ITD INTERPOLATION AND PERSONALIZATION FOR BINAURAL SYNTHESIS USING SPHERICAL HARMONICS

MATTHIEU AUSSAL[1], FRANÇOIS ALOUGES[2], BRIAN FG KATZ[3]

[1]Digital Media Solutions, 45 grande allée du 12 février 1934, 77186 Noisiel, France
[2]CMAP, UMR 7641, Ecole Polytechnique, route de Saclay, 91128 Palaiseau CEDEX, France
[3]LIMSI-CNRS, BP 133, Université Paris Sud, 91403 Orsay, France

Using Head-Related Impulses Responses (HRIR) measurements, the spectral part of the Head-Related Transfer Function (HRTF) and the Interaural Time Delay (ITD) were separated. A method of synthesizing the ITD using individual morphological data and an external HRTF database was evaluated, in order to compute an individual ITD for each listener. To achieve this, spherical harmonics interpolation was performed. This method was evaluated on a set of measurements obtained for a KEMAR dummy head on the full sphere with a 5° spatial resolution, and the effect of modal truncation order was evaluated.

1. INTRODUCTION

The HRTF represents the acoustic signature for human spatial hearing, characterizing the diffraction of sound waves by the anatomy of the listener. HRTFs vary greatly from person to person, representing a major issue in reproduction quality of spatial sources over headphones. As each individual is morphologically different, these functions are particularly difficult to transpose to other individuals without audible artifacts. Current research activities have explored a variety of different approaches to solve this problem, for example, numerical simulations [6, 3], statistical matching and selection [4, 8, 7], learning optimal HRTFs by listening [5], *etc*.

In this paper, we focus on the personalization of the ITD through a principal component analysis (PCA), mainly for the following reasons:

- ITD is the primary acoustic cue for spatial sound localization [14],
- ITD is a real real-valued spatial function, simpler than the HRTF which is a complex spatial function in the frequency domain,
- PCA analysis on ITD provides interesting results, because output modes are physically interpretable. While individual values vary, the underlying cause of the ITD is common across people, that being two ears on either side of a somewhat spherical head sitting atop a torso.

The Listen database [9], with HRIR measurement on 51 subjects, was used to compute modal values of ITD by PCA. As this measurement set has a hole, with no data below -45° elevation, spherical harmonic interpolation was evaluated in order to fill this void in the measurement grid. So, the aim of this paper is first to evaluate spherical harmonic interpolation applied on ITD values, with full-sphere measurements on a KEMAR mannequin. Second, an optimal order of interpolation is determined to build a full sphere model of ITD from PCA applied to the Listen database. Finally, this model may be compared with actual ITD KEMAR measurements in future work, to assess performance.

2. SPHERICAL HARMONIC DECOMPOSITION

In this section, a theoretical summary of spherical harmonic decomposition is presented. The main purpose is to introduce classical notations and knowledge foundations for the practical use of this decomposition. Actually, spherical harmonic basis is nothing more than an equivalent for the $\mathbb{R}^3$ sphere of the $(e^{in\theta})_{n\in\mathbb{Z}}$ basis for the $\mathbb{R}^2$ circle case [12]. For this reason, the term "Fourier Bessel expansion" is sometimes used to describe this tool [10]. This basis is widely used in many domains of computer graphics and physics, including in particular that of acoustics.

2.1 Theoretical construction

Spherical harmonics of order l can be introduced as the trace on the unit sphere $\mathbf{S^2}$ of harmonic homogeneous polynomials of degree l. They have two notable properties:

1. Spherical harmonics of different orders are orthogonal for the scalar product in $\mathbf{L^2(S^2)}$, the space of square integrable functions on the unit sphere.

2. Spherical harmonics of order l form a vector space whose dimension is $\mathbf{2l + 1}$.

These two points lead to the following fundamental theorem, key to the practical use of spherical harmonics.

Theorem: Let $\forall\, l \in \mathbb{N}$, $-l \leq m \leq l$ and $-1 < x < 1$, $\mathbf{P}_{lm}$ the Legendre function associated to the Legendre polynomials $\mathbf{P}_l$ defined by

$$P_l^m(x) = (-1)^m (1-x^2)^{m/2} \frac{d^m}{dx^m} P_l(x). \quad (1)$$

In spherical coordinates $\mathbf{(r, \theta, \phi)}$ with colatitude convention, the family $\mathbf{(Y_{lm})}$ defined by

$$Y_l^m(\theta,\phi) = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}} P_l^m(\cos(\phi)) e^{im\theta} \quad (2)$$

form an orthonormal basis of $\mathbf{L^2(S^2)}$, which diagonalizes the Laplace Beltrami operator $\mathbf{\Delta}$ on the sphere $\mathbf{S^2}$:

$$\Delta Y_l^m + l(l+1) Y_l^m = 0. \quad (3)$$

Thus, any function $\mathbf{u} \in \mathbf{L^2(S^2)}$ can be written as

$$u(\theta,\phi) = \sum_{l,m} \alpha_l^m Y_l^m(\theta,\phi) \quad (4)$$

where $\boldsymbol{\alpha_{lm}}$ are the coefficients of the Fourier Bessel expansion.

This orthonormal basis of $\mathbf{L^2(S^2)}$ can now be used to represent many data types measured on a sphere. Indeed, as the ITD is a function of $\mathbf{L^2(S^2)}$, it can be projected onto this basis. In addition, a global interpolation for new data is possible, for example over the hole in the Listen database measurement grid.

2.2 Calculation of the decomposition coefficient

As explained in [2], in numerical practice we truncate the summation at a fixed value $\mathbf{L}$ called the truncation number, in such a way that we take $\boldsymbol{\alpha_{lm}} = 0$, $\forall\, l > \mathbf{L}$. There is a total of $\mathbf{M} = (\mathbf{L} + 1)^2$ terms in the spherical decomposition Eq. (4), and, for a discrete set of measurements, $\mathbf{u(\theta,\phi)}$ is known at $\mathbf{N}$ spatial position $\mathbf{\{u_1,...,u_N\}}$ on the sphere $\mathbf{S^2}$. We therefore fit the $\mathbf{M}$ unknowns $\boldsymbol{\alpha_{lm}}$ using a regularized fitting approach, by writing $\mathbf{N}$ linear equation

$$\begin{aligned} u_1 = u(\theta_1,\phi_1) &= \sum_{l,m}^{L} \alpha_l^m Y_l^m(\theta_1,\phi_1) \\ &\cdots \\ u_N = u(\theta_N,\phi_N) &= \sum_{l,m}^{L} \alpha_l^m Y_l^m(\theta_N,\phi_N). \end{aligned} \quad (5)$$

In short form, Eq. (5) is rewritten as $\mathbf{Y\alpha = U}$, where $\mathbf{Y}$ is an $\mathbf{N \times M}$ matrix of spherical harmonics in Eq. (2) at measurement points, $\boldsymbol{\alpha}$ is the unknown $\mathbf{M}$ vector of Fourier Bessel coefficients, and $\mathbf{U}$ is the $\mathbf{N}$ vector of values at measurements points.

2.3 Truncation order

The formulation in spherical harmonics, Eq. (4), has the advantage of condensing all spatial information available on the unit sphere $\mathbf{S^2}$ as a set of coefficients $\boldsymbol{\alpha_{lm}}$, depending on the order $\mathbf{L}$. In theory, the greater $\mathbf{L}$, the more spatially accurate the decomposition in spherical harmonics will be. In contrast, as $\mathbf{L}$ decreases, the decomposition in spherical harmonics will represent a spatial averaging. Mathematically, this principle is inherent to Eq. (2), where the width of the lobes depends on the order $0 < l < \mathbf{L}$.

In our approach (Eq. (5)), it is necessary to pay special attention to the order of truncation. Indeed, the resolution matrix of the system depends on the number of measurement points, defined for each database. Also, it is necessary to ensure that an increased precision given by a high truncation order $\mathbf{L}$ for reconstructed values does not induce errors, due to solving an overdetermined system, in regions where no measurements are available. Numerous studies exist on this subject, and the determination of optimal order for the decomposition is still an active research question.

2.4 Matrix conditioning and regularization

As the system in Eq. (5) is solved in the least squares sense, it is important to note that the condition number of matrix $\mathbf{Y}$ also plays an important role, being primarily dependent on the spatially distribution of the measured data. For example, most databases have a hole at low elevations (Listen, CIPIC, *etc*.), which leads in practice to an ill-conditioned matrix $\mathbf{Y}$, and requires regularization methods [13].

Tikhonov Regularization: Let us consider the linear system Eq. (5) that we rewrite under the form $\mathbf{Y\alpha} = \mathbf{U}$, where $\mathbf{Y}$ is generally not square. The classical approach to solve an over/underdetermined system of linear equations is to minimize the norm of the residual $\|\mathbf{Y\alpha} - \mathbf{U}\|^2$. In order to favor a particular solution with properties that seem relevant, a regularization term $\mathbf{\Gamma}$ is introduced in this minimization:

$$\|Y\alpha - U\|^2 + \epsilon\|\Gamma\alpha\|^2, \epsilon \in [0,1]. \quad (6)$$

The "Tikhonov matrix" $\mathbf{\Gamma}$ must be carefully chosen for the given problem. To smooth solutions, including large instabilities in areas where information is missing, regularization by Eq. (6) is often performed using a normalization of the variations. In the present case, as $\boldsymbol{\alpha}$ is the weighting vector of spherical harmonics, it was decided to emphasize the "low" spatial frequencies to avoid abrupt variations, such as mentioned in [2]. The linear system of Eq. (5) then becomes

$$(Y^T Y + \epsilon D)\alpha = Y^T U \quad (7)$$

with $\mathbf{D} = \mathbf{\Gamma}^\mathbf{T}\mathbf{\Gamma} = diag(1+\boldsymbol{l}(\boldsymbol{l}+1))$, (*diag* being the diagonal operator), and $\boldsymbol{l}$ the degree of the corresponding spherical harmonic Eq. (2).

3. CASE STUDY ON PRECISION

As the fundamental theory of interpolation by spherical harmonics is now introduced, we focus on the application of this interpolation for the KEMAR mannequin ITD, measured on a full-sphere using a particularly dense grid. In order to study the effects of truncation order, incomplete spherical grids, and interpolation, it was necessary to have a complete HRTF measurement set over the entire sphere measured with a high spatial resolution.

3.1 Reference dataset

The HRTF of a dummy head was measured in an anechoic chamber. Microphones (DPA 4060) where placed at the entrance of the ear canals (blocked meatus) of a KEMAR mannequin. Three loudspeakers, spaced at 5° intervals, were mounted on a movable arm at a distance of 1.95 m from the center of the head. The mannequin was installed on a rotating turntable and wore a fleece jumper and felt brimless cap in order to provide realistic absorption characteristics. Measurements were made for elevations of -45° to +90°, in 5° steps, with the turntable rotating in 15° increments. Using the three speakers, this resulted in a grid roughly 5°×5°. HRIRs were obtained using a sweepsine excitation signal [18] and recorded at 192 kHz sample rate (RME Fireface 800).

In order to obtain a complete spherical dataset, the mannequin was then installed in an inverted position, with the inter-aural axis at the same position, and the measurement was repeated. Following equalization for the measurement chain, the two datasets were merged. This was accomplished through a panning between the two sets at coincident measurement points. A linear panning was employed, after assuring time alignment of the two HRIRs using the peak of the cross-correlation to minimize any potential comb filtering errors. The resulting dataset comprised 1710 positions (see Fig. 1a).

The ITD was estimated from the measured HRIR by calculating the centroid value of the inter-aural cross-correlation between the envelopes of left and right HRIRs for each position (see [16]), which offers a more stable ITD estimate than the more commonly used maximum IACC method (e.g. [15]).

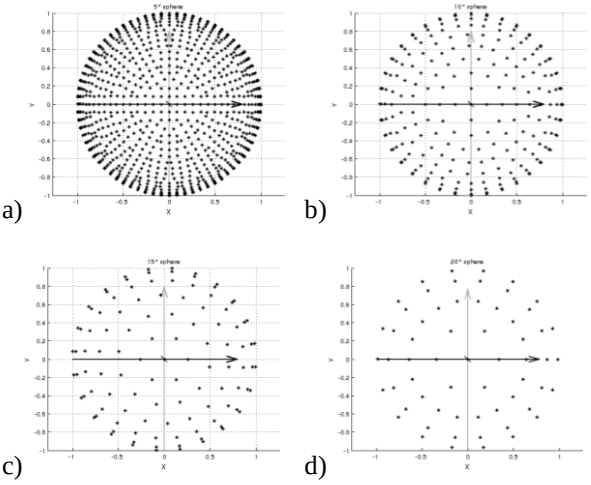

Figure 1 - Top view of KEMAR measurement grid (a) every 5°, and coarse versions at (b) 10°, (c) 15°, and (d) 20° resolution.

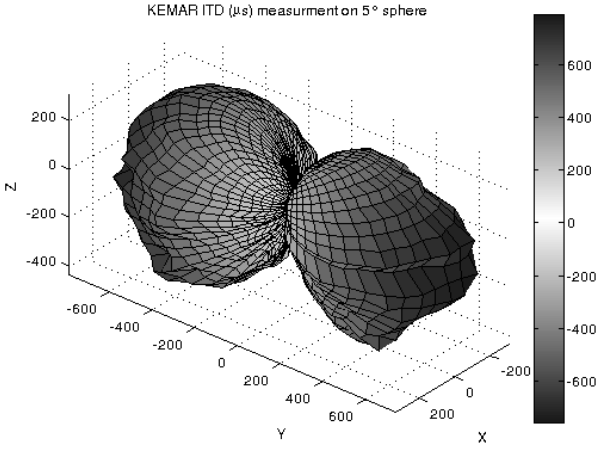

Figure 2 - Measured KEMAR ITD (µs).

3.2 Errors measurements

For the following section, in order to evaluate the spherical harmonic decomposition for ITD, some error measurements need to be properly defined.

First, as seen in Fig. 2, the maximum ITD for the KEMAR dataset is ~790 μs. The localization precision of human hearing is of the order of 2° at best. Therefore, the just-noticeable-difference (JND) for ITD can be estimated to be ~8 μs/°. Even if the JND increases with azimuth, the maximal acceptable error can be considered to be on the order of 16 μs, and we hereafter use this value as a reference.

Second, as previously discussed, it would be relevant to compare these errors to the condition number of the linear system Eq. (5). Indeed, the condition number measures the sensitivity of the solution to errors in the data, and gives an indication of the accuracy of the results. Values of conditioning near unity indicate a well-conditioned matrix, while greater values are signs of ill-conditioned systems, with poorer results.

3.3 Spatial resolution and truncation order

In order to meet the assumptions of the theoretical considerations detailed in the previous section, we first study harmonic interpolation in the case of a complete sphere, without regularization. For this, the initial mesh has been coarsened (Fig. 1b, 1c, and 1d), and two types of points are now considered.

1. The first set of points consists of Reconstructed Points (**RP**), which are the points retained as input data for the spherical harmonic decomposition in order to compute the Fourier Bessel coefficients $\boldsymbol{\alpha}_{lm}$ in Eq. (5). It is expected that the measured error of these points decreases with the truncation order **L**.

2. The second set contains the Interpolated Points (**IP**), defined as all points present in the original mesh not contained in the set **RP**, and therefore, not used for the computation of $\boldsymbol{\alpha}_{lm}$ in Eq. (5). Errors related to this set reflect the effectiveness of the spherical harmonics for interpolation.

Measures of error **e** were calculated for both sets of points, **RP** and **IP** (Fig. 3).

As expected, **e(RP)** converge with spherical harmonic truncation order **L**. The coarser the mesh, the smaller the size of set **RP**, which results in a quicker converging process, and is consistent with the theory. However, for coarsened meshes, **e(IP)** is unstable for some truncation orders **L**. Indeed, the problem becomes ill-conditioned when too many spherical harmonics are used in relation to the number and density of points on the measurement grid.

To overcome this problem, the Tikhonov regularization in Eq. (7) has been performed, and results are shown in Fig. 4. The effectiveness of this regularization method is seen due to the fact that **e(RP)** and **e(IP)** are convergent, even if $\mathbf{e}_{max}$**(IP)** can be further improved. The choice of the matrix $\boldsymbol{\Gamma}$ and the weight $\boldsymbol{\epsilon}$ in Eq. (7) should be studied more thoroughly in order to improve this convergence. However, data used as references are physical measures, and suffer from inaccuracy. Also, the inherent smoothing of the interpolation by spherical harmonics may produce "errors", which can ultimately prove beneficial. Finally, it is interesting to notice that **e(RP)** does not tend towards zero when the Tikhonov regularization is used. Indeed, as this method performs a smoothing of the results, reconstructed values **RP** are included. But these errors are well below the proposed limit of 16 μs, which remains acceptable.

3.4 Effect of a data hole

Having addressed the case of the complete sphere, we are now interested in the accuracy of spherical harmonic decomposition when the dataset has a hole. Because of the complexity of measurements of HRIRs at low elevations, most databases do not have measurements in this area. The linear system in Eq. (5) is therefore ill-conditioned, which is a real problem.

The results for two mesh resolutions, the original one and a mesh coarsened to 15° resolution, are presented here (see Fig. 5 and 6). In both cases, the Tikhonov regularization in Eq. (7) has been applied. **e(RP)** and **e(IP)** are now not only functions of truncation order **L**, but also of the hole size (varied here as a function of elevation below which data are excluded from set **RP**. In the first case (Fig. 5), **e(IP)** applies only to points in the hole to focus attention solely on the presence of this hole in the harmonic interpolation. The second case (Fig. 6) is more general since it allows one to obtain an optimal truncation order **L** empirically, in the case of the Listen database (measurements every 15°, hole at $\boldsymbol{\phi}$ = -45°).

Conditioning: Both cases (Fig. 5 and 6) show that despite regularization, the conditioning of linear system Eq. (7) deteriorates with increased hole size and also with increasing order. This is not surprising since we are attempting to solve a problem more precisely with **L** increasing, though using less and less information.

Reconstruction error: $\mathbf{e_{max}(RP)}$ clearly converges in both cases (see Fig. 5 and 6), even if the truncation order $\mathbf{L} = 30$ is insufficient to see $\mathbf{e_{max}(RP)} < 16$ μs on the original mesh. Moreover, it converges more quickly when the number of points used for reconstruction decreases, as predicted by theory.

Interpolation error on 5 Mesh: In Fig. 5, **e(IP)** generally follows the conditioning of the matrix, with some exceptions. We find that the larger the hole, the poorer the interpolated values in the hole. Indeed, the interpolation over the hole does not correspond to reality, but to the model that we have imposed by the Tikhonov regularization in Eq. (7).

Interpolation error on 15 Mesh: In Fig. 6, **e(IP)** is the combination of errors generated by the hole, as well as errors generated by the data surrounded by points of reconstruction. It appears that the error generated by the hole is still present, but it is compensated by the accuracy of the interpolation in the area where measurements exist. So, even if $\mathbf{e_{max}(IP)}$ does not converge, $\mathbf{e_{mean}(IP)}$ is rather good at $\mathbf{L} \sim 18$. We therefore use this value in the following section.

4. ITD INDIVIDUALIZATION VIA PCA

In considering the ITD as a simple single valued function at any given point in space (ITD as a function of frequency is not considered in this initial study) it can be imagined that it would be possible to predict the ITD for an individual given sufficient morphological information. This phase of the study concerns the prediction of approximate ITD values using a limited number of morphological parameters and statistical analysis of an HRTF/morphological database.

4.1 PCA

Principal Component Analysis (PCA) is a powerful statistical tool that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of uncorrelated variables called principal components, hierarchically ordered. This procedure can be used to reduce the complexity of datasets while finding commonalities across observations.

Using the publicly available Listen HRTF dataset [9], the ITD of each individual in the dataset was calculated. This dataset (at the time of this study) contained HRIRs for 51 individuals, sampled at 187 positions from -45° to +90° in elevation, with sampling spacing of approximately 15° in azimuth and elevation. The ITD was estimated using the same method described in Sec. 3.

Applying the PCA analysis to this set of ITD results in the generation of a series of spatial basis functions and associated scalar weightings for each individual, which when recombined allow for the reconstruction of the original dataset. The first three PCA components, or spatial functions, are shown in Fig. 7. These functions have a value range of ~±2. Their contributions to the reconstructed ITD can be examined by values of the associated weighting functions. The corresponding three individual weighting functions have respective means(standard deviations) of 357.5(16.7), -0.13 (22.8), and 0.13(18.1) for an ITD in μs. Visual inspection of these results highlights that the first component contains the primary and fundamental information associated with the ITD, clearly defining variations with respect to the median plane, with extreme values being near the extreme azimuthal positions. The second and third components appear to offer refinements in the ITD spatial function along either the generalized front/back or up/down directions. The mean error in the reconstructed ITD over all subjects and positions using only these three PCA components was -0.75±25.52 μs, with the mean unsigned error being 16.47±19.51 μs.

Overall, the reconstructed ITD results via PCA are within the perceptual error limits described previously, though there is some degree of variation that extends beyond these limits. Inspection of the estimated ITD from measurements highlights some variations which are possibly due to individual subject errors. The goal of the PCA method is to provide a *generalized* solution to an ITD prediction method, and therefore some error may exist for a given individual. If, for example, there are any anomalies in the measurement results for an individual, at some given position, the PCA reconstruction will act as a form of smoothing, removing this anomaly (unless it is present in all subjects). As such, the calculated error could be high, while the actual perceived error for the subject could be lower, if not even improved with respect to the measurements.

4.2 Morphological prediction

In addition to the measured HRIR data, the Listen database [9] also included morphological data measured on the various subjects following the parameters defined in the CIPIC database [17]. Using the available morphological parameters, the individual weights are predicted using a directed linear regression analysis technique. Of the 51 subjects in the database, 38 subjects had sufficient associated head and torso morphological data (parameters relative to the pinnae morphology were not included).

The procedure consisted of calculating the best linear regression fit using all available morphological

data (using the MatLab function *robustfit*). From these first results, those parameters whose contribution was judged to be relevant to the prediction (p-value < 0.15) were retained and the robust regression was recalculated with only these select parameters. This procedure was performed for the three sets of individual weights. Results showed that the morphological parameters that were most correlated with the different weights, and therefore identified as contributing to the respective spatial functions were:

- PCA 1: *Head_Circumference*_x16 and *Shoulder_Circumference_x17*,
- PCA 2: *Head_Depth_x3* and *Torso_Top_Width_x9*,
- PCA 3: *Head_Depth_x3*, *Pinna_Offset_Back_x5*, *Neck_Width_x6*, and *Neck_Height_x7*.

Correlations of the resulting prediction models to the computed PCA weighting function resulted in values of 0.86, 0.34, and 0.50 for the three functions respectively. The prediction model for the first and primary principal component performs well, requiring only two simple parameters which seem relevant to the acoustics inherent in the ITD. The second component, associated with front/back variations, presents the poorest performance. The third component performs moderately, with several additional parameters. These results, combined with the presence of somewhat unexpected morphological parameters such as neck height, may be indicative of the difficulty in obtaining reliable values for the more "obvious" parameters. We note that while the CIPIC protocol presents numerous measures, the precise definitions are not apparent when applied to real-world individuals. As such, parameters such as head depth, or ear offset, may exhibit more variability due to experimenter protocol which may render the results less pertinent. At the same time, simpler to measure parameters, such as those relating to the neck and torso, while not having obvious acoustical impact may be anatomically correlated to the more difficult to measure parameters.

The mean error across all participating subjects was -0.63±37.56 μs, with the mean unsigned error being 28.04±25.00 μs, slightly higher but still quite comparable to the results for the PCA reconstruction. Mean results over subjects of a correlation analysis between the overall estimated ITD and both the PCA decomposition and the PCA reconstruction using the morphological predictions for the individual weighting functions are equal to 0.99.

4.3 Spherical harmonics applied to PCA ITD

The goal of the application of a spherical harmonic decomposition to the proposed ITD prediction method is to improve the interpolation of ITD values far from those points in the measured dataset. The main advantage of this method over linear interpolation is that it is a global interpolation, not local. Moreover, it is particularly suited to the spherical distribution of data, which is generally the case for measurements of HRIR. Finally, the Tikhonov regularization provides a smoothing of interpolated values, controllable by the matrix regularization. This can impose some desired properties on the solution sought. Using the results from Sec. 3, spatial decomposition of the three PCA spatial functions has been performed with the order limits determined according to the spatial resolution and spatial distribution corresponding to the Listen measurement dataset (roughly 15° spacing with no data below -45° elevation). This corresponds to a model order of $\mathbf{L} = 18$, as determined in the previous section.

The ITD predication process for the simple PCA recomposition (Sec. 4.1) and the recomposition using PCA and the morphological model for the PCA weighting functions (Sec. 4.2) was repeated using the reconstructed three PCA basis functions using spherical harmonics. The mean error relative to the measured ITD across all subjects and positions for the PCA reconstruction using the spherical harmonic decomposition was found to be -0.78 ± 25.81 μs, with the unsigned mean error being 16.59 ± 19.78 μs, almost identical to the results for the pure PCA reconstruction. The mean error for the ITD reconstructed using the morphological method for individual weighting functions was -0.66±37.74 μs, with the mean unsigned error being 28.23±25.06 μs. These results show little to no degradation due to the spherical harmonic decomposition of the PCA basis functions.

A more detailed analysis of the error is provided as a function of azimuth angle in Fig. 8. Results show good performance for the proposed method for azimuths near the median plane, while errors for azimuths approaching the interaural axis exhibited the largest errors, with the maximum errors a lateral angle of 105°. This region offers a variety of difficulties in ITD analysis, which are beyond the scope of this paper.

The results here suggest that for the majority of potential source positions, using the spherical harmonic decomposition of the PCA model combined with the morphological model for the individual weighting functions provides an architecture for the prediction of individual ITD values for any desired source position. The errors are typically below 20 μs, near the

approximate JND limit of 16 μs. While some variations beyond this limit are observed, it remains to be verified if these errors are actually perceptible.

5. CONCLUSION

The decomposition of ITD using PCA resulted in the first three components offering spatial variations which present variations relevant to physical interpretations. Correlations of weighting functions for PCA components with individual morphological data showed high correlations with physically relevant parameters. While the determination of *high* relevance of some parameters was not always obvious, variability and difficulty in the measurement of some of the finer parameters may explain their absence in the statistical results, and biological correlations between indirect measurements may offer justification for these final results.

Otherwise, results for the morphological prediction of PCA weighting functions are encouraging. While not perfect, the proposed method offers a new approach for addressing the need for individualized HRTF data (currently only addressing the ITD) in the case when actual measured data for the individual is not available.

As a final evaluation of the proposed methodology, the ITD estimate based on the morphology of the KEMAR mannequin has to be compared to the corresponding measured ITD. As this measured ITD employed a finer resolution spatial grid over the entire sphere, the accuracy of the predicted ITD should be compared both at intermediate positions relative to the Listen measurement grid for which the PCA basis functions were calculated, as well as at lower elevations where the Listen dataset had no data. Our results for spherical harmonic interpolation are a significant step towards achieving this in the near future.

Finally, additional studies should be carried out concerning the perceptual validation of the method with regards to prediction errors, as well as interpolation errors.

ACKNOWLEDGMENTS

This work was performed in the context of a CIFRE industry-linked research funding program with Digital Media Solutions (DMS). The authors would like to thank both Gaëtan Parseihian, LIMSI-CNRS, for his assistance, and IRCAM for use of their measurement facilities, regarding the acquisition of the KEMAR HRTF. Finally, thanks to Omnihead (http://www.univ-brest.fr/mstis/omnihead/accueil.htm) for the use of their KEMAR dummy head.

REFERENCES

[1] M. Bruneau. *Introduction aux Théories de l'Acoustique.* Université du Maine, Le Mans, France - 1983.

[2] R. Duraiswami, D. Zotkin, and N. Gumerov. *Interpolation and range extrapolation of HRTFs.* In IEEE ICASSP'04, pages IV45–IV48, Montreal, Canada - 2004.

[3] N. A. Gumerov, A. 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 representation via the spherical harmonic spectrum*. Technical Report CS-TR-4936 (also UMIACSTR-2009-06), Department of Computer Science, University of Maryland, College Park - 2009.

[4] V. Larcher. *Techniques de spatialisation des sons pour la réalité virtuelle*. PhD thesis - 2001.

[5] B.F.G Katz and G. Parseihian. *Perceptually based head-related transfer function database optimization*. Journal of the Acoustical Society of America, vol. 131(2), pp. EL99-EL105 - 2012.

[6] B.F.G. Katz. *Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation*. Journal of the Acoustical Society of America, vol. 110 (5), pp. 2440-2448 - 2001.

[7] D. Schönstein and B.F.G. Katz. *HRTF selection for binaural synthesis from a database using morphological parameters*. Proc. 20th International Congress on Acoustics (ICA), Sydney - 2010.

[8] Y. Iwaya. *Individualization of head-related transfer functions with tournament-style listening test: Listening with other's ears*. Acoustical Science and Technology, 27(6):340–343 - 2006.

[9] IRCAM Listen HRTF database. http://recherche.ircam.fr/equipes/salles/listen/ (date last viewed 10/28/11).

[10] A. Laborie, R. Bruno, and S. Montoya. *A New Comprehensive Approach of Surround Sound Recording*. AES 114th Convention, Amsterdam - 2003.

[11] K.-V. Nguyen, T. Carpentier, M. Noisternig, and O. Warusfel. *Calculation of head related transfer functions in the proximity region using spherical harmonics decomposition : Comparaison with the measurements and evaluation*. Proc of the 2nd Int. Sym. on Ambisonics and Spherical Acoustics - 2010.

[12] J.C Nédélec. *Acoustics and Electromagnetic equations. Integral representations for harmonic problems*. Applied Mathematics Science, 144, Springer Verlag, New-York - 2001.

[13] D. N. Zotkin, R. Duraiswami, and N. A. Gumerov. *Regularized HRTF fitting using spherical harmonics*. Proceedings of the 2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pp. 257–260 - 2009.

[14] J. Blauert. *Spatial Hearing*. MIT Press, Cambridge - 1996.

[15] D.J. Kistler and F.L. Wightman. *A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction*. Journal of the Acoustical Society of America, Vol 91(3), pp. 1637–1647 - 1992.

[16] P. Minnaar, J. Plogsties, S.K. Olesen, F. Christensen and H. Møller. *The Interaural Time Difference in Binaural Synthesis*. In Proceedings 108th Convention of the Audio Engineering Society, Paris, France, preprint 5133 - 2000.

[17] V.R. Algazi, R.O. Duda, D.M. Thompson and C. Avendano. *The CIPIC HRTF database*. Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 99–102 (New Platz, NY) - 2001.

[18] P. Fausti and A. Farina. *Acoustic measurements in opera houses: comparison between different techniques and equipment.* J. Sound Vib. 232, 213–229 - 2000.

FIGURES

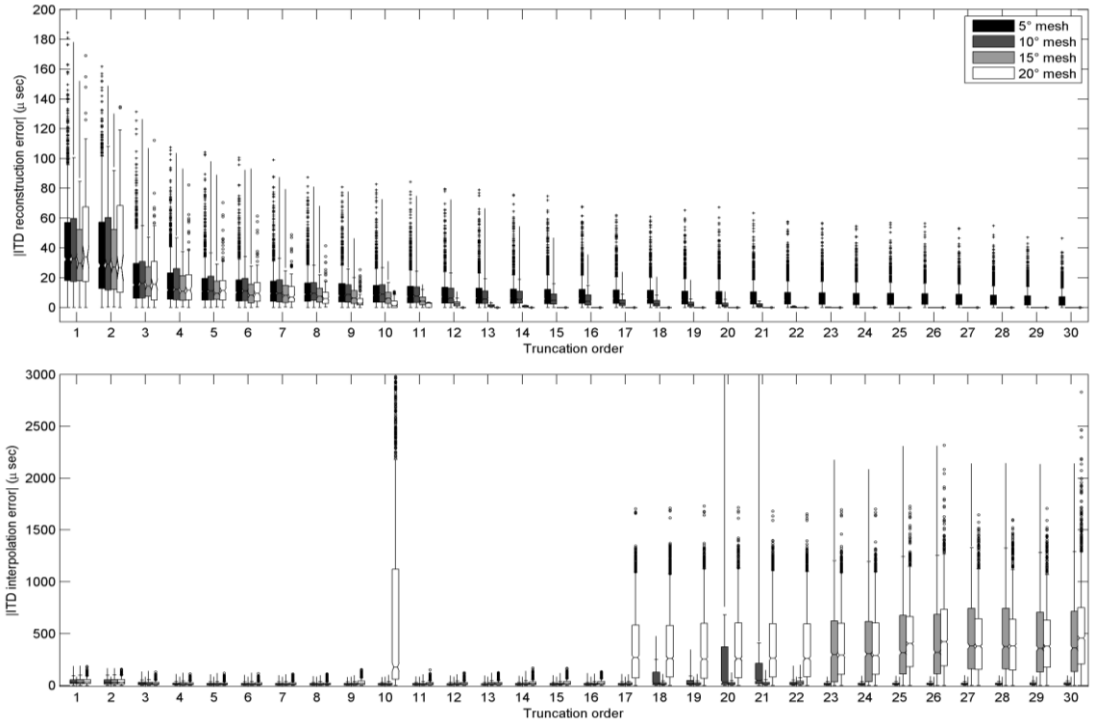

Figure 3 - Absolute error of reconstruction e(RP) (top) and interpolation e(IP) (bottom) on the full sphere, without Tikhonov regularization.

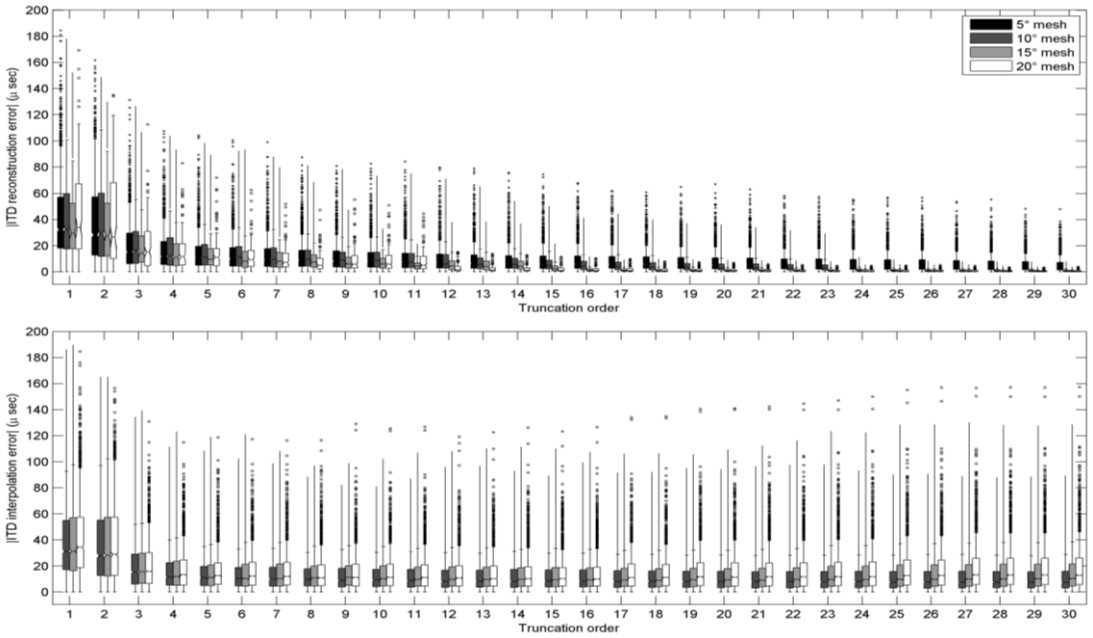

Figure 4 - Absolute error of reconstruction e(RP) (top) and interpolation e(IP) (bottom) on the full sphere, after Tikhonov regularization.

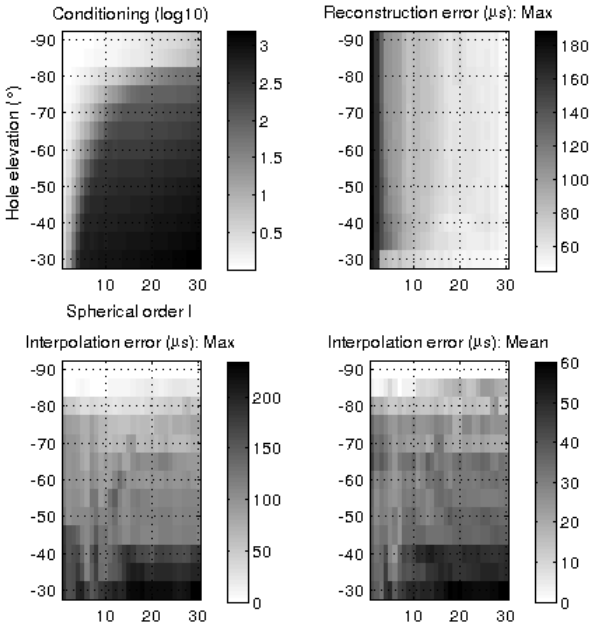

Figure 5 - Errors due to a data hole on the original mesh (5°), as a function of truncation order L and hole elevation.

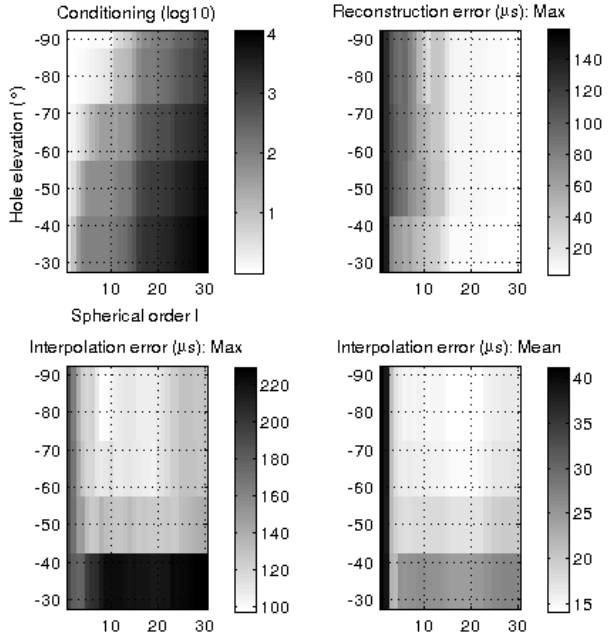

Figure 6 - Errors due to a data hole on a coarse mesh (15°), as a function of truncation order L and hole elevation.

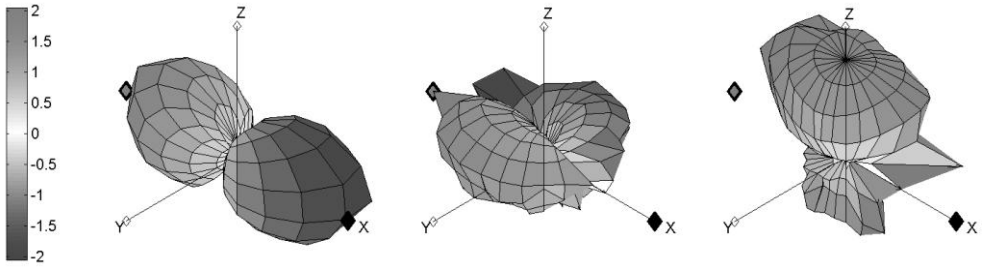

Figure 7 - First three PCA components for ITD reconstruction, from left to right.

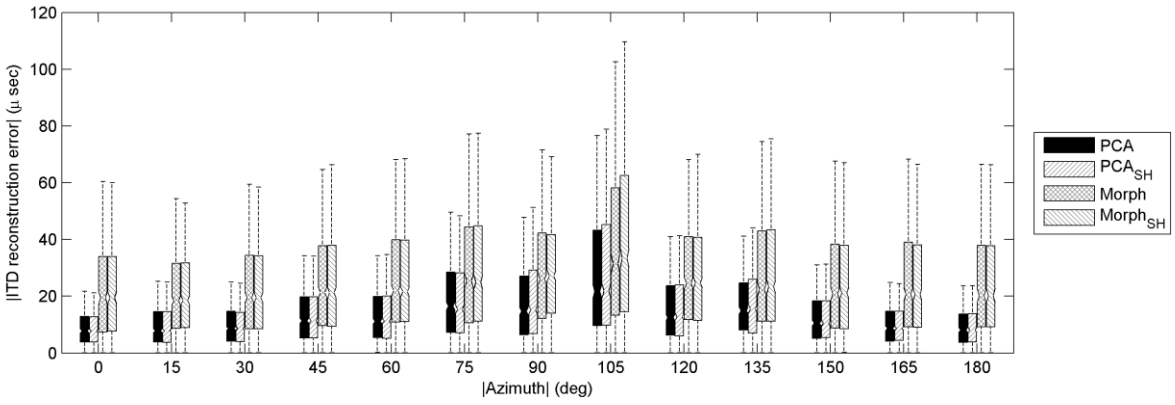

Figure 8 - Boxplots of unsigned errors in ITD reconstruction relative to measured data over all subjects vs. azimuth angle magnitude: direct reconstructions (*PCA*), and reconstructions using morphologically predicted individual weighting function (*Morph*) using the original PCA basis function or spherical harmonic (*SH*) decompositions of the PCA basis functions.