

A LOW-COST SOLUTION TO 3D PINNA MODELING FOR HRTF PREDICTION

L. Bonacina, A. Canclini, F. Antonacci, M. Marcon, A. Sarti, S. Tubaro

Politecnico di Milano
Dipartimento di Elettronica, Informazione, e Bioingegneria
Via Ponzio 34/5, 20133, Milano, Italy

ABSTRACT

We propose an infrared (IR) stereo-vision system for estimating the 3D model of the pinna, based on low-cost devices. A commercial IR calibrated stereo camera is used in conjunction with a structured IR light projector, to acquire highly textured snapshots of the pinna. A point cloud is computed for each snapshot by triangulating the stereo correspondences detected in the acquired IR images. A complete 3D model is computed by aligning and merging the point clouds, and then creating a polygonal mesh surface. The nominal accuracy of the proposed system turns to be about 1 mm, which enables an accurate prediction of the Head Related Transfer Function (HRTF) through numerical acoustic simulation.

Index Terms— 3D pinna model, HRTF, binaural audio

1. INTRODUCTION

Sound reproduction over headphones has become more and more popular in the last few years, thanks to the increased availability of high-quality transducers at a relative low cost. Binaural audio represents one of the most promising technologies in this scenario, as it enables 3D sound spatialization using only two reproduction channels. It is based on the knowledge of the Head Related Transfer Function (HRTF), which represents the acoustic response of any source in space with respect to the listener's ears. The HRTF embeds information about the response of the listener's body, which is mostly determined by the shape of the auricles (pinnae) and by the reflections produced by the torso, the shoulders and the head. Although the literature is plenty of methods attempting to derive general listener-independent acoustic models [1, 2, 3, 4, 5], accurate 3D sound reproduction is generally possible only when a customized HRTF is available.

Individual HRTFs are typically measured capturing the sound in the ear canals by means of a pair of in-ear microphones. One or more compact loudspeakers are driven with wideband signals to identify the source-ear acoustic channels, at every frequencies within the audible range. In order to span all the possible directions, either the sources or the listener are rotated by means of precision rotating arms or turntables. To obtain accurate results, measurements must be carried under anechoic conditions, where the listener is required to sit still for the entire acquisition session. These practical limitations make the acoustic measurement of the HRTF a "privilege" for few selected listeners, mainly belonging to the acoustic research community.

An interesting alternative to acoustic measurements is that of predicting the HRTF from the knowledge of the 3D model of the

pinnae and the head of the listener. Techniques based on the Finite Element Method (FEM), as well on the Boundary Element Method (BEM), can be used for simulating the source-ear acoustic propagation to obtain an estimate of the HRTF [6, 7, 8]. Recently, C. Jin et al released the SYMARE database [9], i.e., a collection of high-resolution surface meshes of 61 subjects obtained from magnetic resonance imaging (MRI) data, along with the measured HRTFs and their predictions obtained through a Fast Multipole-Boundary Element Method (FM-BEM) technique. They showed that the predicted HRTFs are highly correlated with the measured ones, especially at frequencies below 10 kHz.

The valuable work in [9] suggests that the acquisition of individualized HRTFs from 3D models could potentially become an affordable task, thus paving the way to consumer personalized binaural audio applications. Clearly, the only technical limitation is represented by the 3D acquisition system. In fact, MRI data provide extremely accurate models, but turn out to be prohibitive for low-end commercial applications. Systems such as laser-scanners (used for instance in [10] to acquire the pinna 3D model) are less expensive than MRI, but not cheap enough to be considered as consumer electronics devices. A practically costless solution, on the other end, would be that of using simple 2D data (i.e., images) to infer the HRTF. To this end, authors in [11] investigated the relationship between the measured HRTF and some anthropometric features extracted from pinna pictures. Some interesting results were found relatively to the frontal medial plane, however the intuitions in [11] are still far from enabling an accurate prediction of the whole HRTF.

In this paper we propose a low-cost alternate solution for acquiring the 3D model of the pinna, using standard low-end devices. In particular, we adopt an infrared (IR) calibrated stereo camera to acquire pairs of images of the pinna. To enrich the natural texture of the cartilage, the pinna is illuminated with a structured IR pattern. The projected pattern increases the number of matching points extracted from each image pair, which are triangulated to obtain a dense 3D point cloud for each of the stereo snapshots. The 3D point clouds are then aligned and merged to obtain a complete 3D model, from which a surface mesh is finally produced. The experimental results demonstrate that the resolution of the proposed method is sufficient for accurately predicting the HRTF through FM-BEM simulation.

The rest of the paper is structured as follows. Section 2 describes the methodology developed to extract the 3D model of the pinna, along with the hardware components and the computer vision algorithms employed to realize the proposed system. In Section 3, the method is validated from the geometrical standpoint. In Section 4 we present the results of some acoustical simulations, aimed at assessing the effectiveness of the method for HRTF prediction. Some final considerations are reported in Section 5.

The authors would like to thank Prof. Craig Jin and the CARLab research team for sharing the SYMARE database.

2. METHODOLOGY AND SYSTEM IMPLEMENTATION

The goal of the proposed system is to build a 3D surface mesh of the pinna. To this end, we devise a stereo-vision methodology for recovering a 3D model from synchronized IR image pairs. A textured IR light pattern is projected onto the ear surface, in order to produce a controlled illumination. This solution turns out to be advantageous with respect to using standard cameras (i.e., sensible to the light in the visible range), as it makes the system independent from the natural illumination conditions. Consequently, it allows the acquisition in any kind of environment.

The acquisition process can be summarized as follows. After a preliminary calibration of the stereo camera, we take a stereo snapshot of the pinna and we build a point cloud by triangulating the point correspondences extracted from the two views. In order to cope with the self-occlusions caused by the complex geometric structure of the pinna, the procedure is repeated for multiple stereo snapshots taken at different positions around the head of the subject. The resulting point clouds are then aligned and merged to form a global cloud, which is eventually meshed into a connected network of polygons approximating the 3D surface of the pinna.

In the following we provide details about the proposed methodology, describing the algorithms and the hardware components adopted for the system implementation.

Stereo-image acquisition The core component of the proposed acquisition system is the LeapMotion[®] controller¹, a low-cost commercial device designed for 3D motion and gesture control, originally thought for virtual reality and gaming applications. To acquire stereo images of the pinna, we exploit the built-in IR stereo camera, constituted by a pair of hardware-synchronized IR CMOS sensors with wide-angled lenses. As the two images of the stereo pair are acquired simultaneously, the subject is not forced to stay still during the acquisition.

The LeapMotion[®] controller is also equipped with three IR leds for illuminating the scene. Unfortunately, the intense illumination produced by these leds prevents the observation of the fine veining structure composing the ear cartilage, which would be useful to detect matching points in the two imaged views. To overcome this issue, we have chosen to deactivate the leds and replacing them with an external IR projector. More specifically, we project a structured IR pattern by means of a Microsoft Kinect[®] device, with the aim of producing a high number of point correspondences in the acquired stereo pair².

To obtain stereo images suitable for 3D reconstruction, it is necessary to compensate the radial distortion introduced by the wide-angled lenses. Indeed, distorted stereo images do not satisfy the geometric constraints arising from ideal rectilinear projection, and would lead to severely impaired 3D reconstructions. A preliminary correction is performed by remapping the raw image $\mathbf{I}_R(x, y)$ onto a grid of compensated pixel positions, namely

$$\mathbf{I}_U(x, y) = \mathbf{I}_R(f_x(x, y), f_y(x, y)),$$

where the compensation functions $f_x(x, y)$ and $f_y(x, y)$ are provided as lookup tables by the manufacturer. A finer compensation procedure is then accomplished in the calibration step described in the next paragraph. An example of undistorted stereo snapshot is shown in Figure 1.

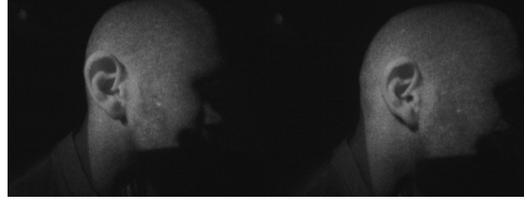


Fig. 1: Example of undistorted stereo snapshot.

Stereo-camera calibration To recover 3D metric information about the imaged object, a calibration of the stereo system is in order. More specifically, we need to determine the camera projection parameters, along with the mutual pose of the two cameras. This information is embedded in the camera projection matrices \mathbf{P}_i and \mathbf{P}_j , which describe the mapping from 3D world points to 2D points in the images; and in the fundamental matrix \mathbf{F} , which defines the geometric relationship between the pairs of corresponding points (epipolar geometry) [12]. To estimate these quantities, we follow a two-step state-of-the-art calibration procedure [13], using a planar checkerboard as a calibration object with known 3D geometry (12×7 black and white squares of size 25 mm). In a first stage, we obtain independent estimates of \mathbf{P}_i and \mathbf{P}_j , along with an accurate estimation of the residual radial distortion parameters, used to apply a finer radial correction. In a second stage, the epipolar geometry constraints [12] are exploited to iteratively compute the fundamental matrix \mathbf{F} and refining the camera projection matrices estimates.

Finding stereo correspondences The IR pattern projected onto the pinna allows us to easily determine the point-to-point correspondences between the two views. After correcting the radial distortion using the parameters estimated during the calibration stage, points³ $\mathbf{x}_i = [x_i, y_i, 1]^T$ in the left view are matched against points $\mathbf{x}_j = [x_j, y_j, 1]^T$ in the second view among those lying on the epipolar line, i.e., satisfying the epipolar constraint $\mathbf{x}_i^T \mathbf{F} \mathbf{x}_j = 0$ [12]. More specifically, we employ a template matching technique acting on the neighborhood of the tested pixel locations. Matching pairs are selected as those minimizing the Sum of Squared Distances (SSD) similarity metric.

Point cloud computation The acquisition process consists in shooting multiple stereo snapshots, changing several times the viewpoint for accurately capturing the folds and cavities constituting the pinna. For each of them, we collect N_v pairs of matching points from the two views, $v = 1, \dots, V$ denoting the index of the snapshot. Knowing the camera matrices \mathbf{P}_i and \mathbf{P}_j , the stereo correspondences are triangulated to estimate the 3D locations of the imaged points. Robust triangulation is achieved through the Direct Linear Transformation (DLT) algorithm [12]. The resulting 3D points are then collected to build V point clouds, defined as

$$\mathcal{P}_v = \{\mathbf{X}_{n,v}\}_{n=1}^{N_v},$$

where $\mathbf{X}_{n,v} = (x_{n,v}, y_{n,v}, z_{n,v})$ denotes the 3D Cartesian coordinates of the n th point extracted from the v th snapshot.

Point clouds merging As the camera movement around the pinna is unknown, it follows that the resulting V point clouds are referred to different local coordinate systems. A procedure to correctly merge the point clouds is therefore in order. To this end, we adopt the Iterative Closest Point (ICP) [14] registration algorithm. It iteratively searches for the rigid transformation that minimizes the misalignment of a pair of point clouds. The first point cloud \mathcal{P}_1 is

³Point locations are here expressed in homogeneous coordinates.

¹www.leapmotion.com

²Note that the resolution of the depth map produced by the Kinect[®] is too low to be used for extracting the 3D model of the pinna.

selected as a reference, and the ICP algorithm is run on the $V - 1$ pairs that include it. The resulting realigned clouds $\mathcal{P}_2^* \dots \mathcal{P}_V^*$ now share a common reference system, thus they can be safely merged into the global point cloud $\overline{\mathcal{P}} = \mathcal{P}_1 \cup \mathcal{P}_2^* \cup \dots \mathcal{P}_V^*$.

Surface meshing To obtain a connected model of the pinna, the global point cloud is finally converted to a surface mesh model. This task is accomplished through the Poisson meshing algorithm described in [15], implemented in most of the free 3D modeling softwares. It is worth noticing that a mesh model represents the input of any FEM/BEM-based algorithms used to predict the HRTF via acoustic simulation.

3. GEOMETRICAL VALIDATION

In this section we assess the validity of the proposed system from the geometrical standpoint. We first provide an estimation of the nominal accuracy of the system, analyzing the results obtained during the calibration phase. Then, we analyze the 3D models extracted acquiring the left pinnae of two different subjects.

3.1. Nominal system accuracy

We are first interested in evaluating the system in a controlled scenario, in order to empirically derive its nominal accuracy. To do so, we consider the 3D reconstructions of the calibration checkerboard. In this simple scenario, it is reasonable to assume that no errors are introduced in finding the point correspondences in the two views, as the corners of the checkerboard's squares can be easily detected through basic image processing techniques (i.e., straight line fitting and intersection) [16]. Therefore, we assume the 3D reconstruction error to be merely determined by the resolution of the camera and the residual (uncompensated) radial distortion.

To infer the accuracy of the system, we analyze the mean reprojection error [12] of the 3D reconstruction, as follows. Consider a pair of point correspondences $\mathbf{x}_i = [x_i, y_i, 1]^T$ and $\mathbf{x}_j = [x_j, y_j, 1]^T$, generated by the 3D point $\mathbf{X} = [X, Y, Z, 1]^T$, all expressed in homogeneous coordinates. Let $\hat{\mathbf{X}}$ be the estimate of \mathbf{X} obtained through the DLT method. The reprojections of $\hat{\mathbf{X}}$ in the two views are given by $\hat{\mathbf{x}}_i = [\hat{x}_i, \hat{y}_i, 1]^T = \mathbf{P}_i \hat{\mathbf{x}}_i$ and $\hat{\mathbf{x}}_j = [\hat{x}_j, \hat{y}_j, 1]^T = \mathbf{P}_j \hat{\mathbf{x}}_j$, respectively. The Euclidean distance between \mathbf{x}_i (\mathbf{x}_j) and $\hat{\mathbf{x}}_i$ ($\hat{\mathbf{x}}_j$) is called reprojection error.

Averaging the results obtained during the calibration phase (30 snapshots of the checkerboard), we obtained a mean reprojection error of 0.31 pixel. Following the procedure outlined in Chapter 12.6 of [12], we estimated the uncertainty of the 3D reconstruction. Assuming the reprojection error to be Gaussian and identically distributed in the x and y directions in both the views, the maximum reconstruction error equals 1.07 mm; and its mean value is 0.61 mm.

3.2. Accuracy of the extracted 3D pinna models

The proposed methodology has been tested by acquiring the 3D model of the left pinna of two people (denoted as subjects A and B throughout the paper). A ground-truth 3D model of the pinnae was obtained through a ROMER RS2 high-precision laser scanner, mounted on a 6-DOF absolute positioning arm, whose nominal accuracy is $30 \mu\text{m}$. The subjects were lying on the right side, on a rigid table, with their head resting on a pillow. This position facilitated them in remaining perfectly still during the scanning operation, which lasted about 5 minutes per person. Differently, the acquisition of the pinnae with the proposed technique was accomplished with the subjects sitting on a chair, as in this case no stillness is required. 20 stereo snapshots were acquired from different viewpoints of the

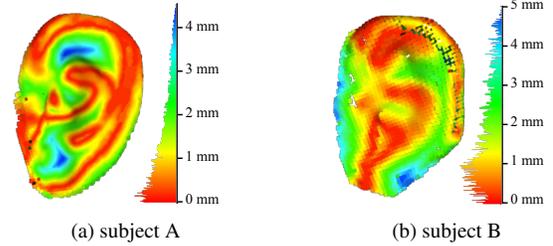


Fig. 2: 3D reconstruction error.

two pinnae, keeping the LeapMotion[®] controller and the Kinect[®] projector at a distance of about 30 cm from the subject's head.

To enable a comparison with the ground-truth, we first realigned the extracted models with those obtained through the laser scanner, using the ICP algorithm. Then, for each point $\hat{\mathbf{X}}_n$ of the extracted model, $n = 1, 2, \dots, N$, we computed the 3D reconstruction error as

$$E(\hat{\mathbf{X}}_n) = \min_{g=1,2,\dots,G} \|\hat{\mathbf{X}}_n - \overline{\mathbf{X}}_g\|_2,$$

where $\overline{\mathbf{X}}_g$, $g = 1, 2, \dots, G$, are the points constituting the ground-truth model. The distribution of the reconstruction error for the two subjects is shown in Figures 2-a and 2-b, respectively. The error distribution is almost independent from the subject, and the average error is below 2mm; the maximum error is 4.47 mm for subject A; and 5.02 mm for subject B. Notice that, in general, the reconstruction error is higher than the nominal accuracy of the system reported in Section 3.1. This is mainly due to the fact that the resolution of the IR pattern projected by the Kinect[®] is lower than that of the IR cameras, thus causing a slight worsening in the matching procedure to find stereo correspondences. Notice also that the regions exhibiting the highest error correspond to the most occluded parts of the pinna (e.g., deep cavities). Instead, the principal contours (i.e., helix, antihelix, anterior notch, etc.), which are much more representative of the shape of the pinna (and reasonably more related to perception) are reconstructed with higher precision.

4. IMPACT ON HRTF PREDICTION

We now analyze the suitability of the proposed 3D reconstruction method for predicting the HRTF using acoustic simulation. We do so following the same approach as in [9], i.e., using the FM-BEM solver provided by the Coustyx software. To simulate the effect of the head, the extracted pinna models were manually merged with a head model, randomly selected from the SYMARE database [9]. In particular, we removed the original left pinna, which was replaced with the one extracted through the proposed method. This operation was repeated for the two subjects, building the meshes for both the estimated and the ground-truth pinna models. An example is reported in Figure 3, showing the mesh model of subject A, observed from different viewpoints.



Fig. 3: Different views of the mesh model of subject A.

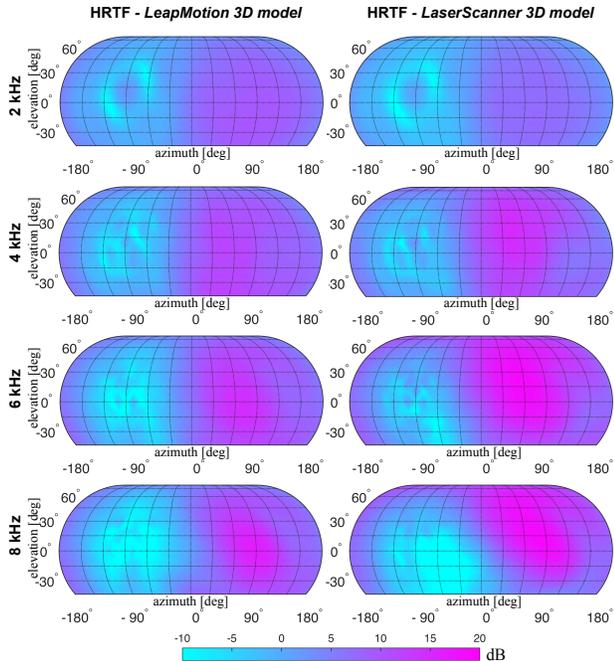


Fig. 4: Examples of SFRSs computed for subject B.

We predicted the left HRTF of the two subjects considering 393 source directions. Simulations were limited to the frequency range 20 Hz – 8 kHz, in which most of the localization cues are included [17]. Notice also that, within this range, a good matching between the measured HRTF and its prediction is expected [9]. For further details about the acoustic simulation setup, we refer the reader to Section II-D in [9]. In Figure 4 we show some examples of spatial frequency response surfaces (SFRSs), relative to the HRTF predicted for subject B. More specifically, the SFRS maps the magnitude of the HRTF for every direction in space (azimuth ϕ and elevation θ), for a selected frequency. The horizontal plane with respect to the listener’s head is given by $\theta = 0$; the median plane is defined for $\phi = 0$; the left side is mapped for $\phi > 0^\circ$ and the right side for $\phi < 0^\circ$. The left column in Figure 4 corresponds to the HRTF predicted from the 3D model extracted with the proposed LeapMotion-based method; the right column is relative to the HRTF predicted from the ground-truth 3D model. We notice that, up to 6 kHz, the two 3D models lead to very similar predictions; at 8 kHz some deviations are noticeable in the two SFRSs.

In order to objectively quantify the deviations between the HRTFs we resort to the spectral distortion (SD) metric, widely accepted in the literature as it gives some insights about spatial sound perception [18, 19, 11]. Given a predicted HRTF function $H(\phi, \theta, f_k)$, defined on the spatial angular domain (ϕ, θ) and discretized at frequencies f_k , $k = 1, \dots, F$, the spectral distortion with respect to a reference HRTF function $\bar{H}(\phi, \theta, f_k)$ is computed as [11]

$$\text{SD}(\phi, \theta) = \sqrt{\frac{1}{K} \sum_{k=1}^F \left(20 \log_{10} \frac{|\bar{H}(\phi, \theta, f_k)|}{|H(\phi, \theta, f_k)|} \right)^2} \quad [\text{dB}].$$

The resulting SDs from the two subjects, computed in the analyzed frequency range 20 Hz – 8 kHz, are shown in Figure 5. The spectral distortion turns out to be almost constant, around 2 dB, except

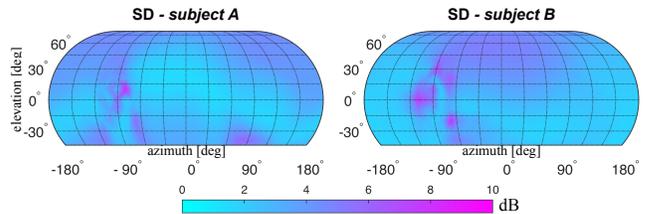


Fig. 5: Spectral distortion of the predicted HRTFs.

for a few isolated directions around $\phi = -90^\circ$ where $\text{SD} > 6$ dB. Averaging the SD over all the spatial directions, we obtained a mean value of 3.01 dB for subject A and 2.98 dB for subject B. Notice that values of SD below 4 dB generally denote perceptually indistinguishable differences between the HRTFs, especially in the horizontal plane [20].

Finally, in Table 1 we report the average and maximum SD for the two subjects, dividing the frequency axis into sub-bands. We notice that, for frequencies below 5 kHz, the average SD is very low for both the subjects. Results for subject B are slightly worse as far as the maximum value is concerned, due to isolated peaks of the SD function. Above 5 kHz the SD tends to increase. This is not unexpected, as the resolution of the 3D mesh acquired with the proposed technique is much lower than that of the ground-truth model, thus impacting on the FM-BEM simulation at high frequencies. Nevertheless, the SD still remains acceptable on the average.

Table 1: Average and maximum values of SD, for the two subjects, at different frequency ranges.

frequency range	subject A		subject B	
	avg.	max.	avg.	max.
20 Hz - 500 Hz	2.70	2.84	2.85	3.14
1 kHz - 2 kHz	2.23	2.85	2.39	3.49
2 kHz - 3 kHz	1.33	3.43	1.65	8.95
3 kHz - 4 kHz	0.85	6.69	1.59	12.09
4 kHz - 5 kHz	1.58	4.06	2.18	9.69
5 kHz - 8 kHz	4.12	13.49	3.69	12.61

5. CONCLUSIONS

We proposed a low-cost solution to recover the 3D model of the pinna, based on state-of-the-art stereo vision techniques. The accuracy of the proposed system, although being inherently lower than that achieved by high-precision scanning devices, is adequate for the prediction of the HRTF through numerical acoustic simulation. Indeed, the spectral distortion of the predicted HRTF, measured with respect to the HRTF computed from laser-scanned pinna models, turns to be acceptable in a wide frequency range, for most of the spatial directions.

Besides its cheapness, the proposed system presents several advantages over MRI or laser scanning devices. At first, it does not require any particular expertise nor medical qualification to accomplish the acquisition. Moreover, the subject is not forced to remain still during the acquisition process. Keeping these facts in mind, this work proves the feasibility of a future scenario in which customized HRTF measurement will be at everyone’s hand.

We are currently working on a deeper testing of the system, acquiring the model of more subjects and comparing the predicted HRTFs with their acoustic measurements. We are also planning formal listening tests to provide a perceptual validation of the system.

6. REFERENCES

- [1] C. P. Brown and R. O. Duda, "A structural model for binaural sound synthesis," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 5, pp. 476–488, Sep 1998.
- [2] Y. Haneda, S. Makino, Y. Kaneda, and N. Kitawaki, "Common-acoustical-pole and zero modeling of head-related transfer functions," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 2, pp. 188–196, Mar 1999.
- [3] 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, pp. 2053 – 2053, 2002.
- [4] M. Geronazzo, S. Spagnol, and F. Avanzini, "A head-related transfer function model for real-time customized 3-d sound rendering," in *Seventh International Conference on Signal-Image Technology and Internet-Based Systems (SITIS)*, Nov 2011, pp. 174–179.
- [5] C. Ahuja and R. M. Hegde, "Fast modelling of pinna spectral notches from HRTFs using linear prediction residual cepstrum," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 4458–4462.
- [6] 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, no. 3, pp. 1280–1290, 2009.
- [7] N. A. Gumerov, A. E. O'Donovan, R. Duraiswami, and D. N. Zotkin, "Computation of the head-related transfer function via the fast multipole accelerated boundary element method and its spherical harmonic representation," *The Journal of the Acoustical Society of America*, vol. 127, no. 1, pp. 370–386, 2010.
- [8] M. Joshi, N. Gupta, and L. V. Hmurcik, "Modeling of pinna related transfer functions (prtf) using the finite element method (fem)," in *COMSOL conference*, 2013.
- [9] 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, Jan 2014.
- [10] P. Guillon, *Individualisation des indices spectraux pour la synthèse binaurale: recherche et exploitation des similarités inter-individuelles pour l'adaptation ou la reconstruction de HRTF*, Ph.D. dissertation, Université du Maine, Le Mans, France, 2009.
- [11] 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, March 2013.
- [12] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge books online. Cambridge University Press, 2003.
- [13] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.
- [14] P. J. Besl and N. D. McKay, "A method for registration of 3-d shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, Feb 1992.
- [15] M. Kazhdan, M. Bolitho, and H. Hoppe, "Poisson surface reconstruction," in *Proceedings of the Fourth Eurographics Symposium on Geometry Processing*, Aire-la-Ville, Switzerland, Switzerland, 2006, SGP '06, pp. 61–70, Eurographics Association.
- [16] R. C. Gonzalez and R. E. Woods, *Digital Image Processing (3rd Edition)*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2006.
- [17] A. W. Bronkhorst, "Localization of real and virtual sound sources," *The Journal of the Acoustical Society of America*, vol. 98, no. 5, 1995.
- [18] N. Takatori, I. Naoya, T. Kazuya, and I. Fumitada, "Estimation of HRTFs on the horizontal plane using physical features," *Applied Acoustics*, vol. 68, no. 8, pp. 897 – 908, 2007.
- [19] K. J. Faller II, A. Barreto, and M. Adjouadi, "Augmented hankel total least-squares decomposition of head-related transfer functions," *The Journal of the Acoustical Society of America*, vol. 58, no. 1/2, pp. 3–21, 2010.
- [20] F. C. Tommasini, O. A. Ramos, M. X. Hüg, and F. R. Bermejo, "Usage of spectral distortion for objective evaluation of personalized HRTF in the median plane," *The International Journal of Acoustics and Vibration*, vol. 20, no. 2, pp. 81 – 89, 2015.