REGULARIZED HRTF FITTING USING SPHERICAL HARMONICS

Dmitry N. Zotkin, Ramani Duraiswami, and Nail A. Gumerov

Institute for Advanced Computer Stidues (UMIACS)
University of Maryland, College Park
College Park, MD 20742 USA

ABSTRACT

By the Helmholtz reciprocity principle, the head-related transfer function (HRTF) is equivalent to an acoustic field created by a transmitter placed at the ear location. Therefore, it can be represented as a *spherical harmonics spectrum* – a weighted sum of spherical harmonics. Such representations are useful in theoretical and computational analysis. Many different (often severely undersampled) grids are used for HRTF measurement, making the spectral reconstruction difficult. In this paper, two methods of obtaining the spectrum are presented and analyzed both on synthetic (ground-truth data available) and real HRTF measurements.

Index Terms— Acoustic fields, spatial audio, circular arrays, array signal processing, head-related transfer function, HRTF.

1. INTRODUCTION

Sound localization cues used by humans are produced by scattering of the acoustic wave off the listener's anatomical parts (in particular, outer ear (pinna), head, and torso) [1]. The head related transfer function (HRTF) quantifies how the signal is modified by this scattering as a function of arrival direction and of source range. It is usually assumed that the range effects are small at typical measurement distances. If the HRTF of a particular person is known, the illusion of a sound arriving from a specific direction can be created [2]. However, HRTFs are different for different individuals and consequently HRTF measurement is necessary for accurate spatial audio reproduction. During such a measurement, a sound source (e.g., a loudspeaker) is moved to many positions on a certain grid around the subject, and at each position a test signal is produced by the source and is recorded by a microphone placed in the subject's ear, directly giving the desired transfer function. It can be shown that the infinite-range HRTF for direction (θ, φ) is just the acoustic field potential at the ear when the incoming acoustic field is a plane wave arriving from (θ, φ) [3].

Recently, an alternative way of measuring HRTF was proposed and tested in [4]. Assume that in a linear time-invariant scene the source and the receiver positions are swapped. By the Helmholtz reciprocity principle, the recordings done before and after the swap must be identical. It follows that the HRTF can be measured reciprocally by placing the (miniature) loudspeaker in the person's ear and recording the test signal in parallel by a number of microphones arranged around the subject. That means that HRTF is equivalent to the potential of an acoustic field created by the loudspeaker and sampled by microphones (many authors use this fact when computing HRTF using numerical methods on a subject's mesh, e.g. [3]). As any acoustic field, this field must comply with the Helmholtz (wave) equation, and the natural basis for representing it is the set of the elementary solutions of that

equation. These elementary solutions are a product of spherical harmonics in (θ, φ) and of spherical Hankel functions in range, which become constant if the range is fixed. The remaining spherical harmonics basis is well-suited for HRTF representation; an advantage of such representation is that the interpolation results are guaranteed to be physically valid, which can not be said about other ad hoc (e.g., spline-based) HRTF interpolation methods.

Several authors worked before on HRTF representation with spherical harmonics. In [5], an 18-order representation is used for manikin HRTF measured on a 648-point grid that is particularly well-suited for the spherical harmonics framework. It is shown that HRTF is represented well and that the interpolated data is reasonably consistent with actual HRTF measured at interpolated positions. However, the specific grid used (Gaussian grid) is inconvenient in practice due to high node density near the poles; further, it includes positions with very low elevations, where the measurements can not be done for human subjects. Also, a large number of measured HRTFs exist, and these measurements are all on different grids. In [6], the above-mentioned problem of not having the data at low elevations is addressed, and an iterative algorithm is proposed to obtain the decomposition while simultaneously filling in the missing data. Simulation results are presented for low-accuracy (order 4) data. It is shown that the ground-truth is recovered successfully despite having no data on a quarter of the grid. No experimental results are presented, and order 4 representation appears inadequate for dealing with realistic fine-detail HRTF.

In the current work, two methods of computing spherical harmonics representation are described and tested. The first is the traditional method of direct integration (DI) over the sphere (used e.g. in [5]). The second is the least-squares (LS) fitting method introduced in [7]. It is shown in simulations that DI method requires specific quadrature and fails if arbitrary grid is used whereas the LS method works well with arbitrary grid; however, the fitting matrix appears to be ill-conditioned if a grid with cut-off bottom is used. Regularization is introduced to control the magnitude of spherical harmonics expansion coefficients in such a case, and experimental results are presented.

2. SPHERICAL HARMONICS DECOMPOSITION

The orthonormal spherical harmonics $Y_n^m(\theta, \varphi)$ are defined as [8]

$$Y_n^m(\theta, \varphi) = (-1)^m \sqrt{\frac{2n+1}{4\pi} \frac{(n-|m|)!}{(n+|m|)!}} P_n^{|m|}(\cos\theta) e^{im\varphi},$$

(1)

where $n = 0, 1, 2, ...; m = -n, ..., n$; and $P_n^m(\mu)$ are the associated Legendre functions. Assume that the HRTF measurement grid consists of L_q directions s'_q and that the potential measured at s'_q is $\Psi(k; s'_q)$, where $k = 2\pi f/c$ is the wavenumber, f is the frequency, and c is the sound speed. The decomposition is done

separately at each frequency bin; the value of $\Psi(k; s'_q)$ is just the complex Fourier coefficient at the corresponding frequency f. The argument k is omitted from now on for clarity. Remember that the HRTF measurements are just the samples of a valid acoustic field $\Psi(s)$, where s is the continuous vector of directions. The desired representation of $\Psi(s)$ is

$$\Psi(s) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} C_n^m Y_n^m(s), \qquad (2)$$

where C_n^m are the complex expansion coefficients. The set of C_n^m is also called *spherical harmonic spectrum*. In practice, the outer summation is truncated:

$$\Psi(s) = \sum_{n=0}^{p-1} \sum_{m=-n}^{n} C_n^m Y_n^m(s). \qquad (3)$$

The parameter p is called the truncation number; the total number of C_n^m coefficients is p^2. The truncation error obviously depends on p. Note that the higher is k, the faster can the field vary spatially and the higher accuracy is needed. Many authors use $p = ka$ as the value reasonably balancing the representation size and the truncation error, where a is the scatterer radius.

Direct Integration (DI) Method: Similarly to 1-D Fourier transform, the spectrum can be computed directly from the data:

$$C_n^m = \int_{S_u} \Psi(s) Y_n^{m*}(s) dS(s), \qquad (4)$$

where the star denotes complex conjugation and the integration is done over the surface S_u of unit sphere. In the discrete case, summation is done over quadrature points s'_q with weights w'_q:

$$C_n^m = \sum_{q=1}^{L_q} w'_q \Psi(s'_q) Y_n^{m*}(s'_q). \qquad (5)$$

Selection of an optimal quadrature for the sphere is a subject of considerable debate. It must not only approximate the integral well but should also be feasible for practical implementation. Any exact quadrature of order p must have at least p^2 points [9], and a *product Gaussian quadrature* uses $2p^2$ points and is easy to construct mathematically [5]; however, it is inconvenient to use in practical applications. An exact quadrature with points spread more or less evenly requires $4p^2$ points. Recently a method for generating *approximate* quadrature for the sphere was proposed by Fliege and Mayer in [10]. The obtained point distribution is uniform, and despite having only p^2 points the Fliege quadrature provides very low approximation error for many integrands [10] including spherical harmonics [11] and plane waves [12]. Still, the general problem with the DI method is that the measurements must be performed on a pre-defined (quadrature) grid, and all the existing HRTF measurements can not be processed through this framework.

Least-Squares Fitting (LS) Method: Another way to proceed is to note that a linear system of equations is formed if equation (3) is repeated L_q times for all the measurement directions. In the matrix form, the system becomes $YC = \Psi$, where Ψ is $L_q \times 1$ vector of $\Psi(s'_q)$, Y is $L_q \times p^2$ matrix of values $Y_n^m(s'_q)$, and C is a $p^2 \times 1$ vector of C_n^m. The spectrum values are stored by order n and then by degree m, so the vector $C = [C_0^0 \ C_1^{-1} \ C_1^0 \ C_1^1 \ C_2^{-2} \ ...]$. Similarly, the qth row of Y looks like $[Y_0^0(s'_q) \ Y_1^{-1}(s'_q) \ Y_1^0(s'_q) \ Y_1^1(s'_q) \ Y_2^{-2}(s'_q) \ ...]$. This linear system can be solved for C using least-squares fitting (assuming $p^2 \leq L_q$).

Spatial Sampling Issues: To recover the spatial structure of the acoustic field up to order p, the number of measurement directions must be at least p^2 (applying spatial Nyquist criteria and assuming even distribution of points over the sphere). Indeed, the number of coefficients in the expansion is p^2, so with less than p^2 measurements there is not enough information available. For a fixed measurement grid, this places a limit on accurate representation of HRTF at higher frequencies. For reference, at 20 kHz $k = 366.2 \ m^{-1}$ and $ka = 36$ assuming head radius of $0.1 \ m$. At $p = 36$ the spectrum has 1296 coefficients, which is approximately equal to the number of measurement directions in state-of-the-art HRTF measurement systems (e.g., in CIPIC database). If a coarser grid is used, fine details of HRTF will not be captured. It is not clear whether these fine details matter for localization accuracy and for scene perception in general; it wouldn't be possible to answer this question without careful psychoacoustic study, which is beyond the scope of the current work.

Grids Used: Define the following concepts and abbreviations:

- Open grid: A grid with the bottom part cut off (i.e., measurements are not available for low elevations);
- Closed grid: A grid fully covering the sphere of directions;
- FC: Closed Fliege grid (100 points);
- LO: Low-resolution open grid used in [4] for HRTF measurement (397 points);
- LC: Low-resolution closed grid constructed from LO grid by reflecting the points as necessary (452 points);
- GC: Closed Gaussian quadrature grid used in [5] for HRTF measurement (648 points);
- HO: High-resolution open grid used in [13] for HRTF measurement (823 points);
- HC: High-resolution closed grid constructed from HO grid by reflecting the points as necessary (1074 points).

The HC grid is actually a regular 5-degree double polar grid. The LC grid is irregular; the points are spread as even as possible subject to the constraint that they must be located in the nodes of a mesh-like structure surrounding the subject.

3. LS METHOD ANALYSIS

In preliminary studies, it was noted that the LS method fails in case of open grid; more specifically, both the spectrum magnitude and $\Psi(s)$ computed become very large for the directions in the cut-off grid area. In order to fix this problem, regularization analysis was performed using Regularization Toolbox 4.1 [14], which has a number of tools for analyzing ill-posed problems, applying various regularization techniques, and automatically selecting the regularization parameters. It was found that the matrix Y is indeed ill-conditioned for the open grid. For example, for $k = 189.0 \ m^{-1}$ ($ka = 19$) the condition number for Y computed over LC grid is 8.56, while the same over LO grid is $1.66 \cdot 10^5$.

Further insights into the characteristics of ill-posed problem are given by so-called Picard plots. The discrete Picard condition that must be satisfied for the problem $YC = \Psi$ to be well-defined is that the Fourier coefficients of the right-hand side $|u_j^T \Psi|$ must decay to zero faster than the singular values σ_j of Y, where u is the left singular vector of Y. The Picard plot shows the values of $|u_j^T \Psi|$ and σ_j and the ratio between them as j increases. It is often possible to isolate a group of very small singular values by looking at the Picard plot and to regularize the problem by removing (truncated SVD regularization, TSVD) or suppressing (damped SVD regularization, DSVD) the contribution to the solution corresponding to those small values. Figure 1 shows the Picard plots for the
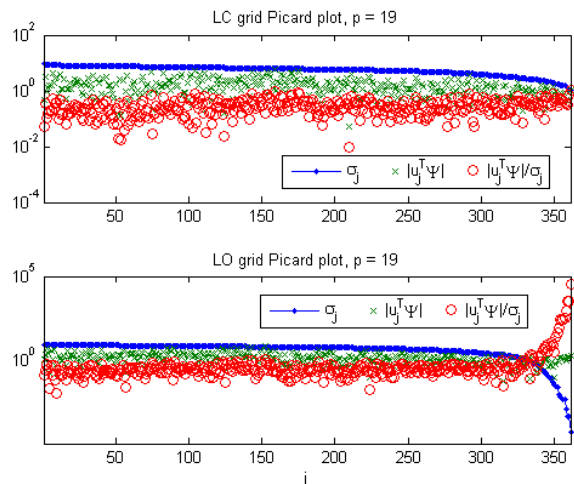
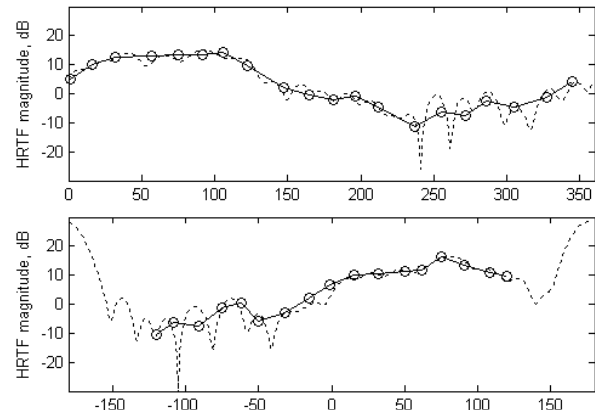Figure 1: Sample Picard plots for closed (top) and open (bottom) grids.

Figure 2: KU-100 manikin HRTF variation along equator (top) and along a circle in frontal plane (bottom). The value on X axis is the angle around the circle. Solid line: measured HRTF (circles denote measurement points); dashed line: evaluation of spherical harmonics representation of the HRTF obtained by the LS method.

LC grid (top) and for the LO grid (bottom). In the top plot, the discrete Picard condition holds everywhere, whereas in the bottom plot it is violated for about last 10% of singular values whose magnitude drops sharply, which suggests to set the TSVD regularization parameter to about 0.9 (i.e., one should keep about 90% of the components and drop the rest).

4. SIMULATED RESULTS

The algorithms were tested against the ground-truth data to objectively evaluate their performance. Fast boundary elements method (BEM) was used on the mesh of Georg Neumann KU-100 dummy head to numerically compute the HRTF as the acoustic field generated by the transmitter placed at the ear location [3]. The BEM software output the HRTF spectrum of order $p_s = ka + 4(ka)^{1/3} + 3$, which was used as the ground-truth data. The BEM software also computed the infinite-range HRTF data for all grids used. The DI and the LS methods were then used to derive the spherical harmonic spectrum of the HRTF data, and the obtained spectrum was compared against the ground-truth. Implementation was done both in C and in Matlab; the least-squares fitting was done by "gaussj" routine from Numerical Recipes in C and by simple backslash division in Matlab. It was found that the results produced by these two approaches were identical. Computations were performed for $k = 56.7, 81.9, 107.1, 138.6, 189.0,$ and $236.3 \ m^{-1}$.

DI Method: In experiments, the DI method behaved consistently as predicted above. The spectrum reconstruction was observed to be very accurate (up to the fourth significant digit) when FC grid and GC grid were used (with proper weights). The Fliege grid weights are very close to uniform, so an experiment was also done with FC grid using uniform weights. The accuracy of reconstruction was slightly worse in this case but still quite acceptable. With all other (open and closed) grids, the spectrum reconstruction failed. The obtained spectrum somewhat resembled the ground-truth data with respect to positions of peaks, but the peak heights were not correct. The DI method was robust to noise on all grids (i.e., small input perturbations caused only small changes in the output). In summary, while good reconstruction quality was observed on grids created specifically for use with the DI method, an

arbitrary grid can not be used and the method can not be applied to existing HRTF measurements.

LS Method: It was found that the LS method achieved *perfect spectrum reconstruction* on *all* grids, assuming no noise, $p = p_s$, and $p^2 \leq L_q$, for p as high as 33. Differences between grids started to show up if these conditions were violated, as it would be the case in any practical application.

LS Method, Closed Grid: Generally it was found that a reasonable approximation to the ground-truth data was obtained with all closed grids tested. When small white noise (std = 0.01 of max of data) was added to the data, the changes in the output were minimal. When $p < p_s$ (e.g., $p = ka$) was used, the output values still stayed within 2% of the ground-truth data. As expected, the closed grid solution was stable and did not require regularization.

LS Method, Open Grid: In this case, any disturbance to the data (even very small amounts of noise, or setting $p \neq p_s$) caused the spectrum magnitude to become very large ($\sim 10^4$). Interestingly, the quality of fit was still maintained over the data points in the grid (i.e., the fitted surface passed through or near all the data points); however, the computed field values in the cut-off part of the grid were nonsense. This was consistent with the above observation that Y is badly conditioned and needs regularization. Several regularization methods included in [14] were tried, and the results obtained with different methods were quite similar. In particular, Tikhonov regularization with λ determined automatically by toolbox routine "l_curve" was successful in limiting spectrum magnitude and obtaining the solution that was consistently within 5% of the ground-truth data. TSVD regularization dropping last 10% of components as suggested by Picard plots provided similar results. In summary, regularized fitting was successful in recovering the spectrum on *arbitrary* grid. Note that the LS procedure is done in one step and iterative approach as in [6] is not required.

5. EXPERIMENTAL RESULTS

The HRTF of the actual KU-100 dummy head was measured on HO grid using the setup of [13] and on LO grid using the reciprocal setup [4]. The LS algorithm was run with $p = ka$ at a number of frequencies. The obtained HRTF spectrum was used to compute
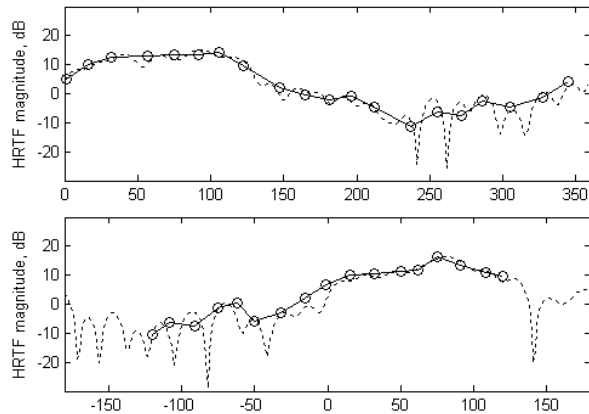
Figure 3: Same as Figure 2, regularized.

continuous (with one-degree resolution) reconstructed HRTF over the whole sphere. Visual comparison of 3-D plots of the original data and of obtained reconstruction provided useful insights into behavior of algorithms. Slices of these plots are shown below.

Figure 2 shows the LS method results without regularization for $k = 107.1 \ m^{-1}$. The top plot shows the right ear HRTF varying along the circle of zero elevation (the sphere equator). Zero angle is in front of the subject, and positive values are to the right. The bottom plot shows the same HRTF varying along the circle in the frontal plane (a plane passing through both ears and through the top of the head). Here, angle of $-180°$ (or $180°$) corresponds to the point below; $-90°$ – directly to the left; zero – above; and $90°$ – directly to the right of the subject. Note that for this plot no measured data were available for angles $< -120°$ or $> 120°$, which allowed evaluation of the behavior of HRTF reconstruction from spectrum for the directions in that area. In the plots, the fit quality is good on the equator and on most of the frontal circle; however, the HRTF magnitude is very large at the bottom of the circle. Figure 3 shows the same plots obtained with Tikhonov regularization using λ provided by the "l_curve" routine. The fit quality is the same on existing data while more meaningful results are obtained in the area where the measurements were missing. It can be hypothesized that use of spherical harmonics for extrapolation of the measured data is the physically-correct method to use, as the fitting is done globally and all information about the scatterer available through measurements is captured and is projected into the cut-off portion of the measurement grid by this procedure.

6. CONCLUSION AND FUTURE WORK

In this paper, a method for computing spherical harmonic spectrum of HRTF (or, in fact, of arbitrary scattered acoustic field) was presented. The method was tested on simulated data and showed successful recovery of the ground-truth spectrum. Unlike the existing methods, the method appears to work well for arbitrary grid and does not require iterative procedure. The computed representation of HRTF can be used in spatial audio playback systems to ease the implementation of space/user rotation and translation.

Some improvements can be possibly made to the procedure. In particular, the evaluation of the spectrum shows some oscillatory behavior in low-energy area. A smoothing algorithm can be applied either before or after computing the spectrum to get better approximation. Another idea is to use iterative fitting procedure

and to include the surface roughness into the cost function; similar algorithms are in active use in shape reconstruction in computer vision (e.g., [15]). Yet another approach is to revert to the direct integration method for obtaining the spectrum and to develop a procedure for computing proper quadrature weights for arbitrary grids. Future work is planned targeting these open questions.

7. REFERENCES

[1] W. M. Hartmann (1999). "How we localize sound", Physics Today, November 1999, pp. 24-29.

[2] D. N. Zotkin, R. Duraiswami, and L. S. Davis (2004). "Rendering localized spatial audio in a virtual auditory space", IEEE Transactions on Multimedia, vol. 6, no. 4, pp. 553-564.

[3] N. A. Gumerov, A. O'Donovan, R. Duraiswami, and D. N. Zotkin (2009). "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 UMIACS-TR-2009-06), Department of Computer Science, University of Maryland, College Park.

[4] D. N. Zotkin, R. Duraiswami, E. Grassi, and N. A. Gumerov (2006). "Fast head-related transfer function measurement via reciprocity", Journal of the Acoustical Society of America, vol. 120, no. 4, pp. 2202-2215.

[5] M. J. Evans, J. A. S. Angus, and A. I. Tew (1998). "Analyzing head-related transfer function measurements using surface spherical harmonics", Journal of the Acoustical Society of America, vol. 104, no. 4, pp. 2400-2411.

[6] W. Zhang, R. A. Kennedy, and T. D. Abhayapala (2008). "Iterative extrapolation algorithm for data reconstruction over sphere", Proc. IEEE ICASSP 2008, Las Vegas, NV, April 2008, pp. 3733-3736.

[7] R. Duraiswami, D. N. Zotkin, and N. A. Gumerov (2004). "Interpolation and range extrapolation of HRTFs", Proc. IEEE ICASSP 2004, Montreal, QC, Canada, May 2004, vol. 4, pp. 45-48.

[8] N. A. Gumerov and R. Duraiswami (2004). "Fast multipole methods for the Helmholtz equation in three dimensions", Elsevier Science, The Netherlands.

[9] M. Taylor (1995). "Cubature for the sphere and the discrete spherical harmonic transform", SIAM Journal of Numerical Analysis, vol. 32, no. 22, pp. 667-670.

[10] J. Fliege and U. Maier (1999). "The distribution of points on the sphere and corresponding cubature formulae", IMA Journal of Numerical Analysis, vol. 19, pp. 317-334.

[11] R. Duraiswami, Z. Li, D. N. Zotkin, E. Grassi, and N. A. Gumerov (2005). "Plane-wave decomposition analysis for spherical microphone arrays", Proc. IEEE WASPAA 2005, New Paltz, NY, October 2005, pp. 150-153.

[12] D. N. Zotkin and R. Duraiswami (2009). "Plane-wave decomposition of acoustical scenes via spherical and cylindrical microphone arrays", to appear in IEEE Transactions on Audio, Speech, and Language Processing.

[13] E. Grassi, J. Tulsi, and S. A. Shamma (2003). "Measurement of head-related transfer functions based on the empirical transfer function estimate", Proc. IEEE ICAD 2003, Boston, MA, July 2003, pp. 119-122.

[14] http://www2.imm.dtu.dk/~pch/Regutools/

[15] J. Li and A. O. Hero (2004). "A fast spectral method for active 3D shape reconstruction", Journal of Mathematical Imaging and Vision, vol. 20, pp. 73-87.