HEADPHONE-BASED REPRODUCTION OF 3D AUDITORY SCENES CAPTURED BY SPHERICAL/HEMISPHERICAL MICROPHONE ARRAYS

Zhiyun Li, Ramani Duraiswami

Perceptual Interfaces and Reality Laboratory, UMIACS, Univ. of Maryland, College Park, MD 20742
zli@cs.umd.edu; ramani@umiacs.umd.edu

ABSTRACT

We propose a method to reproduce 3D auditory scenes captured by spherical microphone arrays over headphones. This algorithm employs expansions of the captured sound and the head related transfer function over the sphere and uses the orthonormality of the spherical harmonics. Using a spherical microphone array, we first record the 3D auditory scene, then the recordings are spatially filtered and reproduced through headphones in the orthogonal beam-space of the head related transfer functions (HRTFs). We use the KEMAR HRTF measurements to verify our algorithm. In experiments, we use a hemispherical array for recording. The reproduction results are posted online.

1. INTRODUCTION

Currently available headphone-based personal audio systems can only recreate a limited 3D auditory scene because when the user rotates his head, the auditory scene moves also. In another words, the auditory scene is fixed to his head by the headphones. That is different from the real-world experience where the auditory scene is independent on the head rotation. Some HRTF-based technologies mainly aim to use headphone to create virtual sound sources at user specified spatial positions, but are unable to recreate scenes from real-world 3D recordings [1][4].

To reproduce real-world 3D auditory scenes through headphones from 3D recordings, a straightforward method is to localize, track and beamform the sound sources and then use the corresponding HRTF measurements to filter the beamformed signals before playback over headphones. This is reasonable for a few sound sources in simple scenes. However, for complex scenes with many sources and much reverberation (and thus thousands of virtual sources), this method will fail. Even worse, in complex scenes with more sound sources, the localization and near real-time tracking become very difficult. Recently an alternate, heuristically based approach for which some convincing demonstrations have been produced, has been proposed [3]. It has been used to produce quite convincing reproductions. In it one simply chooses two microphones at locations approximately corresponding to the ear positions of a listener from a set of microphones on a spherical array, and then just play back the recordings through headphones. Here, we seek to extend this idea rigorously and incorporate HRTF cues in the playback.

In this paper, we will develop a coupled theory based on the orthonormality of the spherical harmonics[1]. By using a spherical microphone array, we first decompose the recorded 3D soundfield in orthogonal beam-space, then we use the resulting beampattern to approximate the HRTFs for all 3D directions. Our method is independent of the locations of the sound sources and the surrounding environment, except that it is assumed that the microphone array does not disrupt the acoustics in the recording room. In our experiments, we first use KEMAR HRTF measurements to verify our algorithm, which is then applied to the real-world 3D auditory scenes recorded by our hemispherical microphone array as described in [9].

2. PRINCIPLE OF SPHERICAL BEAMFORMING

The basic principle of a spherical beamformer is to make use of the orthonormality of spherical harmonics to decompose the soundfield arriving at a spherical array. Then the orthogonal components of the soundfield are linearly combined to approximate a desired beampattern [11].

For a unit magnitude plane wave $\mathbf{k}$, incident from direction (θ_k, φ_k), the complex pressure field on the surface $(\theta_s, \varphi_s, r_s = a)$ of the rigid sphere is [12]:

$$p_t = 4\pi \sum_{n=0}^{\infty} i^n b_n(ka) \sum_{m=-n}^{n} Y_n^m(\theta_k, \varphi_k) Y_n^{m*}(\theta_s, \varphi_s), \tag{1}$$

$$b_n(ka) = j_n(ka) - \frac{j_n^{'}(ka)}{h_n^{'}(ka)} h_n(ka), \tag{2}$$

where j_n is the spherical Bessel function of order n, Y_n^m is the spherical harmonics of order n and degree m. * denotes the complex conjugation. h_n is the spherical Hankel function of the first kind.

If we assume that the pressure recorded at each point (θ_s, φ_s) on the surface of the sphere Ω_s, is weighted by

$$W_{n'}^{m'}(\theta_s, \varphi_s, ka) = \frac{Y_{n'}^{m'}(\theta_s, \varphi_s)}{4\pi i^{n'} b_{n'}(ka)}. \tag{3}$$

Then making use of orthonormality of spherical harmonics:

$$\int_{\Omega_s} Y_n^{m*}(\theta_s, \varphi_s) Y_{n'}^{m'}(\theta_s, \varphi_s) d\Omega_s = \delta_{nn'} \delta_{mm'} \tag{4}$$

the total output from a pressure-sensitive spherical surface is:

$$P = \int_{\Omega_s} p_t W_{n'}^{m'}(\theta_s, \varphi_s, ka) d\Omega_s = Y_{n'}^{m'}(\theta_k, \varphi_k) \tag{5}$$

This work was partially supported by NSF Award 0205271.

[1] An equivalent but theoretically stricter approach based on band-limited Herglotz wave function was introduced in [6].

This shows the gain of the plane wave coming from (θ_k, φ_k), for a continuous pressure-sensitive spherical microphone, is $Y_{n'}^{m'}(\theta_k, \varphi_k)$. Since an arbitrary real function $F(\theta, \varphi)$ can be expanded in terms of complex spherical harmonics, we can implement arbitrary beampatterns. For example, an ideal beampattern looking at the direction (θ_0, φ_0) can be modeled as a delta function:

$$F(\theta, \varphi) = \delta(\theta - \theta_0, \varphi - \varphi_0), \tag{6}$$

which can be expanded into an infinite series of spherical harmonics [2]:

$$F(\theta, \varphi) = 2\pi \sum_{n=0}^{\infty} \sum_{m=-n}^{n} Y_n^{m*}(\theta_0, \varphi_0) Y_n^m(\theta, \varphi). \tag{7}$$

So the weight at each point (θ_s, φ_s) to achieve this beampattern is:

$$w = \sum_{n=0}^{\infty} \frac{1}{2i^n b_n(ka)} \sum_{m=-n}^{n} Y_n^{m*}(\theta_0, \varphi_0) Y_n^m(\theta_s, \varphi_s). \tag{8}$$

The advantage of this system is that it can be steered into any 3D directions *digitally* with the same beampattern. This is for an ideal continuous microphone array on spherical surface.

For discrete arrays with finite number of microphones, the practical beampattern is a truncated version of (7) to some limited order N:

$$F_N(\theta, \varphi) = 2\pi \sum_{n=0}^{N} \sum_{m=-n}^{n} Y_n^{m*}(\theta_0, \varphi_0) Y_n^m(\theta, \varphi). \tag{9}$$

3. IDEAL HRTF SELECTION

In an ideal case, we assume the HRTF is already measured continuously on the spherical surface of radius r. Our goal is to select the correct HRTF for a specified direction. Although it seems trivial for an ideal case, we will use this as a starting point and extend it to more practical cases in the following sections.

We drop the arguments k and r for simplicity, the HRTF for the sound of wave number k from the point (r, θ, φ) is [5]:

$$\psi(\theta, \varphi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \alpha_{nm} h_n(kr) Y_n^m(\theta, \varphi), \tag{10}$$

where h_n and Y_n^m have the same definitions as in the last section, and α_{nm} are the fitting coefficients which can be determined using real-world discrete HRTF measurements [5].

Suppose we want to select the HRTF for the direction (θ_k, φ_k), we apply the following delta function (ideal beampattern) to each measured HRTF:

$$F(\theta, \varphi) = \delta(\theta - \theta_k, \varphi - \varphi_k), \tag{11}$$

we have:

$$\int_{\Omega_s} \psi(\theta, \varphi) F(\theta, \varphi) d\Omega_s = \psi(\theta_k, \varphi_k), \tag{12}$$

where Ω_s is the spherical surface. Obviously, the delta function simply selects the value we need and discards everything else.

To present another viewpoint of the HRTF selection, we rewrite (12) into a more "complicated" form by using (7):

$$\begin{aligned} &\int_{\Omega_s} \left[\sum_{n=0}^{\infty} \sum_{m=-n}^{n} \alpha_{nm} h_n(kr) Y_n^m(\theta, \varphi) \right] \\ &\quad \times \left[2\pi \sum_{n=0}^{\infty} \sum_{m=-n}^{n} Y_n^{m*}(\theta, \varphi) Y_n^m(\theta_k, \varphi_k) \right] d\Omega_s \\ &= \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \alpha_{nm} h_n(kr) Y_n^m(\theta_k, \varphi_k). \end{aligned} \tag{13}$$

Alternatively, this can be easily proven by using the orthonormality of spherical harmonics (4).

4. HRTF APPROXIMATION IN ORTHOGONAL BEAM-SPACE

In practice, however, HRTFs are measured on discrete points. In this case, (13) and (4) can only hold approximately and to finite order. In addition, using a practical spherical array with finite number of microphones, the beampattern is (9).

The HRTF for the sound of wave number k from the measurement point (r, θ_l, φ_l) is:

$$\psi(\theta_l, \varphi_l) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \alpha_{nm} h_n(kr) Y_n^m(\theta_l, \varphi_l), \tag{14}$$

$$(\ l = 1, ..., B\ )$$

where B is the number of HRTF measurements.

The weighted combination of HRTFs then becomes:

$$\sum_{l=1}^{B} \psi(\theta_l, \varphi_l) F_N(\theta_k, \varphi_k, \theta_l, \varphi_l). \tag{15}$$

If the HRTF measurement points $(\theta_l, \varphi_l), l = 1, ...B$, are approximately uniformly distributed on a spherical surface so that the orthonormality of spherical harmonics holds up to order N', then the HRTF can be expanded into two groups:

$$\psi(\theta_l, \varphi_l) = \psi_0^{N'}(\theta_l, \varphi_l) + \psi_{N'+1}^{\infty}(\theta_l, \varphi_l), \tag{16}$$

where

$$\psi_0^{N'}(\theta_l, \varphi_l) = \sum_{n=0}^{N'} \sum_{m=-n}^{n} \alpha_{nm} h_n(kr) Y_n^m(\theta_l, \varphi_l), \tag{17}$$

$$\psi_{N'+1}^{\infty}(\theta_l, \varphi_l) = \sum_{n=N'+1}^{\infty} \sum_{m=-n}^{n} \alpha_{nm} h_n(kr) Y_n^m(\theta_l, \varphi_l). \tag{18}$$

So (15) can be rewritten as:

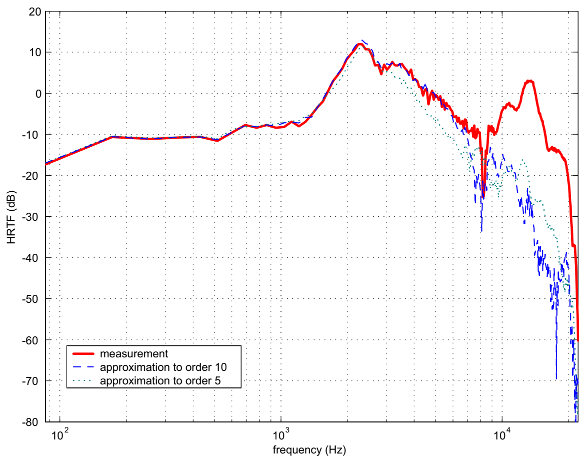

Fig. 1. HRTF approximations to orders 5 and 10. Plot shows the magnitude in dB scale.

$$\sum_{l=1}^{B}\left[\psi_0^{N'}(\theta_l,\varphi_l)+\psi_{N'+1}^{\infty}(\theta_l,\varphi_l)\right]F_N(\theta_k,\varphi_k,\theta_l,\varphi_l)$$

$$=\sum_{l=1}^{B}\psi_0^{N'}(\theta_l,\varphi_l)F_N(\theta_k,\varphi_k,\theta_l,\varphi_l) \tag{19}$$

$$+\sum_{l=1}^{B}\psi_{N'+1}^{\infty}(\theta_l,\varphi_l)F_N(\theta_k,\varphi_k,\theta_l,\varphi_l) \tag{20}$$

$$=\psi_0^{\min(N',N)}(\theta_k,\varphi_k)+\epsilon \tag{21}$$

which is the approximation of HRTF up to the order $\min(N', N)$. Here the error ϵ consists of two parts: one is the orthonormality error from (19) which is supposed to be small according to the discrete orthonormalities; the other is from (20) which is also small with well-chosen N' because of the convergence of the series expansion in (14). In general, this is a quadrature problem over the spherical surface for spherical harmonics. More details can be found in [7][8][10][6].

If HRTFs are not measured on uniformly distributed angular points, which is the case for all currently available measurements, we can first obtain a uniform version via interpolation [5]. In practice the HRTF measurement points are significantly more than microphones on a spherical array. In this case, the HRTF approximation at (θ_k, φ_k) depends only on the order of beampattern N, which is:

$$\sum_{l=1}^{B}\psi(\theta_l,\varphi_l)F_N(\theta_k,\varphi_k,\theta_l,\varphi_l)=\psi_0^N(\theta_k,\varphi_k)+\epsilon. \tag{22}$$

Therefore, if there is a plane wave incident from (θ_k, φ_k) in the original auditory scene, it will be automatically filtered with the corresponding HRTF, in the approximation of order N.

5. REPRODUCTION ALGORITHM

Suppose we have built a spherical microphone array to record a 3D auditory scene. The spherical beamformer for this array has the beampattern as in (9). To reproduce the 3D auditory scene from the recordings, there are three steps:

1. beamform the recordings to (θ_l, φ_l) for $l = 1, ..., B$ (the "uniformly" interpolated point);
2. filter the beamformed signal at (θ_l, φ_l) with the measured HRTF $\psi(\theta_l, \varphi_l)$ for $l = 1, ..., B$;
3. superimpose the resulted signals for $l = 1, ..., B$.

Suppose we have sufficient HRTF measurements, the only factor that determines reproduction quality is the beampattern order N of the spherical microphone array.

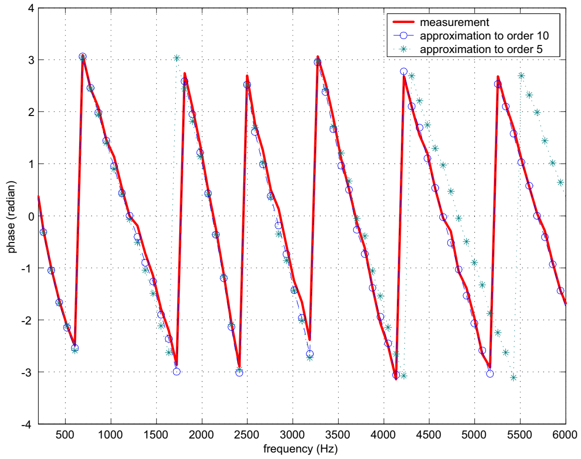

Fig. 2. Phases of the approximations to orders 5 and 10.

6. VERIFICATION AND EXPERIMENTS

We use the KEMAR HRTF measurements [1] to demonstrate our algorithm. In Fig. 1, the red (solid) line shows the HRTF measurement at the position just in front of the manikin. The green (dot) line shows the approximation to order five supposing we have a spherical microphone array of order five. It is a good approximation for frequencies until about 2KHz. It is also a relatively close approximation until 4KHz which may be used in spatial speech acquisition and reproduction. The blue (dash) line shows the approximation to order 10, which closely matches the measurement until about 6KHz. The phases are compared in Fig. 2.

For efficient implementation in practice, the beamformer should be approximated at different orders for different frequency bands. In [9], we described a hemispherical microphone array as shown in Fig. 3, which is used to record 3D auditory scenes in our experiments. The experimental results using a hemispherical microphone array are posted online[2].

7. SUMMARY

In summary, we have developed the theory of reproducing 3D auditory scene using headphones from recordings of a spherical microphone array. We use the spherical microphone array since it

[2] http://www.umiacs.umd.edu/~zli/hemisphere/

provides a natural way to decompose the 3D soundfield in orthogonal beam-space which will be used to approximate the HRTF measurements. The advantage of our method lies in its independence of the sound source locations and the surrounding environment, only if under the far-field assumption. Preliminary design examples are presented to justify our approach. Experimental results using the recordings from our hemispherical array are presented online. Future work may include reduced-dimensional description of HRTF measurements, efficient data structure, extension to near-field case, etc.

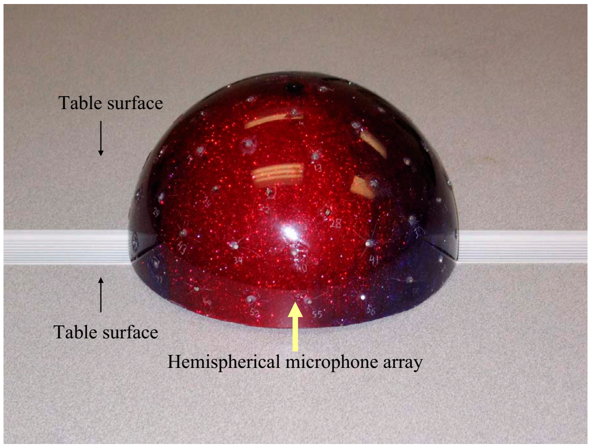

Fig. 3. A hemispherical microphone array built on the surface of a half bowling ball. Its radius is 10.925cm.

8. REFERENCES

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

[2] M. Abramowitz and I. A. Stegun, editors. *Handbook of Mathematical Functions*. U.S. Government Printing Office, 1964.

[3] V. R. Algazi, R. O. Duda, and D. Thompson. Dynamic binaural sound capture and reproduction. US Patent No: US20040076301A1, Apr. 2004.

[4] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano. The CIPIC HRTF database. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics(WASPAA'01)*, pages 99–102, New Paltz, NY, Oct. 2001.

[5] R. Duraiswami, D. Zotkin, and N. Gumerov. Interpolation and range extrapolation of HRTFs. In *IEEE ICASSP'04*, pages IV45–IV48, Montreal, Canada, May 17-21 2004.

[6] R. Duraiswami, D. N. Zotkin, Z. Li, E. Grassi, N. A. Gumerov, and L. S. Davis. System for capturing of high-order spatial audio using spherical microphone array and binaural head-tracked playback over headphones with head related transfer function cues. In *AES 119th Convention*, New York, NY, Oct. 2005.

[7] J. Fliege and U. Maier. The distribution of points on the sphere and corresponding cubature formulae. *IMA Journal on Numerical Analysis*, 19:317–334, 1999.

[8] R. H. Hardin and N. J. A. Sloane. McLaren's improved snub cube and other new spherical designs in three dimensions. *Discrete and Computational Geometry*, 15:429–441, 1996.

[9] Z. Li and R. Duraiswami. Hemispherical microphone arrays for sound capture and beamforming. In *IEEE WASPAA'05*, pages 106–109, New Paltz, New York, Oct. 2005.

[10] Z. Li and R. Duraiswami. A robust and self-reconfigurable design of spherical microphone array for multi-resolution beamforming. In *IEEE ICASSP'05*, volume IV, pages 1137–1140, Mar. 2005.

[11] J. Meyer and G. Elko. A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield. In *IEEE ICASSP'02*, volume 2, pages 1781–1784, May 2002.

[12] E. G. Williams. *Fourier Acoustics*. Academic Press, San Diego, 1999.