox for reproducible research in auditory modeling," *Acta Acustica*, vol. 6, p. 19, 2022.

[11] H. Møller, M. F. Sørensen, D. Hammershøi, and C. B. Jensen, "Head-related transfer functions of human subjects," *Journal of the Audio Engineering Society*, vol. 43, pp. 300–321, may 1995.

[12] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," in *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No.01TH8575)*, (New York), pp. 99–102, 2001.

[13] V. R. Algazi, R. O. Duda, R. Duraiswami, N. A. Gumerov, and Z. Tang, "Approximating the head-related transfer function using simple geometric models of the head and torso," *The Journal of the Acoustical Society of America*, vol. 112, no. 5, pp. 2053–2064, 2002.

[14] V. R. Algazi, C. Avendano, and R. O. Duda, "Elevation localization and head-related transfer function analysis at low frequencies," *The Journal of the Acoustical Society of America*, vol. 109, no. 3, pp. 1110–1122, 2001.

[15] V. C. Raykar, R. Duraiswami, and B. Yegnanarayana, "Extracting the frequencies of the pinna spectral notches in measured head related impulse responses," *The Journal of the Acoustical Society of America*, vol. 118, no. 1, pp. 364–374, 2005.

[16] H. Takemoto, P. Mokhtari, H. Kato, R. Nishimura, and K. Iida, "Mechanism for generating peaks and notches of head-related transfer functions in the median plane," *The Journal of the Acoustical Society of America*, vol. 132, pp. 3832–41, dec 2012.

[17] P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato, "Vertical normal modes of human ears: Individual variation and frequency estimation from pinna anthropometry," *The Journal of the Acoustical Society of America*, vol. 140, no. 2, pp. 814–831, 2016.

[18] E. A. Macpherson and J. C. Middlebrooks, "Listener weighting of cues for lateral angle: the duplex theory of sound localization revisited," *The Journal of the Acoustical Society of America*, vol. 111, pp. 2219–2236, May 2002.

[19] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization using nonindividualized head-related transfer functions," *The Journal of the Acoustical Society of America*, vol. 94, no. 1, pp. 111–123, 1993.

[20] H. Feldmann and G. Steimann, "Die bedeutung des äußeren ohres für das hören im wind," *Archiv für klinische und experimentelle Ohren-, Nasen-und Kehlkopfheilkunde*, vol. 190, pp. 69–85, 1968.

[21] V. Best, S. Carlile, C. Jin, and A. van Schaik, "The role of high frequencies in speech localization," *The Journal of the Acoustical Society of America*, vol. 118, no. 1, pp. 353–363, 2005.

[22] M. B. Gardner and R. S. Gardner, "Problem of localization in the median plane: effect of pinnae cavity occlusion," *The Journal of the Acoustical Society of America*, vol. 53, no. 2, pp. 400–408, 1973.

[23] R. Baumgartner, D. K. Reed, B. Tóth, V. Best, P. Majdak, H. S. Colburn, and B. Shinn-Cunningham, "Asymmetries in behavioral and neural responses to spectral cues demonstrate the generality of auditory looming bias," *Proceedings of the National Academy of Sciences*, vol. 114, pp. 9743–9748, aug 2017.

[24] J. Hebrank and D. Wright, "Spectral cues used in the localization of sound sources on the median plane," *The Journal of the Acoustical Society of America*, vol. 56, pp. 1829–34, Dec. 1974.

[25] D. W. Batteau, "The role of the pinna in human localization," *Proceedings of the Royal Society of London. Series B. Biological Sciences*, vol. 168, no. 1011, pp. 158–180, 1967.

[26] H. G. Fisher and S. J. Freedman, "The role of the pinna in auditory localization," *Journal of Auditory research*, 1968.

[27] A. D. Musicant and R. A. Butler, "The influence of pinnae-based spectral cues on sound localization," *The Journal of the Acoustical Society of America*, vol. 75, no. 4, pp. 1195–1200, 1984.

[28] B. U. Seeber and H. Fastl, "Subjective selection of non-individual head-related transfer functions," in *Proceedings of the International Conference on Auditory Display*, pp. 259–262, Georgia Institute of Technology, 2003.

[29] P. Majdak, R. Baumgartner, and B. Laback, "Acoustic and non-acoustic factors in modeling listener-specific performance of sagittal-plane sound localization," *Frontiers in psychology*, vol. 5, p. 319, 2014.

[30] C. Jenny and 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*, vol. 8, p. e17576, Sep 2020.

[31] J. C. Middlebrooks, "Individual differences in external-ear transfer functions reduced by scaling in frequency," *The Journal of the Acoustical Society of America*, vol. 106, no. 3, pp. 1480–1492, 1999.

[32] R. Algazi, C. Avendano, and R. O. Duda, "Estimation of a spherical-head model from anthropometry," *Journal of the Audio Engineering Society*, vol. 49, 2001.

[33] C. Guezenoc and R. Seguier, "Hrtf individualization: A survey," in *Audio Engineering Society Convention 145*, p. Convention Paper 10129, 2018.

[34] D. Hammershøi and H. Møller, "Sound transmission to and within the human ear canal," *The Journal of the Acoustical Society of America*, vol. 100, no. 1, pp. 408–427, 1996.

[35] S. Li and J. Peissig, "Measurement of Head-Related Transfer Functions: A Review," *Applied Sciences*, vol. 10, p. 5014, jan 2020. Number: 14 Publisher: Multidisciplinary Digital Publishing Institute.

[36] P. Majdak, P. Balazs, and B. Laback, "Multiple exponential sweep method for fast measurement of head-related transfer functions," *Journal of the Audio Engineering Society*, vol. 55, pp. 623–637, aug 2007.

[37] G. Yu, R. Wu, Y. Liu, and B. Xie, "Near-field head-related transfer-function measurement and database of human subjects," *The Journal of the Acoustical Society of America*, vol. 143, no. 3, pp. EL194–EL198, 2018.

[38] E. A. Torres-Gallegos, F. Orduna-Bustamante, and F. Arámbula-Cosío, "Personalization of head-related transfer functions (hrtf) based on automatic photo-anthropometry and inference from a database," *Applied Acoustics*, vol. 97, pp. 84–95, 2015.

[39] H. Hu, L. Zhou, J. Zhang, H. Ma, and Z. Wu, "Head related transfer function personalization based on multiple regression analysis," in *2006 International Conference on Computational Intelligence and Security*, vol. 2, pp. 1829–1832, 2006.

[40] C. Guezenoc and R. Seguier, "A wide dataset of ear shapes and pinna-related transfer functions generated by random ear drawings," *The Journal of the Acoustical Society of America*, vol. 147, no. 6, pp. 4087–4096, 2020.

[41] R. Zolfaghari, N. Epain, C. T. Jin, J. Glaunes, and A. Tew, "Large deformation diffeomorphic metric mapping and fast-multipole boundary element method provide new insights for binaural acoustics," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2863–2867, 2014.

[42] C. T. Jin, R. Zolfaghari, X. Long, A. Sebastian, S. Hossain, J. Glaunés, A. Tew, M. Shahnawaz, and A. Sarti, "Considerations regarding individualization of head-related transfer functions," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6787–6791, 2018.

[43] F. Grijalva, L. C. Martini, D. Florencio, and S. Goldenstein, "Interpolation of head-related transfer functions using manifold learning," *IEEE Signal Processing Letters*, vol. 24, no. 2, pp. 221–225, 2017.

[44] J. C. Middlebrooks, "Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency," *The Journal of the Acoustical Society of America*, vol. 106, no. 3, pp. 1493–1510, 1999.

[45] K. Iida, T. Aizaki, and T. Kikuchi, "Toolkit for individualization of head-related transfer functions using parametric notch-peak model," *Applied Acoustics*, vol. 189, p. 108610, 2022.

[46] M. Zhang, R. Kennedy, T. Abhayapala, and W. Zhang, "Statistical method to identify key anthropometric parameters in hrtf individualization," in *2011 Joint Workshop on Hands-free Speech Communication and Microphone Arrays*, pp. 213–218, IEEE, 2011.

[47] M. Zhang, Z. Ge, T. Liu, X. Wu, and T. Qu, "Modeling of individual hrtfs based on spatial principal component analysis," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 785–797, 2020.

[48] D. Lu, X. Zeng, X. Guo, and H. Wang, "Personalization of head-related transfer function based on sparse principle component analysis and sparse representation of 3d anthropometric parameters," *Acoustics Australia*, pp. 1–10, 2019.

[49] F. C. Tommasini, O. A. Ramos, M. X. Hüg, and F. Bermejo, "Usage of spectral distortion for objective evaluation of personalized hrtf in the median plane," *International Journal of Acoustics & Vibration*, vol. 20, no. 2, 2015.

[50] Q. Huang and Q. Zhuang, "Hrir personalisation using support vector regression in independent feature space," *Electronics letters*, vol. 45, no. 19, pp. 1002–1003, 2009.

[51] P. Bilinski, J. Ahrens, M. R. Thomas, I. J. Tashev, and J. C. Platt, "Hrtf magnitude synthesis via sparse representation of anthropometric features," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4468–4472, IEEE, 2014.

[52] S. Spagnol, R. Miccini, M. G. Onofrei, R. Unnthorsson, and S. Serafin, "Estimation of spectral notches from pinna meshes: Insights from a simple computational model," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 2683–2695, 2021.

[53] G. W. Lee, J. H. Lee, S. J. Kim, and H. K. Kim, "Directional audio rendering using a neural network based personalized hrtf," in *INTERSPEECH*, pp. 2364–2365, 2019.

[54] I. D. Gebru, D. Marković, A. Richard, S. Krenn, G. A. Butler, F. De la Torre, and Y. Sheikh, "Implicit hrtf modeling using temporal convolutional networks," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3385–3389, 2021.

[55] H. Hu, L. Zhou, H. Ma, and Z. Wu, "Hrtf personalization based on artificial neural network in individual virtual auditory space," *Applied Acoustics*, vol. 69, no. 2, pp. 163–172, 2008.

[56] L. Li and Q. Huang, "Hrtf personalization modeling based on rbf neural network," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3707–3710, 2013.

[57] Y. Shu-Nung, T. Collins, and C. Liang, "Head-related transfer function selection using neural networks," *Archives of Acoustics*, vol. 42, no. 3, pp. 365–373, 2017.

[58] F. Grijalva, L. Martini, S. Goldenstein, and D. Florencio, "Anthropometric-based customization of head-related transfer functions using isomap in the horizontal plane," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4473–4477, 2014.

[59] R. Miccini and S. Spagnol, "A hybrid approach to structural modeling of individualized hrtfs," in *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 80–85, 2021.

[60] Y. Zhou, H. Jiang, and V. K. Ithapu, "On the predictability of hrtfs from ear shapes using deep networks," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 441–445, IEEE, 2021.

[61] J. Lu and X. Qi, "Pre-trained-based individualization model for real-time spatial audio rendering system," *IEEE Access*, vol. 9, pp. 128722–128733, 2021.

[62] B. Zhi, D. N. Zotkin, and R. Duraiswami, "Towards fast and convenient end-to-end hrtf personalization," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 441–445, IEEE, 2022.

[63] B. F. Katz and G. Parseihian, "Perceptually based head-related transfer function database optimization," *The Journal of the Acoustical Society of America*, vol. 131, no. 2, pp. EL99–EL105, 2012.

[64] S. Spagnol, S. Scaiella, M. Geronazzo, F. Avanzini, *et al.*, "Subjective evaluation of a low-order parametric filter model of the pinna for binaural sound rendering," in *Proceedings of the 22nd International Congress on Sound and Vibration, ICSV'15*, 2015.

[65] R. Pelzer, M. Dinakaran, F. Brinkmann, S. Lepa, P. Grosche, and S. Weinzierl, "Head-related transfer function recommendation based on perceptual similarities and anthropometric features," *The Journal of the Acoustical Society of America*, vol. 148, no. 6, pp. 3809–3817, 2020.

[66] R. Miccini and S. Spagnol, "Hrtf individualization using deep learning," in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 390–395, IEEE, 2020.

[67] 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.

[68] R. Barumerli, P. Majdak, M. Geronazzo, D. Meijer, F. Avanzini, and R. Baumgartner, "A bayesian model for human directional localization of broadband static sound sources," *Acta Acustica*, vol. 7, p. 12, 2023.

[69] B. Xie, X. Zhong, and N. He, "Typical data and cluster analysis on head-related transfer functions from chinese subjects," *Applied Acoustics*, vol. 94, pp. 1–13, 2015.

[70] M. G. Onofrei, R. Miccini, R. Unnthorsson, S. Serafin, and S. Spagnol, "3d ear shape as an estimator of hrtf notch frequency," in *17th Sound and Music Computing Conference*, pp. 131–137, Sound and Music Computing Network, 2020.

[71] S. Ghorbal, T. Auclair, C. Soladie, and R. Seguier, "Pinna morphological parameters influencing hrtf sets," in *Proceedings of the 20th International Conference on Digital Audio Effects (DAFx-17)*, pp. 353–359, 2017.

[72] S. Spagnol, M. Geronazzo, and F. Avanzini, "On the relation between pinna reflection patterns and head-related transfer function features," *IEEE transactions on audio, speech, and language processing*, vol. 21, no. 3, pp. 508–519, 2012.

[73] R. J. Otte, M. J. Agterberg, M. M. Van Wanrooij, A. F. Snik, and A. J. Van Opstal, "Age-related hearing loss and ear morphology affect vertical but not horizontal sound-localization performance," *Journal of the Association for Research in Otolaryngology*, vol. 14, pp. 261–273, 2013.

[74] T. Nishino, N. Inoue, K. Takeda, and F. Itakura, "Estimation of HRTFs on the horizontal plane using physical features," *Applied Acoustics*, vol. 68, pp. 897–908, aug 2007.

[75] 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* (D. B. F. Katz and D. P. Majdak, eds.), ch. 4, pp. 1–36, Rijeka: IntechOpen, 2022.

[76] H. Ziegelwanger, W. Kreuzer, and P. Majdak, "Mesh2HRTF: Open-source software package for the numerical calculation of head-related transfer functions," in *Proceedings of the 22nd International Congress on Sound and Vibration*, (Florence, IT), pp. 1–8, July 2015.

[77] A. Andreopoulou, D. R. Begault, and B. F. Katz, "Inter-laboratory round robin hrtf measurement comparison," *IEEE Journal of selected topics in signal processing*, vol. 9, no. 5, pp. 895–906, 2015.

[78] H. Ziegelwanger, P. Majdak, and W. Kreuzer, "Numerical calculation of listener-specific head-related transfer functions and sound localization: Microphone model and mesh discretization," *The Journal of the Acoustical Society of America*, vol. 138, pp. 208–222, July 2015.

[79] H. Ziegelwanger, A. Reichinger, and P. Majdak, "Calculation of listener-specific head-related transfer functions: Effect of mesh quality," in *Proceedings of Meetings on Acoustics*, vol. 19, (Montreal, Canada), p. 050017, 2013.

[80] R. Klumpp and H. Eady, "Some measurements of interaural time difference thresholds," *The Journal of the Acoustical Society of America*, vol. 28, no. 5, pp. 859–860, 1956.

[81] M. A. Senova, K. I. McAnally, and R. L. Martin, "Localization of virtual sound as a function of head-related impulse response duration," *Journal of the Audio Engineering Society*, vol. 50, no. 1/2, pp. 57–66, 2002.

[82] A. W. Mills, "On the minimum audible angle," *The Journal of the Acoustical Society of America*, vol. 30, no. 4, pp. 237–246, 1958.

[83] J. C. Middlebrooks, "Spectral shape cues for sound localization," *Binaural and spatial hearing in real and virtual environments*, pp. 77–97, 1997.

[84] J. C. Makous and J. C. Middlebrooks, "Two-dimensional sound localization by human listeners," *The Journal of the Acoustical Society of America*, vol. 87, no. 5, pp. 2188–2200, 1990.

[85] D. R. Perrott and K. Saberi, "Minimum audible angle thresholds for sources varying in both elevation and azimuth," *The Journal of the Acoustical Society of America*, vol. 87, pp. 1728–1731, Apr. 1990.

[86] J. C. Middlebrooks and D. M. Green, "Sound localization by human listeners," *Annual review of psychology*, vol. 42, no. 1, pp. 135–159, 1991.

[87] P. Poirier, S. Miljours, M. Lassonde, and F. Lepore, "Sound localization in acallosal human listeners," *Brain*, vol. 116, no. 1, pp. 53–69, 1993.

[88] P. Voss, M. Lassonde, F. Gougoux, M. Fortin, J.-P. Guillemot, and F. Lepore, "Early-and late-onset blind individuals show supra-normal auditory abilities in far-space," *Current Biology*, vol. 14, no. 19, pp. 1734–1738, 2004.

[89] P. Senn, M. Kompis, M. Vischer, and R. Haeusler, "Minimum audible angle, just noticeable interaural differences and speech intelligibility with bilateral cochlear implants using clinical speech processors," *Audiology and Neurotology*, vol. 10, no. 6, pp. 342–352, 2005.

[90] V. Pulkki, "Localization of Amplitude-Panned Virtual Sources II: Two- and Three-Dimensional Panning," *Journal of the Audio Engineering Society*, vol. 49, no. 4, pp. 753–767, 2001.

[91] P. Bremen, M. M. van Wanrooij, and A. J. van Opstal, "Pinna cues determine orienting response modes to synchronous sounds in elevation," *Journal of Neuroscience*, vol. 30, pp. 194–204, Jan. 2010.

[92] C. Pörschmann and J. M. Arend, "Obtaining dense hrtf sets from sparse measurements in reverberant environments," in *Audio Engineering Society Conference: 2019 AES International Conference on Immersive and Interactive Audio*, Audio Engineering Society, 2019.

[93] A. O. Hogg, M. Jenkins, H. Liu, I. Squires, S. J. Cooper, and L. Picinali, "Hrtf upsampling with a generative adversarial network using a gnomonic equiangular projection," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2024.

[94] P. Lladó, K. Pollack, and N. Meyer-Kahlen, "Toward a standard listener-independent hrtf to facilitate long-term adaptation," *Journal of the Audio Engineering Society*, vol. 72, no. 4, pp. 188–192, 2024.

[95] B. F. Katz, "Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation," *The Journal of the Acoustical Society of America*, vol. 110, pp. 2440–8, nov 2001.

[96] R. Coifman, V. Rokhlin, and S. Wandzura, "The fast multipole method for the wave equation: a pedestrian prescription," *IEEE Antennas and Propagation Magazine*, vol. 35, pp. 7–12, June 1993.

[97] W. Kreuzer, P. Majdak, and Z. Chen, "Fast multipole boundary element method to calculate head-related transfer functions for a wide frequency range," *The Journal of the Acoustical Society of America*, vol. 126, pp. 1280–1290, sep 2009.

[98] W. Kreuzer, K. Pollack, P. Majdak, and F. Brinkmann, "Mesh2hrtf / numcalc: An open-source project to calculatehrtfs and wave scattering in 3d," (Aalborg, Denmark), pp. 443–452, 2022.

[99] S. Marburg, "Six boundary elements per wavelength. Is that enough?," *Journal of Computational Acoustics*, vol. 10, pp. 25–51, 2002.

[100] M. Botsch and L. Kobbelt, "A remeshing approach to multiresolution modeling," in *Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing*, pp. 185–192, 2004.

[101] T. Palm, S. Koch, F. Brinkmann, and M. Alexa, "Curvature-adaptive mesh grading for numerical approximation of head-related transfer functions," pp. 1111–1114, 08 2021.

[102] F. Brinkmann, A. Lindau, and S. Weinzierl, "On the authenticity of individual dynamic binaural synthesis," *The Journal of the Acoustical Society of America*, vol. 142, pp. 1784–1795, Oct. 2017.

[103] 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.

[104] F. Brinkmann, A. Lindau, S. Weinzierl, M. Müller-Trapet, R. Opdam, M. Vorländer, *et al.*, "A high resolution and full-spherical head-related transfer function database for different head-above-torso orientations," *Journal of the Audio Engineering Society*, vol. 65, no. 10, pp. 841–848, 2017.

[105] I. the effect of skin and hair absorption on numerically simulated HRTFs, "Brinkmann, fabian and wenneman, tim niklas and weinzierl, stefan," in *11th Forum Acusticum (EAA)*, 2025.

[106] B. F. Katz, "Boundary element method calculation of individual head-related transfer function. II. Impedance effects and comparisons to real measurements," *The Journal of the Acoustical Society of America*, vol. 110, pp. 2449–55, nov 2001.

[107] M. Otani and S. Ise, "A fast calculation method of the head-related transfer functions for multiple source points based on the boundary element method," *Acoustical science and technology*, vol. 24, no. 5, pp. 259–266, 2003.

[108] M. Otani and S. Ise, "Fast calculation system specialized for head-related transfer function based on boundary element method," *The Journal of the Acoustical Society of America*, vol. 119, pp. 2589–2598, may 2006.

[109] K. Pollack, F. Di Giusto, D. Sinev, and P. Majdak, "Spectral and psychoacoustic evaluation of headrelated transfer functions calculated at the blocked ear canal and the eardrum," in *Proceedings of Forum Acusticum*, vol. 2023, 2023.

[110] 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 *Proceedings of Forum Acusticum*, (Turin), pp. 1–8, 2023.

[111] K. Pollack and P. Majdak, "Importance of the acquisition accuracy of various pinna regions on the median-plane sound-localisation performance," in *Audio Engineering Society Convention 154*, pp. 1–7, Audio Engineering Society, 2023. 2.

[112] P. Minnaar, J. Plogsties, and F. Christensen, "Directional resolution of head-related transfer functions required in binaural synthesis," *J. Audio Eng. Soc*, vol. 53, no. 10, pp. 919–929, 2005.

[113] F. Brinkmann, W. Kreuzer, J. Thomsen, S. Dombrovskis, K. Pollack, S. Weinzierl, and P. Majdak, "Recent advances in an open software for numerical hrtf calculation," *Journal of the Audio Engineering Society*, 2023.

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

[115] M. Dinakaran, F. Brinkmann, S. Harder, R. Pelzer, P. Grosche, R. R. Paulsen, and S. Weinzierl, "Perceptually motivated analysis of numerically simulated head-related transfer functions generated by various 3d surface scanning systems," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 551–555, 2018.

[116] I. Engel, R. Daugintis, T. Vicente, A. O. Hogg, J. Pauwels, A. J. Tournier, and L. Picinali, "The SONICOM HRTF dataset," *Journal of the Audio Engineering Society*, vol. 71, no. 5, pp. 241–253, 2023.

[117] R. Greff and B. F. Katz, "Round robin comparison of hrtf simulation systems: Preliminary results," in *Audio Engineering Society Convention 123*, p. Convention Paper 7188, Audio Engineering Society, 2007.

[118] C. T. Jin, P. Guillon, N. Epain, R. Zolfaghari, A. Van Schaik, A. I. Tew, C. Hetherington, and J. Thorpe, "Creating the sydney york morphological and acoustic recordings of ears database," *IEEE Transactions on Multimedia*, vol. 16, no. 1, pp. 37–46, 2013.

[119] R. Roden and M. Blau, "The IHA database of human geometries including torso, head and complete outer ears for acoustic research," Sept. 2021. Funded by the Deutsche Forschungsgemeinschaft (DFG) – Project Nr. 352015383 – SFB 1330 HAPPAA C1.

[120] M. Dellepiane, N. Pietroni, N. Tsingos, M. Asselot, and R. Scopigno, "Reconstructing head models from photographs for individualized 3d-audio processing," in *Computer Graphics Forum*, vol. 27, pp. 1719–1727, Wiley Online Library, 2008.

[121] K. Iida, O. Nishiyama, and T. Aizaki, "Estimation of the category of notch frequency bins of the individual head-related transfer functions using the anthropometry of the listener's pinnae," *Applied Acoustics*, vol. 177, p. 107929, 2021.

[122] K. Pollack, "Evaluation of photogrammetric reconstruction for numerical computation of perceptually valid HRTFs." wissenschaftliches Toningenieur-Projekt, 2019.

[123] K. Pollack, F. Brinkmann, P. Majdak, and W. Kreuzer, "Von fotos zu personalisierter räumlicher audiowiedergabe," *e & i Elektrotechnik und Informationstechnik*, pp. 1–6, 2021.

[124] S. Ullman and S. Brenner, "The interpretation of structure from motion," *Proceedings of the Royal Society of London. Series B. Biological Sciences*, vol. 203, pp. 405–426, Jan. 1979. Publisher: Royal Society.

[125] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 2, pp. 1150–1157 vol.2, 1999.

[126] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[127] L. Donato, R. Cecchi, M. Goldoni, and D. H. Ubelaker, "Photogrammetry vs ct scan: Evaluation of accuracy of a low-cost three-dimensional acquisition method for forensic facial approximation," *Journal of Forensic Sciences*, vol. 65, no. 4, pp. 1260–1265, 2020.

[128] A. Mäkivirta, M. Malinen, J. Johansson, V. Saari, A. Karjalainen, and P. Vosough, "Accuracy of photogrammetric extraction of the head and torso shape for personal acoustic hrtf modeling," in *Audio Engineering Society Convention 148*, Audio Engineering Society, 2020.

[129] F. Di Giusto, S. van Ophem, E. Deckers, and W. Desmet, "Evaluation of the accuracy of photogrammetry for head-related transfer functions acquisition using numerical methods," in *Proceedings of DAGA*, pp. 1–4, 2021.

[130] 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, p. 53, 2023.

[131] P. J. Besl and N. D. McKay, "Method for registration of 3-d shapes," in *Sensor fusion IV: control paradigms and data structures*, vol. 1611, pp. 586–606, Spie, 1992.

[132] M. P. Reed, T. R. Vallier, and A. C. Bonifas, "Development of a parametric model of adult human ear geometry," *International Journal of Industrial Ergonomics*, vol. 107, p. 103738, 2025.

[133] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," *Computer vision and image understanding*, vol. 61, no. 1, pp. 38–59, 1995.

[134] D. Fantini, F. Avanzini, S. Ntalampiras, and G. Presti, "Hrtf individualization based on anthropometric measurements extracted from 3d head meshes," in *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, pp. 1–10, 2021.

[135] P. Stitt and B. F. G. Katz, "Sensitivity analysis of pinna morphology on head-related transfer functions simulated via a parametric pinna model," *The Journal of the Acoustical Society of America*, vol. 149, pp. 2559–2572, apr 2021.

[136] J. Fels and S. Fingerhuth, "Anthropometric data acquisition using photogrammetric techniques to obtain acoustic head-related transfer functions of children," in *Proceedings of the CTU Conference, Prague*, 2004.

[137] M. Takanen, O. Santala, and V. Pulkki, "Binaural assessment of parametrically coded spatial audio signals," in *The technology of binaural listening*, pp. 333–358, Springer, 2013.

[138] H. G. Hassager, F. Gran, and T. Dau, "The role of spectral detail in the binaural transfer function on perceived externalization in a reverberant environment," *The Journal of the Acoustical Society of America*, vol. 139, no. 5, pp. 2992–3000, 2016.

[139] R. Baumgartner and P. Majdak, "Modelle zur vorhersage der auditiven externalisierung virtueller schallquellen," in *presented at the German Convention on Acoustics (DAGA)*, 2018.

[140] H. Jo, Y. Park, and Y.-s. Park, "Analysis of individual differences in head-related transfer functions by spectral distortion," in *2009 ICCAS-SICE*, pp. 1769–1772, IEEE, 2009.

[141] S. Perrett and W. Noble, "The effect of head rotations on vertical plane sound localization," *The Journal of the Acoustical Society of America*, vol. 102, no. 4, pp. 2325–2332, 1997.

[142] P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato, "Frequency and amplitude estimation of the first peak of head-related transfer functions from individual pinna anthropometry," *The Journal of the Acoustical Society of America*, vol. 137, no. 2, pp. 690–701, 2015.

[143] J. C. Middlebrooks, J. C. Makous, and D. M. Green, "Directional sensitivity of sound-pressure levels in the human ear canal," *The Journal of the Acoustical Society of America*, vol. 86, no. 1, pp. 89–108, 1989.

[144] E. A. G. Shaw, "The External Ear," in *Auditory System* (W. D. Keidel and W. D. Neff, eds.), vol. 5 / 1, pp. 455–490, Berlin, Heidelberg: Springer Berlin Heidelberg, 1974. Series Title: Handbook of Sensory Physiology.

[145] P. A. Nelson and Y. Kahana, "Spherical harmonics, singular-value decomposition and head-related transfer function," *Journal of Sound and Vibration*, vol. 239, pp. 607–637, 2001.

[146] P. M. Hofman, J. G. Van Riswick, and A. J. Van Opstal, "Relearning sound localization with new ears," *Nature neuroscience*, vol. 1, no. 5, pp. 417–421, 1998.

[147] P. Friedrich and M. Schönwiesner, "Adaptation rate and persistence across multiple sets of spectral cues for sound localization," *The Journal of the Acoustical Society of America*, vol. 157, no. 3, pp. 1543–1553, 2025.

[148] B. O. Community, *Blender - a 3D modelling and rendering package*. Blender Foundation, Stichting Blender Foundation, Amsterdam, 2018.

[149] D. Sinev, F. Di Giusto, K. Pollack, K. Mick, and J. Peissig, "Assessment of the directional characteristics of the ear canal using 3d printed replicas and numerical simulations," in *Proceedings of Forum Acusticum*, (Turin), pp. 1–8, 2023.

[150] A. L. Yuille and N. M. Grzywacz, "The motion coherence theory," in *ICCV*, pp. 344–353, 1988.

[151] W. Birkfellner, *Applied medical image processing: a basic course*. CRC Press, 2016.

[152] F. Tajdari, T. Huysmans, and Y. Song, "Non-rigid registration via intelligent adaptive feedback control," *IEEE transactions on visualization and computer graphics*, vol. 30, no. 8, pp. 4910–4926, 2023.

[153] M. Alexa, D. Cohen-Or, and D. Levin, "As-rigid-as-possible shape interpolation," in *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, pp. 165–172, 2023.

[154] A. Agathos and P. Azariadis, "3d reconstruction of skeletal mesh models and human foot biomodel generation using semantic parametric-based deformation," *International Journal of Computers and Applications*, vol. 42, no. 2, pp. 127–140, 2020.

[155] V. Golyanik, B. Taetz, G. Reis, and D. Stricker, "Extended coherent point drift algorithm with correspondence priors and optimal subsampling," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1–9, 2016.

[156] Z. Zhou, J. Tu, C. Geng, J. Hu, B. Tong, J. Ji, and Y. Dai, "Accurate and robust non-rigid point set registration using student'st mixture model with prior probability modeling," *Scientific reports*, vol. 8, no. 1, pp. 1–17, 2018.

[157] M. Saval-Calvo, J. Azorin-Lopez, A. Fuster-Guillo, V. Villena-Martinez, and R. B. Fisher, "3d non-rigid registration using color: Color coherent point drift," *Computer Vision and Image Understanding*, vol. 169, pp. 119–135, 2018.

[158] G. T. G. . W. Sverige, "Human right ear (cropped)," 2016.

[159] D. Benbennick, "A left human ear," 2005.

[160] J. Serafin and G. Grisetti, "Nicp: Dense normal based point cloud registration," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 742–749, IEEE, 2015.

[161] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-squares fitting of two 3-d point sets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-9, no. 5, pp. 698–700, 1987.

[162] D. A. Hirshberg, M. Loper, E. Rachlin, A. Tsoli, A. Weiss, B. Corner, and M. J. Black, "Evaluating the automated alignment of 3d human body scans," in *2nd International Conference on 3D Body Scanning Technologies*, pp. 76–86, Hometrica Consulting, 2011.

[163] Y. Hu, M. Zhou, and Z. Wu, "A dense point-to-point alignment method for realistic 3d face morphing and animation," *International Journal of Computer Games Technology*, vol. 2009, no. 1, p. 609350, 2009.

[164] W. Peng, Z. Feng, C. Xu, and Y. Su, "Parametric t-spline face morphable model for detailed fitting in shape subspace," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6139–6147, 2017.

[165] A. Agathos and P. Azariadis, "Parametric-based reconstruction of 3d mesh models; towards the generation of a parametric human foot biomodel," in *Proceedings of the Eurographics Workshop on Visual Computing for Biology and Medicine*, pp. 221–222, 2015.

[166] M. M. Al Rahhal, M. L. Mekhalfi, M. Guermoui, E. Othman, B. Lei, and A. Mahmood, "A dense phase descriptor for human ear recognition," *IEEE Access*, vol. 6, pp. 11883–11887, 2018.

[167] A. Myronenko and X. Song, "Point set registration: Coherent point drift," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 12, pp. 2262–2275, 2010.

[168] J. Ma, J. Zhao, and A. L. Yuille, "Non-rigid point set registration by preserving global and local structures," *IEEE Transactions on image Processing*, vol. 25, no. 1, pp. 53–64, 2015.

[169] R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (fpfh) for 3d registration," in *2009 IEEE International Conference on Robotics and Automation*, pp. 3212–3217, 2009.

[170] J. G. Darboux, "Lessons on the general theory of surfaces and the geometric applications of infinitesimal calculus," 1887.

[171] O. Hirose, "A bayesian formulation of coherent point drift," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 7, pp. 2269–2286, 2021.

[172] P. Cignoni, M. Callieri, M. Corsini, M. Dellepiane, F. Ganovelli, and G. Ranzuglia, "MeshLab: an Open-Source Mesh Processing Tool," in *Eurographics Italian Chapter Conference* (V. Scarano, R. D. Chiara, and U. Erra, eds.), pp. 129–136, The Eurographics Association, 2008.

[173] J. Danter, R. Siegert, and H. Weerda, "Ultrasonographische haut-und knorpeldickenmessungen an gesunden und rekonstruierten ohren mit einem 20-mhz-ultraschallgerät," *Laryngo-Rhino-Otologie*, vol. 75, no. 02, pp. 91–94, 1996.

[174] D. Pompeiu, "Sur la continuité des fonctions de variables complexes," in *Annales de la Faculté des sciences de Toulouse: Mathématiques*, vol. 7, pp. 265–315, 1905.

[175] R. Rockafellar, M. Wets, and R. Wets, *Variational Analysis*. Grundlehren der mathematischen Wissenschaften, Springer Berlin Heidelberg, 2009.

[176] O. U. Aydin, A. A. Taha, A. Hilbert, A. A. Khalil, I. Galinovic, J. B. Fiebach, D. Frey, and V. I. Madai, "On the usage of average hausdorff distance for segmentation performance assessment: hidden error when used for ranking," *European Radiology Experimental*, vol. 5, no. 1, pp. 1–7, 2021.

[177] M.-P. Dubuisson and A. K. Jain, "A modified hausdorff distance for object matching," in *Proceedings of 12th international conference on pattern recognition*, vol. 1, pp. 566–568, IEEE, 1994.

[178] Y. Gao, "Efficiently comparing face images using a modified hausdorff distance," *IEE Proceedings-Vision, Image and Signal Processing*, vol. 150, no. 6, pp. 346–350, 2003.

[179] A. H. Murphy, "The finley affair: A signal event in the history of forecast verification," *Weather and Forecasting*, vol. 11, no. 1, pp. 3–20, 1996.

[180] G. Carpaneto and P. Toth, "Algorithm 548: Solution of the assignment problem [h]," *ACM Transactions on Mathematical Software (TOMS)*, vol. 6, no. 1, pp. 104–111, 1980.

[181] O. Hirose, "Geodesic-based bayesian coherent point drift," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–18, 2022.

[182] H. Braren and J. Fels, "A high-resolution head-related transfer function data set and 3d-scan of kemar," *Institute for Hearing Technology and Acoustics, RWTH Aachen University, Technical report*, 2020.

[183] R. A. Butler and R. A. Humanski, "Localization of sound in the vertical plane with and without high-frequency spectral cues," *Perception & psychophysics*, vol. 51, no. 2, pp. 182–186, 1992.

[184] K. Pollack, P. Majdak, and H. Furtado, "Application of non-rigid registration to photogrammetrically reconstructed pinna point clouds for the calculation of personalised head-related transfer functions," in *Proceedings of 49th DAGA*, (Hamburg), pp. 1–4, 2023.

[185] B. Xie, X. Zhong, D. Rao, and Z. Liang, "Head-related transfer function database and its analyses," *Science in China Series G: Physics, Mechanics and Astronomy*, vol. 50, pp. 267–280, jun 2007.

[186] B. Xie, *Head-related transfer function and virtual auditory display*. J. Ross Publishing, 2013.

[187] R. Bomhardt, M. de la Fuente Klein, and J. Fels, "A high-resolution head-related transfer function and three-dimensional ear model database," in *Proceedings of Meetings on Acoustics 172ASA*, vol. 29, p. 050002, 2016.

[188] F. Brinkmann, "The fabian head-related transfer function data base," 2017.

[189] S. Ghorbal, X. Bonjour, and R. Séguier, "Computed hrirs and ears database for acoustic research," in *Audio Engineering Society Convention 148*, Audio Engineering Society, 2020.

[190] J. Y.-C. Lee and A. Joneja, "A study on parametric shape modifications of 3d skeletal models," *Computer-Aided Design and Applications*, vol. 11, no. 1, pp. 90–98, 2014.

[191] J. Day and N. Davidenko, "Parametric face drawings: A demographically diverse and customizable face space model," *Journal of vision*, vol. 19, no. 11, pp. 7–7, 2019.

[192] D. Zeng, L. Spreeuwers, R. Veldhuis, and Q. Zhao, "Combined training strategy for low-resolution face recognition with limited application-specific data," *IET Image Processing*, vol. 13, no. 10, pp. 1790–1796, 2019.

[193] S.-Y. Fang and J.-J. Fang, "Automatic head and facial feature extraction based on geometry variations," *Computer-Aided Design*, vol. 43, no. 12, pp. 1729–1739, 2011.

[194] J. Zhang, Y. Luximon, P. Shah, and P. Li, "3d statistical head modeling for face/head-related product design: A state-of-the-art review," *Computer-Aided Design*, vol. 159, p. 103483, 2023.

[195] A. Golovinskiy, W. Matusik, H. Pfister, S. Rusinkiewicz, and T. Funkhouser, "A statistical model for synthesis of detailed facial geometry," *ACM Transactions on Graphics (TOG)*, vol. 25, no. 3, pp. 1025–1034, 2006.

[196] S. Vera, D. Gil, A. Borràs, M. G. Linguraru, and M. A. Gonzalez Ballester, "Geometric steerable medial maps," *Machine vision and applications*, vol. 24, pp. 1255–1266, 2013.

[197] R. Tan, V. Osman, and G. Tan, "Ear size as a predictor of chronological age," *Archives of gerontology and geriatrics*, vol. 25, no. 2, pp. 187–191, 1997.

[198] C. Niemitz, M. Nibbrig, and V. Zacher, "Human ears grow throughout the entire lifetime according to complicated and sexually dimorphic patterns—conclusions from a cross-sectional analysis," *Anthropologischer Anzeiger*, pp. 391–413, 2007.

[199] I. Baran and J. Popović, "Automatic rigging and animation of 3d characters," *ACM Transactions on graphics (TOG)*, vol. 26, no. 3, pp. 72–es, 2007.

[200] N. Sarafianos, B. Boteanu, B. Ionescu, and I. A. Kakadiaris, "3d human pose estimation: A review of the literature and analysis of covariates," *Computer Vision and Image Understanding*, vol. 152, pp. 1–20, 2016.

[201] K. Pollack, P. Majdak, and H. Furtado, "A Parametric Pinna Model for the Calculations of Head-Related Transfer Functions," in *Proceedings of Forum Acusticum*, (Lyon), pp. 1357–1360, 2020.

[202] K. Pollack and P. Majdak, "Evaluation of a parametric pinna model for the calculation of head-related transfer functions," in *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, pp. 1–5, IEEE, 2021.

[203] K. Pollack, F. Pausch, and P. Majdak, "Parametric pinna model for a realistic representation of listener - specific pinna geometry," in *Proceedings: A21, Virtual Acoustics, ICA 2022*, (Gyeongju), pp. 168–178, 2022.

[204] F. Perfler, F. Pausch, K. Pollack, N. Holighaus, and P. Majdak, "Parametric model of the human pinna based on bézier curves and concave deformations," *Computers in Biology and Medicine*, vol. 188, p. 109817, 2025.

[205] P. M. Prendergast, "Anatomy of the external ear," in *Advanced Cosmetic Otoplasty: Art, Science, and New Clinical Techniques*, pp. 15–21, Springer, 2013.

[206] S. Harder, R. R. Paulsen, M. Larsen, S. Laugesen, M. Mihocic, and P. Majdak, "A framework for geometry acquisition, 3-D printing, simulation, and measurement of head-related transfer functions with a focus on hearing-assistive devices," *Computer Aided Design*, vol. 75-76, pp. 39–46, June 2016.

[207] A. G. Hunter and T. Yotsuyanagi, "The external ear: more attention to detail may aid syndrome diagnosis and contribute answers to embryological questions," *American Journal of Medical Genetics Part A*, vol. 135, no. 3, pp. 237–250, 2005.

[208] C. C. Veugen, F. G. Dikkers, and B. S. de Bakker, "The developmental origin of the auricula revisited," *The Laryngoscope*, vol. 130, no. 10, pp. 2467–2474, 2020.

[209] C. S. Karmody and D. J. Annino, "Embryology and anomalies of the external ear," *Facial plastic surgery*, vol. 11, no. 04, pp. 251–256, 1995.

[210] B. M. Carlson, "Special senses—vision and hearing," *The Human Body*, pp. 177–207, 2019.

[211] J. Niamtu III, "Cosmetic otoplasty and related ear surgery," in *Cosmetic facial surgery*, pp. 473–532, Elsevier, 2018.

[212] W. His, *Anatomie menschlicher Embryonen: Zur Geschichte der Organe. 1885. iv, 260 p. 156 illus. Atlas pl. IX-XIV, I*, vol. 3. Vogel, 1885.

[213] G. L. Streeter, "Development of the auricle in the human embryo," *Contrib Embryol*, vol. 69, p. 111, 1922.

[214] F. Wood-Jones and I.-C. Wen, "The development of the external ear," *Journal of anatomy*, vol. 68, no. Pt 4, p. 525, 1934.

[215] J. Midera, "A study on auricular development using scanning electron microscope," *J Jpn Plastic Reconstr Surg*, vol. 2, pp. 1–13, 1982.

[216] J. Davis, "Surgical embryology," in *Aesthetic and Reconstructive Otoplasty: Under the Auspices of the Alfredo and Amalia Lacroze de Fortabat Foundation*, pp. 93–125, Springer, 1987.

[217] C. Park and T. S. Roh, "Congenital upper auricular detachment," *Plastic and reconstructive surgery*, vol. 104, no. 2, pp. 488–490, 1999.

[218] M. Minoux, C. F. Kratochwil, S. Ducret, S. Amin, T. Kitazawa, H. Kurihara, N. Bobola, N. Vilain, and F. M. Rijli, "Mouse hoxa2 mutations provide a model for microtia and auricle duplication," *Development*, vol. 140, no. 21, pp. 4386–4397, 2013.

[219] T. C. Cox, E. D. Camci, S. Vora, D. V. Luquetti, and E. E. Turner, "The genetics of auricular development and malformation: new findings in model systems driving future directions for microtia research," *European journal of medical genetics*, vol. 57, no. 8, pp. 394–401, 2014.

[220] P. Joshi, M. Meyer, T. DeRose, B. Green, and T. Sanocki, "Harmonic coordinates for character articulation," *ACM transactions on graphics (TOG)*, vol. 26, no. 3, pp. 71–es, 2007.

[221] M. Bastioni, S. Re, and S. Misra, "Ideas and methods for modeling 3d human figures: the principal algorithms used by makehuman and their implementation in a new approach to parametric modeling," in *Proceedings of the 1st Bangalore annual compute conference*, pp. 1–6, 2008.

[222] F. Bogo, J. Romero, M. Loper, and M. J. Black, "Faust: Dataset and evaluation for 3d mesh registration," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3794–3801, 2014.

[223] A. Mudry and W. Pirsig, "Auricular hematoma and cauliflower deformation of the ear: from art to medicine," *Otology & Neurotology*, vol. 30, no. 1, pp. 116–120, 2009.

[224] N. S. Scheinfeld, N. B. Silverberg, J. M. Weinberg, and V. Nozad, "The preauricular sinus: a review of its clinical presentation, treatment, and associations," *Pediatric dermatology*, vol. 21, no. 3, pp. 191–196, 2004.

[225] T. Y. Loh and P. R. Cohen, "Darwin's tubercle: review of a unique congenital anomaly," *Dermatology and Therapy*, vol. 6, no. 2, pp. 143–149, 2016.

[226] R. Valeriani, G. Firmani, and M. Valeriani, "Correction of darwin's tubercle with plasma exeresis," *Plastic and Reconstructive Surgery–Global Open*, vol. 10, no. 10, p. e4556, 2022.

[227] T. Yotsuyanagi, Y. Nihei, Y. Shinmyo, and Y. Sawada, "Stahl's ear caused by an abnormal intrinsic auricular muscle," *Plastic and reconstructive surgery*, vol. 103, no. 1, pp. 171–174, 1999.

[228] G. Ferraro, A. Perrotta, F. Rossano, and F. D'Andrea, "Stahl syndrome in clinical practice," *Aesthetic plastic surgery*, vol. 30, no. 3, pp. 348–349, 2006.

[229] A. Nassar, J. Naba, D. Eid, S. Abou Zeid, and J. Demian, "Stahl's ear correction: A review of surgical approaches," *JPRAS open*, 2025.

[230] K. C. Poole, J. Meyer, V. Martin, R. Daugintis, N. Marggraf-Turley, J. Webb, L. Pirard, N. La Magna, O. Turvey, and L. Picinali, "The extended sonicom hrtf dataset and spatial audio metrics toolbox," in *11th Forum Acusticum (EAA)*, 2025.

[231] M. Pizzo, E. Viola, F. Solari, and M. Chessa, "Evaluation of 3d reconstruction techniques for the blending of real and virtual environments," in *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 360–367, IEEE, 2024.

[232] A. Meshram, R. Mehra, H. Yang, E. Dunn, J.-M. Franm, and D. Manocha, "P-hrtf: Efficient personalized hrtf computation for high-fidelity spatial sound," in *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 53–61, IEEE, 2014.

[233] M. Kazhdan and H. Hoppe, "Screened poisson surface reconstruction," *ACM Transactions on Graphics (ToG)*, vol. 32, no. 3, pp. 1–13, 2013.

[234] R. Daugintis, R. Barumerli, M. Geronazzo, and L. Picinal, "Initial evaluation of an auditory-model-aided selection procedure for non-individual hrtfs," in *10th Convention of the European Acoustics Association*, 2023.

[235] D. R. Begault, E. M. Wenzel, and M. R. Anderson, "Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source," *Journal of the Audio Engineering Society*, vol. 49, no. 10, pp. 904–916, 2001.