

APPLICATIONS OF 3D SPHERICAL TRANSFORMS TO PERSONALIZATION OF HEAD-RELATED TRANSFER FUNCTIONS

Archontis Politis^{1*}, Mark R. P. Thomas², Hannes Gamper², Ivan J. Tashev²

¹ Department of Signal Processing and Acoustics, Aalto University, 02150 Helsinki, Finland

² Microsoft Research, One Microsoft Way, Redmond, WA 98052, USA

archontis.politis@aalto.fi, {markth, hagamper, ivantash}@microsoft.com

ABSTRACT

Head-related transfer functions (HRTFs) depend on the shape of the human head and ears, motivating HRTF personalization methods that detect and exploit morphological similarities between subjects in an HRTF database and a new user. Prior work determined similarity from sets of morphological parameters. Here we propose a non-parametric morphological similarity based on a harmonic expansion of head scans. Two 3D spherical transforms are explored for this task, and an appropriate shape similarity metric is defined. A case study focusing on personalisation of interaural time differences (ITDs) is conducted by applying this similarity metric on a database of 3D head scans.

Index Terms— 3D transform, 3D shape similarity, spherical harmonic transform, HRTF personalisation

1. INTRODUCTION

Head-related transfer functions (HRTFs), which are filters modeling the acoustic transfer function from a sound source to the listener's ears, are essential for rendering immersive spatial sound over headphones. Effective spatialisation at arbitrary directions relies strongly on availability of the user's own individual set of HRTFs. However, measurement of HRTFs involves a costly and lengthy measurement procedure, making acquisition of individual HRTFs impossible for massively-deployed immersive spatial sound applications.

Prior research on indirect HRTF personalization can be categorized into three main approaches. The first approach is to parameterize the HRTF filters and let the user adjust prominent parameters through an active listening task [1, 2, 3], while the second relies on acquisition of the user's head scan followed by a wave-based numerical simulation of the filters [4, 5, 6]. The third approach, and the one that has been researched more extensively, is based on acquiring a database of measured HRTFs, associated with the respective anthropometric features of the subjects' head and/or pinna. Based on similarity of the user's features and the closest match in the

database, an HRTF set is selected that is expected to match well the user's own [7, 8, 9, 10, 11, 12, 13, 14]. However, the relation between the various anthropometrics and their effect on the HRTF is still an open research question, making the definition of effective similarity difficult. A review of the various HRTF personalisation approaches can be found in [15].

This study proposes an alternative approach to HRTF selection from a database, based on fast acquisition of the user's head scan using commodity equipment. However, instead of trying to match a few morphological parameters, it considers a non-parametric representation of the user's head shape. Using a harmonic expansion, the similarity of the user's 3D head scan to the scans of subjects in an HRTF database is determined. Following the approach of [16, 17] for finding similar 3D objects in a database, the spherical harmonic transform (SHT), well-known in acoustical processing, seems suitable for this task and has been previously used to compress and coarsely model head meshes in computer graphics [18]. However, it is essentially a 2D transform and therefore unable to model complex shapes with parts that are occluded from the origin, such as the pinna or the shoulders. The authors in [16, 17] overcome this limitation by taking a series of SHTs on a number of concentric spheres intersecting the 3D object mesh. Moreover, they define a similarity measure between 3D shapes, based on the rotationally invariant property of the SHT energy spectrum.

In this work we extend this approach using two full 3D transforms that decompose harmonically both the angular and the radial dimensions, namely the spherical Fourier-Bessel transform (SFBT) [19] and the spherical harmonic oscillator transform (SHOT) [20]. We apply the transforms on a database of head scans and we demonstrate their potential application on personalising HRTFs.

2. BACKGROUND

2.1. Spherical 3D transforms

The spherical harmonic transform (SHT) is defined on the unit sphere of square integrable functions S^2 with harmonic

*Work conducted during a research internship at Microsoft Research

coefficients given by

$$f_{lm} = \int_{\gamma \in S^2} f(\gamma) Y_{lm}^*(\gamma) d\gamma, \quad (1)$$

where $\gamma \equiv (\theta, \phi)$ is a point on S^2 , (θ, ϕ) are the inclination and azimuth angle respectively, and $\int_{\gamma} d\gamma = \int_0^{2\pi} \int_0^\pi \sin\theta d\theta d\phi$. The basis functions $Y_{lm}(\gamma)$ are complex or real orthonormalized spherical harmonics (SHs) of degree $l = 0, \dots, \infty$ and order $m = -l, \dots, l$. The function can be recovered from the coefficients by the inverse SHT

$$f(\gamma) = \sum_{l=0}^{\infty} \sum_{m=-l}^l f_{lm} Y_{lm}(\gamma), \quad (2)$$

A general spherical 3D transform form can be defined as

$$f_{nlm} = \int_{\mathbf{r} \in R^3} f(\mathbf{r}) \psi_{nlm}^*(\mathbf{r}) d^3\mathbf{r}, \quad (3)$$

where $\mathbf{r} \equiv (r, \theta, \phi)$ and $d^3\mathbf{r} = r^2 \sin\theta d\theta d\phi dr$ is the infinitesimal volume element in spherical coordinates. We are interested in basis functions that are separable in the radial and angular dimension as in

$$\psi_{nlm}(\mathbf{r}) = \psi_{nl}(r) \psi_{lm}(\gamma) \quad (4)$$

in which case the angular term are naturally the SHs $\psi_{lm}(\gamma) = Y_{lm}(\gamma)$. Due to (4), the transform of (3) can be split into a radial transform with a nested SHT

$$\begin{aligned} f_{nlm} &= \int_r \left[\int_{\gamma} f(r, \gamma) Y_{lm}^*(\gamma) d\gamma \right] \psi_{nl}^*(r) r^2 dr \\ &= \int_r f_{lm}(r) \psi_{nl}^*(r) r^2 dr. \end{aligned} \quad (5)$$

The function can be reconstructed by the inverse transform as

$$f(r, \gamma) = \sum_{n, l \in \mathbb{Z}_+} \sum_{m=-l}^l f_{nlm} \psi_{nl}(r) Y_{lm}(\gamma) \quad (6)$$

where the indexing of the double summation over the (n, l) wavenumbers depend on the type of the radial transform. For all practical applications, the order of the transform is band-limited to some maximum (N, L) depending on the order of the underlying function that is transformed, or on limitations imposed by finite sampling conditions.

Two spherical 3D transforms of the form of (5) are examined throughout this work, differing only on the radial part of the basis function and their radial domain of integration. The first is the spherical Fourier-Bessel transform (SFBT) [19], with the radial basis functions

$$\psi_{nl}(r) = N_{nl} j_l(k_{nl} r) \quad (7)$$

being the spherical Bessel functions j_l of order l , including the normalisation N_{nl} that preserves orthonormality. If the

domain of the SFBT is restricted to a solid sphere of radius α with $r \in [0, \alpha]$ and a boundary condition of $\psi_{nlm}(\alpha, \gamma) = 0$, then the normalisation N_{nl} and the scaling factor k_{nl} are

$$N_{nl} = \frac{\alpha^3}{2} j_{l+1}^2(x_{ln}) \quad (8)$$

and $k_{nl} = x_{ln}/\alpha$, where x_{ln} is the n -th positive root of $j_l(x) = 0$. Band-limiting the transform to maximum orders N, L requires all coefficients of $n = 1, \dots, N$ and $l = 0, \dots, L$.

The second transform under study is the spherical harmonic oscillator transform (SHOT), familiar in quantum mechanics as its basis functions express the wavefunctions of the 3D isotropic quantum harmonic oscillator. Recently, Pei and Liu [20] introduced it with the name SHOT as a signal processing tool for similar applications to the SFBT, such as compression and reconstruction of 3D data, shape registration and rotation estimation [21]. The radial wavefunctions of the SHOT are given by

$$\psi_{nl}(r) = N_{nl} L_n^{l+1/2}(r^2) r^l e^{-r^2/2} \quad (9)$$

where $L_n^{l+1/2}$ are the associated Laguerre polynomials with $n \in \mathbb{Z}_+$. The normalization factor can be found by enforcing orthonormality on (9) by $\int_0^\infty |\psi_{nl}(r)|^2 r^2 dr = 1$, and by the orthogonality relation of the Laguerre polynomials [22, Eq.22.2.12]

$$N_{nl} = \frac{2n!}{\Gamma(n+l+3/2)}, \quad (10)$$

where $\Gamma(\cdot)$ is the Gamma function. Even though the angular and radial orders $n, l \in \mathbb{Z}_+$ can be considered independent, we follow herein the convention used in [20] that expresses the order of the transform with a single quantum number $p = 2n + l$. A band-limited transform is then defined up to order P , with $p = 0, \dots, P$. Contrary to the SFBT defined above, the radial domain of the SHOT is $r \in [0, \infty)$.

2.2. 3D shape registration and detection

It has been shown in [16, 17] that the energy of the SHT spectrum per angular order l forms a rotationally-invariant descriptor of the transformed shape, suitable for registration and similarity matching of 3D objects [17]. That approach relies on sampling spherically a 3D object by a) voxelizing the boundary of the object, b) finding the intersecting points between these voxels and K concentric spheres expanding from the origin, and c) applying the SHT on each spherical intersection individually up to some order L . Harmonic coefficients $f_{lm}^{(k)}$ are then obtained with $k = 1, \dots, K$. A rotationally-invariant descriptor for each sphere is given by

$$e_l^{(k)} = \sqrt{\sum_{m=-l}^l |f_{lm}^{(k)}|^2}, \quad (11)$$

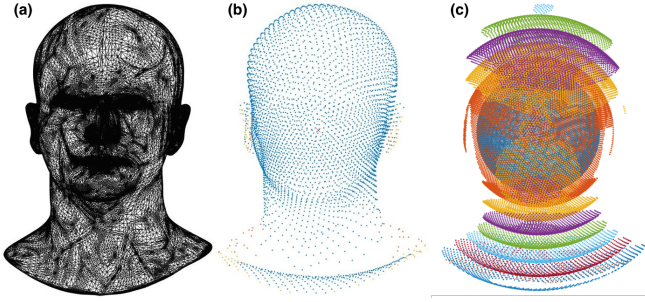


Fig. 1. Illustration of the sampling process: (a) original scanned mesh, (b) raytracing intersections, (c) coarsely sampled example, with a few sampling spheres for visibility.

resulting in an $(L + 1) \times K$ matrix \mathbf{E} that characterises the specific shape and is robust to it being rotated. A shape distance measure between two shapes (i, j) is further defined as

$$d_{ij}^{\text{SHT}} = \|\mathbf{E}_i - \mathbf{E}_j\|_2. \quad (12)$$

This approach treats each intersecting sphere separately, meaning that intersections at each segment can be rotated arbitrarily with no effect to the feature matrix \mathbf{E} . This observation motivated the authors in [19] to use the SFBT instead of separate radial SHTs, obtaining a 3D spectrum unique to the shape under study. Then a rotationally-invariant descriptor can be formulated similar to (11) for the SFBT spectrum, as

$$e_{nl} = \sqrt{\sum_{m=-l}^l |f_{nlm}|^2}, \quad (13)$$

and similarly for the SHOT spectrum, as it is shown in [20].

In this work, we construct a 3D shape similarity measure based on the SFBT/SHOT descriptor of (13), by stacking the spectral energies e_{nl} in a vector \mathbf{e} . The rotationally-invariant distance measure between two shapes (i, j) is then given by

$$d_{ij}^{\text{3DT}} = \|\mathbf{e}_i - \mathbf{e}_j\|_2. \quad (14)$$

3. APPLICATIONS TO HRTF SIMILARITY

Effective spatial rendering relies both on the magnitude and the phase response of the HRTFs, where the phase response is usually approximated with a direction-dependent delay known as the interaural time difference (ITD). While ITD depends mostly on the overall head shape, the magnitude differences rely both on the head and pinna shape. Since the harmonic descriptors obtained for each head are dominated mainly by the head shape, we restricted our preliminary evaluation only on personalization of the ITD.

The 3D transforms are applied to a database of 144 high resolution head scans captured with a Flexscan3D optical scanning setup. Each head scan is associated with its measured HRTFs, captured in the anechoic chamber of Microsoft

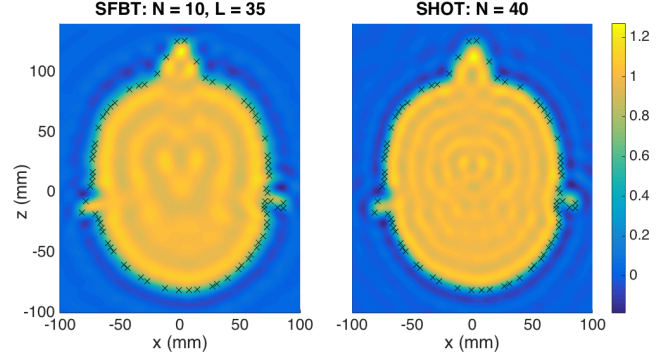


Fig. 2. Head shape reconstruction on a horizontal plane passing through the interaural axis, for a) SFBT, and b) SHOT. The dots represent the actual intersection points on the boundary of the mesh returned by the raytracer.

Research at 400 measurement directions [13]. Assuming that we can capture the user's head scan but we have no access to their HRTFs, our objective is to determine the most similar head in the database, based on the distance metric of (14), and then use the respective non-individual ITDs for the user. To validate this approach we a) apply the SFBT and SHOT transforms to all scans in the database, b) select the most similar head for all subjects, and c) determine a performance based on the similarity between the selected ITD and each subject's true measured one. We compare against two baseline methods for non-individualised HRTFs, the first based on the ITD of a Head and Torso Simulator (HATS), and the second based on the average ITD of the database.

3.1. Implementation

For the application of the SHOT and SFBT to the head scans, a similar sampling approach as the one in [16, 17] was used, but instead of voxelizing the scans, spherical sampling in a uniformly distributed set of directions was performed. 5000 uniform directions were generated as a minimum energy solution to the Thompson problem [23]. The step size for the radial sampling was fixed to 1mm, in order to capture variations on the head shape with high detail. A maximum radius of 165mm, corresponding to the furthest point of all head scans in the database, was used to limit the radial dimension. The head scan was considered as a solid object and all samples in the interior of the mesh were set to a value of one, with the rest set to zero. To assess this interior/exterior condition, a raytracer was used to find the intersections of each sampling direction with the mesh and, based on these, determine if the samples across the ray were inside or outside the head boundary. An example of the sampling process is shown in Fig. 1.

Due to the uniformity of the sampling directions, the dis-

crete SHT in (5) on each radial step r_j reduced to

$$f_{lm}(r_j) = \int_{\gamma} f(r_j, \gamma) Y_{lm}^*(\gamma) d\gamma = \frac{4\pi}{K} \sum_{k=1}^K f(r_j, \gamma_k) Y_{lm}^*(\gamma_k), \quad (15)$$

with γ_k the discrete sampling directions. To obtain the full 3D harmonic coefficients f_{nlm} , the discrete radial transform of (5) was applied to $f_{lm}(r_j)$ with the respective wavefunctions $\psi_{nl}(r_j)$ using trapezoidal integration. The order of the transform was limited to $N = 10$, $L = 35$ for the SFBT, and $P = 40$ for the SHOT. Fig. 2 presents a visual validation of the transforms, where reconstruction by the inverse SFBT and SHOT manages to represent the head shape accurately.

After the SFBT and SHOT spectra were obtained, a distance matrix between all head scans was determined by (13, 14), and for each subject its most similar head scan was selected. The ITD corresponding to this selection was deemed as the non-individual personalized ITD for that subject, returned by the method.

3.2. ITD processing

The ITD of each subject in the database was extracted from the HRTFs as detailed in [14]. To define an ITD similarity measure that considered the ITDs across all directions, we followed an approach similar to the head similarity criterion. A SHT of the ITD was taken, with a maximum order $L = 15$, limited by the specific measurement grid. Since the measurement grid was not uniform to apply directly (15), a weighted least-squares solution to the SHT was used

$$\mathbf{b}^{\text{ITD}} = (\mathbf{Y}_L^H \mathbf{W} \mathbf{Y}_L)^{-1} \mathbf{Y}_L^H \mathbf{W} \mathbf{a}^{\text{ITD}} \quad (16)$$

where \mathbf{a}^{ITD} is the vector of the ITDs at the measurement directions, \mathbf{Y}_L is the matrix of SH values at the same directions up to order L , and \mathbf{W} is a diagonal matrix of weights corresponding to the areas of the Voronoi cells around each measurement point on the sphere. Finally, after the SH spectrum of the ITDs \mathbf{b}^{ITD} was obtained, its rotationally-invariant descriptor \mathbf{e}^{ITD} was computed from (11). This step was applied in order to determine an ITD similarity that is taking the ITD shape into account but not its rotation, which could vary between subjects during measurement. Finally, the ITD distance metric between subjects (i, j) was defined as

$$d^{\text{ITD}} = \|\mathbf{e}_i^{\text{ITD}} - \mathbf{e}_j^{\text{ITD}}\|_2 \quad (17)$$

4. RESULTS

The performance of the proposed approach was evaluated in a leave-one-out cross-validation manner. The following ITD distances were assessed: a) distance between the ITD given by the head similarity and the subject's own ITD (*method*), b) distance between the generic HATS ITD and the subject's

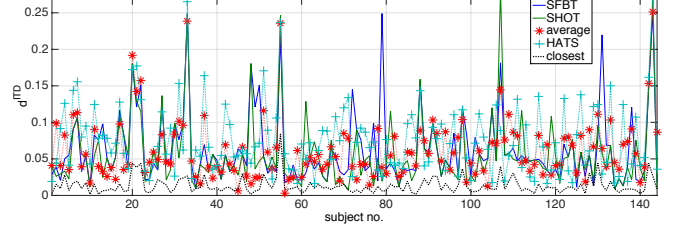


Fig. 3. ITD distances for 144 subjects in the database.

own (*HATS*), c) distance between the average ITD and the subject's own (*average*). Additionally, the smallest ITD distance was computed between each subject and all other subjects in the database (*closest*), which serves as best performance case for any method that selects a single ITD for personalization. The results are shown in Fig. 3. Average performance scores are obtained by counting the number of subjects for which *method* performs better than *average* and *HATS*. These scores are presented in Table 1, given as percentages over the total number of subjects in the database. In addition to the scores of the two 3D transforms, results using the simpler SHT-based similarity are included for comparison.

Table 1. Percentage of cases in which *method* is better predictor of ITD than *average/HATS*.

Method	Average	HATS
SHOT	64%	71%
SFBT	60%	65%
SHT	42%	53%

The scores of the SHT-based approach are significantly lower than the two 3D transforms, justifying the additional complexity of their implementation. Otherwise, the results show that the *method* performs significantly better than the *HATS* ITD for the majority of the subjects, and better than the *average* ITD, for both the SHOT and the SFBT, with SHOT giving the best results. The two transforms agreed in general on the few most similar head candidates, but could differ for certain subjects on the selection of the single most similar one, which could explain their performance difference.

5. CONCLUSIONS

This study introduces the use of two 3-dimensional spherical transforms, the SFBT and SHOT, to determine head shape similarity with applications to individualization of HRTFs. Based on the transform properties, an efficient sampling scheme of 3D head scans is developed. The resulting spectra of the head shapes are used to assess a similarity metric based on a rotationally invariant descriptor, which is applied to a case study on personalization of ITDs with positive results.

6. REFERENCES

- [1] K. Fink and L. Ray, "Tuning principal component weights to individualize HRTFs," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 389–392.
- [2] A. Härmä, R. Van Dinter, T. Svedström, M. Park, and J. Koppens, "Personalization of headphone spatialization based on the relative localization error in an auditory gaming interface," in *132nd Convention of the AES*, 2012.
- [3] K. McMullen, A. Roginska, and G. Wakefield, "Subjective selection of HRTFs based on spectral coloration and interaural time difference cues," in *133rd Convention of the AES*, 2012.
- [4] T. Huttunen, A. Vanne, S. Harder, R. R. Paulsen, S. King, L. Perry-Smith, and L. Kärkkäinen, "Rapid generation of personalized HRTFs," in *55th Int. Conf. of the AES*, 2014.
- [5] P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato, "Computer simulation of HRTFs for personalization of 3D audio," in *2nd Int. Symp. on Universal Communication (ISUC)*, 2008, pp. 435–440.
- [6] A. Meshram, R. Mehra, and D. Manocha, "Efficient HRTF computation using adaptive rectangular decomposition," in *55th Int. Conf. of the AES*, 2014.
- [7] C. Jin, P. Leong, J. Leung, A. Corderoy, and S. Carlile, "Enabling individualized virtual auditory space using morphological measurements," in *IEEE Pacific-Rim Conference on Multimedia*, 2000, pp. 235–238.
- [8] D. Zotkin, J. Hwang, R. Duraiswaini, and L. Davis, "HRTF personalization using anthropometric measurements," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2003, pp. 157–160.
- [9] H. Hu, L. Zhou, J. Zhang, H. Ma, and Z. Wu, "HRTF personalization based on multiple regression analysis," in *Int. Conf. on Computational Intelligence and Security*, 2006, vol. 2, pp. 1829–1832.
- [10] P. Guillon, T. Guignard, and R. Nicol, "HRTF customization by frequency scaling and rotation shift based on a new morphological matching method," in *125th Convention of the AES*, 2008.
- [11] X.-Y. Zeng, S.-G. Wang, and L.-P. Gao, "A hybrid algorithm for selecting HRTF based on similarity of anthropometric structures," *Journal of Sound and Vibration*, vol. 329, no. 19, pp. 4093–4106, 2010.
- [12] D. Schonstein and B. Katz, "HRTF selection for binaural synthesis from a database using morphological parameters," in *Int. Congress on Acoustics (ICA)*, 2010.
- [13] P. Bilinski, J. Ahrens, M. Thomas, I. Tashev, and J. Platt, "HRTF magnitude synthesis via sparse representation of anthropometric features," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 4468–4472.
- [14] H. Gamper, M. R. P. Thomas, and I. J. Tashev, "Anthropometric parameterisation of a spherical scatterer ITD model with arbitrary ear angles," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2015.
- [15] S. Xu, Z. Li, and G. Salvendy, "Individualization of HRTF for 3D virtual auditory display: a review," in *Virtual Reality*, pp. 397–407. Springer, 2007.
- [16] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz, "Rotation invariant spherical harmonic representation of 3D shape descriptors," in *Symposium on geometry processing*, 2003, vol. 6, pp. 156–164.
- [17] T. Funkhouser, P. Min, M. Kazhdan, J. Chen, A. Halderman, D. Dobkin, and D. Jacobs, "A search engine for 3D models," *ACM Transactions on Graphics*, vol. 22, no. 1, pp. 83–105, 2003.
- [18] S. Ertürk and T. Dennis, "Efficient representation of 3D human head models," in *10th British Machine Vision Conference (BMVC)*, 1999, pp. 1–11.
- [19] Q. Wang, O. Ronneberger, and H. Burkhardt, "Rotational invariance based on Fourier analysis in polar and spherical coordinates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 9, pp. 1715–1722, 2009.
- [20] S.-C. Pei and C.-L. Liu, "Discrete spherical harmonic oscillator transforms on the cartesian grids using transformation coefficients," *IEEE Transactions on Signal Processing*, vol. 61, no. 5, pp. 1149–1164, 2013.
- [21] S.-C. Pei and C.-L. Liu, "3D rotation estimation using discrete spherical harmonic oscillator transforms," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 774–778.
- [22] M. Abramowitz and I. Stegun, *Handbook of mathematical functions*, Courier Corporation, 1964.
- [23] J. Fliege and U. Maier, "The distribution of points on the sphere and corresponding cubature formulae," *IMA Journal of Numerical Analysis*, vol. 19, no. 2, pp. 317–334, 1999.