INTERPOLATION AND RANGE EXTRAPOLATION OF HRTFS

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

Perceptual Interfaces and Reality Laboratory, UMIACS, University of Maryland, College Park

ABSTRACT

The Head Related Transfer Function (HRTF) characterizes the scattering properties of a person's anatomy (especially the pinnae, head and torso), and exhibits considerable person-to-person variability. It is usually measured as a part of a tedious experiment, and this leads to the function being sampled at a few angular locations. When the HRTF is needed at intermediate angles its value must be interpolated. Further, its range dependence is also neglected, which is invalid for nearby sources. Since the HRTF arises from a scattering process, it can be characterized as a solution of a scattering problem. In this paper, we show that by taking this viewpoint and performing some analysis we can express the HRTF in terms of a series of multipole solutions of the Helmholtz equation. This approach leads to a natural solution to the problem of HRTF interpolation. Furthermore, we show that the range-dependence of the HRTF in the near-field can also be obtained by extrapolation from measurements at one range.

1. INTRODUCTION

Humans have the remarkable ability to locate a sound source with better than 5° accuracy in both azimuth and elevation, in challenging environments. Multiple cues are involved including those that are produced by sound scattering off the listener themselves [1]. The cues that arise due to scattering off the anatomy of the listener exhibit considerable person-to-person variability. They can be encapsulated in a transfer function that is termed the Head Related Transfer Function (HRTF). To recreate the sound pressure at the eardrums to make a synthetic audio scene indistinguishable from the real one, the virtual audio scene must include the HRTF-based cues to achieve accurate simulation [2].

The HRTF depends on the direction of arrival of the sound, and, for nearby sources, on the source distance, which is usually neglected. If the sound source is located at polar angles (θ, φ), then the (left and right) HRTFs H_l and H_r are defined as the ratio of the complex sound pressure at the corresponding eardrum $\psi_{l,r}$ to the free-field sound pressure at the center of the head as if the listener is absent ψ_f[8]

$$H_{l,r}(\omega, r, \theta, \varphi) = \frac{\psi_{l,r}(\omega, r, \theta, \varphi)}{\psi_f(\omega)}. \quad (1)$$

HRTF interpolation: To synthesize the audio scene given the source location (r, φ, θ) one needs to filter the signal with $H(r, \varphi, \theta)$ and render the result binaurally through headphones. Additionally, the HRTF must be interpolated between discrete measurement positions to avoid audible jumps in sound. Many techniques have been proposed to perform the interpolation of the HRTF, and the correct interpolation is regarded as an open question.

Supported by NSF awards 0086075 and 0205271. Email addresses: { ramani, dz, gumerov}@umiacs.umd.edu

HRTF range dependence: The dependence of the HRTF on the range r is also usually neglected. However, this is known to be incorrect for relatively nearby sources and at lower frequencies. On the other hand, as they are HRTF measurements are relatively tedious and time-consuming procedures, and except for psychophysicists interested in the range dependence effect [8, 5, 6], this effect is neglected and relatively distant sources simulated. For these the range effects can often be synthesized using other cues such as reverberation and intensity [2].

Indeed it might be safe to say that complete range measurements for the HRTF (i.e., for a complete set of values $(r_i, \theta_i, \varphi_i)$ have never been made. However, many applications such as games, auditory user interfaces, entertainment, and virtual reality demand the ability to accurately simulate sounds at relatively close ranges, and some researchers have recently begun measurements of these.

Present contribution: In this paper we present an analysis of the HRTF as a function that is related to the scattering of sound off the human. This analysis enables us to suggest correct answers to both these open problems: we present both the correct interpolation procedure, and a way to obtain the range dependence of the HRTF from existing measurements conducted at a single range!

2. SCATTERING ANALYSIS

When a body with surface S scatters sound from a source located at $(r_1, \theta_1, \varphi_1)$ the complex pressure amplitude ψ at any point (r, θ, φ) is known to satisfy the Helmholtz equation

$$\nabla^2\psi(\mathbf{x}, k) + k^2\psi(\mathbf{x}, k) = 0, \quad k = \omega c^{-1}. \quad (2)$$

Outside a surface S that contains all acoustic sources in the scene, the potential $\psi(\mathbf{x}, k)$ is regular and satisfies Sommerfeld radiation condition at infinity:

$$\lim_{r\to\infty} r\left(\frac{\partial\psi}{\partial r} - ik\psi\right) = 0, \quad r = |\mathbf{x}|. \quad (3)$$

Outside S, we can expand the regular potential $\psi(\mathbf{x}, k)$ that satisfies equation (2) and condition (3) in terms of singular elementary solutions called multipoles [4]. A multipole $\Phi_{lm}(\mathbf{x}, k)$ is characterized by two indices m and l which are called order and degree, respectively. In spherical coordinates, $\mathbf{x} = (r, \theta, \varphi)$

$$\Phi_{lm}(r, \theta, \varphi, k) = h_l(kr)Y_{lm}(\theta, \varphi), \quad (4)$$

where $h_l(kr)$ are the spherical Hankel functions of the first kind, and $Y_{lm}(\theta, \varphi)$ are the spherical harmonics,

$$Y_{lm}(\theta, \varphi) = (-1)^m\sqrt{\frac{2n+1}{4\pi}\frac{(l-|m|)!}{(l+|m|)!}}P_l^{|m|}(\cos\theta)e^{im\varphi} \quad (5)$$

where $P_l^{|m|}(\lambda)$ are the associated Legendre polynomials.

3. FITTING HRTFS

We need a representation of the potential in the region between the head and the many speaker locations. Unfortunately this region contains sources (the speaker), and the scatterer, and thus does not satisfy the conditions for the fitting by multipoles discussed above (source free, and extending to infinity). We remove this difficulty using the reciprocity principle [4]. This states that if the acoustic source at point A in arbitrary complex audio scene creates potential ψ at a point B, then the same acoustic source placed at point B will create the same potential ψ at point A. The acoustic field might be different elsewhere in the scene, but the signal picked up at the receiver is the same if source and receiver locations are interchanged.

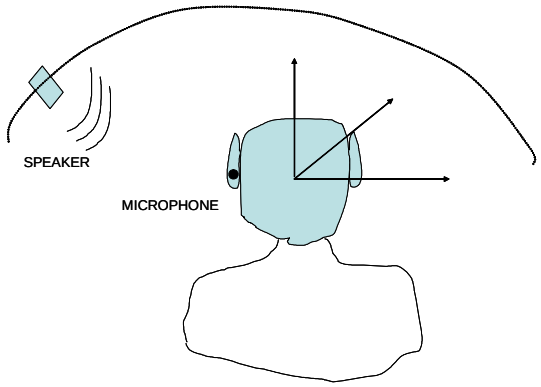

Fig. 1. Typical HRTF measurement set-up. To get the HRTF at a location, a speaker there makes a sound and the sound received by a microphone in the ear is processed. The measurement is repeated with the speaker moved to other locations.

The usual method of HRTF measurement is to place a transmitter (a loudspeaker) at different points in space and a microphone in the ear, emit the signal at the loudspeaker and record it at the microphone. However, by reciprocity an identical recording would be obtained if the transmitter were placed in the ear and the receiver were at the original position of the loudspeaker. Thus we may take the multipath sound from the speaker received at the ear microphone to be the multipath sound at the speaker location, if the idealized point speaker were in the ear. This means we can represent each ψ as

$$\psi = \sum_{l=0}^{\infty} \left(\sum_{m=-l}^{l} \alpha_{lm} h_l\left(kr\right) Y_{lm}\left(\theta, \varphi\right) \right). \tag{6}$$

In practice we truncate the outer summation at some value of l called the truncation number p (summation from $l = 0$ to $p - 1$ only) and ignore terms from p to ∞. After such a truncation, there are a total of $M = p^2$ terms left in the multipole expansion. The values of the potential $\psi_h(\mathbf{x}, k)$ are known at N measurement points at the reference sphere, $\{\mathbf{x}_1, ..., \mathbf{x}_N\}$. Now we can fit the α_{lm} using a regularized fitting approach by writing N linear equations for the M unknowns α_{lm}:

$$\begin{aligned} \psi_h(\mathbf{x}_1, k) &= \sum_{l=0}^{p-1} \sum_{m=-l}^{l} \alpha_{lm} \Phi_{lm}(\mathbf{x}_1, k), \\ &\cdots \\ \psi_h(\mathbf{x}_N, k) &= \sum_{l=0}^{p-1} \sum_{m=-l}^{l} \alpha_{lm} \Phi_{lm}(\mathbf{x}_N, k), \end{aligned} \tag{7}$$

or, in short form, $\Phi A = \Psi$, where the Φ is $N \times M$ matrix of the values of multipoles at measurement points, A is the unknown vector of coefficients of length M, and Ψ is a vector of potential values of length N. This system is usually overdetermined ($N > M$), and solved in the least squares sense.

Once the equations (7) are solved and the set of coefficients α determined, the acoustic field can be evaluated at any desired point outside the sphere. **This means that we can evaluate it at points to be interpolated, and at points with a different range.**

Obviously, a certain spatial resolution is necessary to capture the potential field and the spatial resolution is related to the wavelength by the Nyquist criteria [3]. It can be shown that the number of the measurement points necessary to obtain accurate holographic recording for up to the limit of human hearing is about 2000, which is almost twice as big as the number of HRTF measurement points in any currently existing HRTF measurement system. The sphere radius used in these measurements does not matter, because by our reciprocity analysis the only requirement is that all sources are contained within a sphere S of a small radius, and outside this sphere only the angular resolution matters due to fitting with multipoles.

4. IMPLEMENTATION DETAILS

Choice of Truncation Number: The primary parameter that affects the quality of the fitting is the truncation number p. A higher truncation number results in better quality of fitting for a fixed r, but p being too large leads to overfitting. The general rule of thumb is that the truncation number should be roughly equal to the wavenumber for good interpolation quality [9]. Such rules are also used in the fast multipole method literature. Indeed, if the wavenumber is small, the potential field cannot vary fast and high-degree multipoles are unnecessary for a good fit. However, high-degree multipoles can have devastating effect when the potential field approximated at r_h is evaluated at $r < r_h$ because of exponential growth of the spherical Bessel functions of the first kind $j_l(kr)$ as the argument kr approaches zero. Thus, we set

$$p = \text{integer}(kr) + 1. \tag{8}$$

When doing resynthesis, this can lead to artifacts when two adjoint frequency bins are processed with different truncation numbers, and a solution must be developed for this.

Tikhonov Regularization: Use of regularization helps avoid blow-up of the approximated function in areas where no data is available (usually at low elevations) and thus the function is not constrained. With Tikhonov regularization the equation becomes

$$(\Phi^T \Phi + \varepsilon D) A = \Phi^T \Psi. \tag{9}$$

Here ε is the regularization coefficient, D is the diagonal damping or regularization matrix. In our computations we set

$$D = (1 + l(l+1))I \tag{10}$$

where l is the degree of the corresponding multipole coefficient and I is the identity matrix. In this way, high-degree harmonics are penalized more than low-degree ones which was seen to improve interpolation quality, and avoid excessive "jagging" of the approximation. Even small values of ε prevent blowup in the unconstrained area, so we set $\varepsilon = 10^{-6}$ for our experiments.

5. RESULTS

We compared the proposed technique for both range and angular interpolation. First, some analytical solutions were considered, and next some experimental HRTF measurements.

Analytical Solutions: We tested the proposed method of range HRTF interpolation on synthetic data first. We used the well-known analytical sphere HRTF model presented in [8] and a grid consisting of 1636 points on the sphere. (Care must be taking in choosing points, since for the sphere the HRTF does not depend on the elevation. Thus, to choose a meaningful grid enough different azimuths must be sampled). We compute ψ at each grid point at a distance of 1 m and use it in the interpolation of the multipole decomposition of the potential field. We then evaluate the resulting decomposition at 0.5, 0.25 and 0.125 m. We set the truncation number at every bin using (8) with r taken to be the radius of the sphere which encloses the source, and outside which the decomposition is to be used, here 0.125m; the highest p is about 28, at 12 kHz. In Figure 2, the magnitudes of the analytical HRTF for the sphere at different ranges at the equator are shown on the left, compared to the reconstruction on the right.

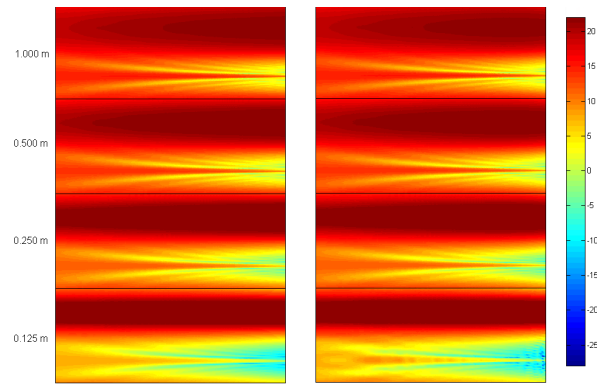

Fig. 2. Left: Analytical solution of a sphere HRTF at 1.0, 0.5, 0.25 and 0.125m. Right: Results of HRTF extrapolation at the same ranges. Horizontal axis is frequency; vertical axis is azimuth.

The plots reveal that the reconstruction of the HRTF at 1 meter and extrapolation of it to 0.5m and 0.25m are perfect. There are a few artifacts at the closest range due to the discrete changes in the truncation number along the frequency axis. These may be alleviated by blending the reconstructions in adjacent bins or with adjacent truncation numbers. It can be concluded from these experiments on synthetic data that the method works in this simple case.

To better understand the dependence of the error on the truncation number and the grid density, we studied the average HRTF magnitude error per grid point for different grids. (Plot omitted due to space). For both grids, the error grows with wavenumber, and a denser grid gives better approximation for all wavenumbers.

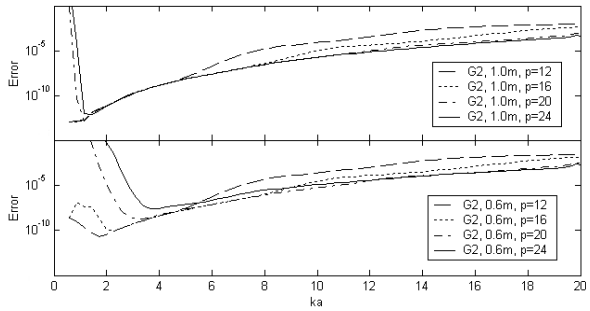

Fig. 3. Top: HRTF reconstruction error at 1.0 m versus frequency for two-sphere case for different truncation numbers. Bottom: HRTF extrapolation error from 1.0 m to 0.6 m for the same case.

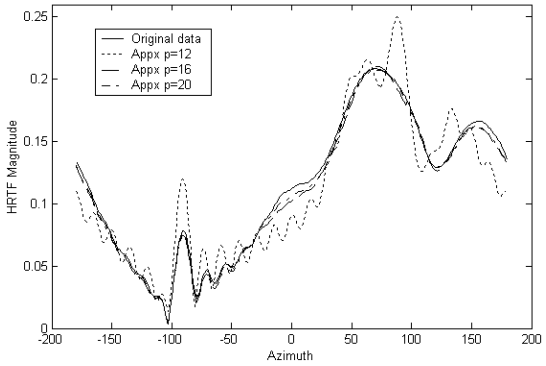

Fig. 4. Sample behavior of reconstructed HRTF in the equatorial plane for different truncation numbers for the two-sphere case.

As the wavenumber reaches that in (8), the error starts to increase. Reconstruction of the HRTF with a truncation number set too high increases the error, with higher p resulting in larger errors.

We also analyzed the error behavior in a more complex synthetic case involving 2 close-by spheres. We compute the "ground truth" solution for four different ranges using the software of [9], decompose the HRTF at 1.0 m to obtain the α_{lm}, compute the result of range interpolation at closer ranges and compare. The results exhibit similar behavior to the one-sphere case described above. Agreement of HRTF contour plots (similar to Figure 2; plots are not shown) is good, and plots of the average extrapolation error per grid point versus ka for two-sphere case are shown in Figure 3 (top plot is HRTF reconstruction error at 1.0 meter, bottom plot is the HRTF extrapolation error from 1.0 to 0.6 meters). As before, the approximation and extrapolation quality depends on the truncation number and starts to degrade when p is approximately equal to kr, and higher truncation numbers cause exponential growth of the error at low kr, similar to the one-sphere case.

A sample plot showing the angular interpolation of the HRTF within the same range is shown in Figure 4. For this plot, the HRTF magnitude is plotted on the equatorial circle for two-sphere case described above at 7.4 kHz. The solid line is the original analytically computed data, and the broken lines are the approximation with truncation numbers of 12, 16 and 20, respectively. It can be seen that if truncation number is insufficient the jagging

artefacts appear and angular interpolation quality is poor, but as it becomes smooth as it approaches the estimate in equation (8).

Real Data: We used KEMAR mannikin measurements because these are the only available HRTF range measurements. We wanted to reconstruct HRTFs from measured KEMAR HRTF data [7] and to compare it with the measured range HRTFs from [5] where measurements of the KEMAR HRTF at distances of 1, 0.5, 0.25 and 0.125 meters are presented. In [5] the acquisition of the HRTF is not performed at the whole sphere but is rather done only on the equator. The data from [7] that have lower resolution (710 points over the sphere) and are taken farther away (at 1.4 meters). We convert the head-related impulse responses (HRIR) $h(r,t)$ stored in the database to HRTFs $H(r,\omega)$, which we took to be equal to the potential $\psi(r,\omega)$ that satisfies equation (2). The decomposition is performed with fixed $p = 25$ and $(p+1)^2 = 676$ coefficients. The expansion is evaluated at 1.0, 0.5, 0.25 and 0.125 meters, dropping coefficients with $l > kr$ to prevent divergence. The original HRTF and reconstruction results at different ranges are shown on the left side of Figure 5, and the results of KEMAR HRTF measurements at different distances. The experimental results are scanned in from [5], since we could not obtain the original data despite several attempts. These are shown on the right side of the corresponding figure. The colormaps and scales in the two panels were attempted to be matched, but are slightly different due to scanning difficulties.

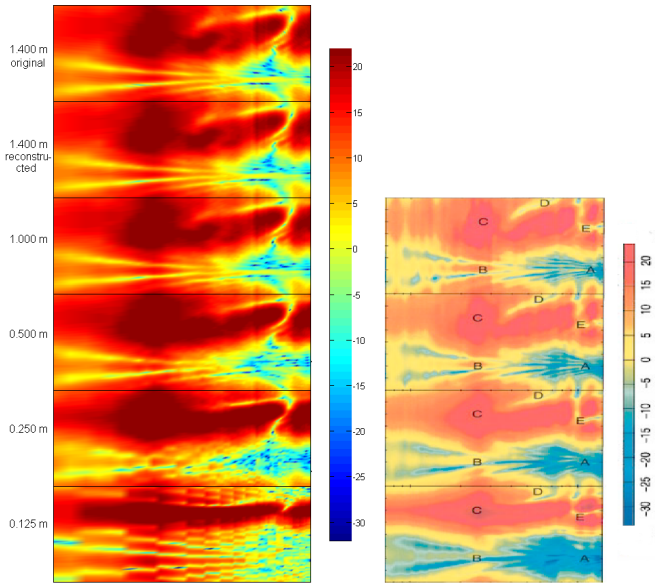

Fig. 5. Left: KEMAR HRTF measured at 1.4 meters and results of HRTF reconstruction at 1.4 meters and HRTF extrapolation to 1.0, 0.5, 0.25 and 0.125 meters. Right: KEMAR HRTF measured at 1.0, 0.5, 0.25 and 0.125 meters (scanned in from [5]).

The general tendencies observed in the reconstructed data agree well with the trends observed on measured data sets. As the source moves toward the head, the magnitude of HRTF increases when a direct path exists between the source and the ear. When the ear is in the acoustic shadow of the head the shadow becomes deeper. The shadow region also grows as the source approaches the head, which can be expected from simple geometric observations, and overall the magnitude growth is bigger at lower frequencies than at higher ones, so the source signal gets effectively low-pass filtered as it approaches the head. This is true both for the spherical model and the KEMAR measurements.

However, the agreement between the reconstructed and measured data is not as exact for the synthetic case of the sphere. That can be attributed to several reasons. One is the different sources of measurements (where it is known different speaker types were used); the other is inadequate spatial resolution of the original data which does not satisfy the Nyquist criteria for higher frequencies. Also, the reconstruction at 0.125 m is heavily influenced by use of different truncation numbers at different frequencies.

It is likely that the perceptual features of the HRTF at close range are correctly captured using the proposed reconstruction because the features generally follow what is expected from physical arguments (growth of the magnitude of HRTF in the direct path region, enlargement of the shadow region and low-pass filtering of HRTF as source approaches the head). These are also the cues that are responsible for evolution of the cones of confusion into tori of confusion at close range (see [6]).

6. CONCLUSIONS

Our technique is a powerful method for computation of range HRTF from a given set of HRTF measurements at a fixed distance. We obtain good preliminary results using both synthetic data and real mannikin head measurements; the agreement between measured and predicted HRTFs is excellent for the synthetic data and good for the real data set, except for the closest range of 0.125 meters. In the near future, we plan to perform experiments to measure the mannikin HRTF over the whole sphere at several ranges using dense measurement grid and to further evaluate performance of the spherical holography using those measurements, and test the perceptual fidelity of the range HRTFs generated.

7. REFERENCES

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

[2] D.N. Zotkin, R. Duraiswami, L.S. Davis (2003). "Creation of Virtual Auditory Spaces," accepted IEEE Trans. Multimedia. (available off authors' homepages).

[3] J. D. Maynard, E. G. Williams, Y. Lee (1985). "Nearfield acoustic holography: Theory of generalized holography and the development of NAH", J. Acoust. Soc. Am., 78, pp. 1395-1413.

[4] P. M. Morse, K. U. Ingard (1968). "Theoretical acoustics", Princeton University Press, New Jersey.

[5] D. S. Brungart, W. M. Rabinowitz (1998). "Auditory localization of nearby sources. Head-related transfer functions", J. Acoust. Soc. Am., 106, pp. 1465-1479.

[6] B.G. Shinn-Cunningham, S. G. Santarelli, and N. Kopco (2000). "Tori of confusion: Binaural localization cues for sources within reach of a listener", J. Acoust. Soc. Am., 107, pp. 1627-1636.

[7] http://sound.media.mit.edu/KEMAR.html

[8] R. O. Duda and W. M. Martens (1998). "Range dependence of the response of a spherical head model", J. Acoust. Soc. Am., 104, pp. 3048-3058.

[9] N. A. Gumerov and R. Duraiswami (2002). "Computation of scattering from N spheres using multipole reexpansion", J. Acoust. Soc. Am., 112, pp. 2688-2701.