Modeling Head Related Transfer Functions

Richard O. Duda
Department of Electrical Engineering
San Jose State University
San Jose, CA 95192
email: duda@ai.sri.com

Abstract

Head Related Transfer Functions (HRTF's) characterize the transformation of a sound source to the sounds reaching the eardrums, and are central to binaural hearing. Because they are the result of wave propagation and diffraction, they can only be approximated by finitely parameterized filters. The functional dependence of the HRTF on azimuth and elevation is described, the requirements for a model are discussed, and existing models are reviewed.

1 Introduction

When sound waves are propagated from a vibrating source to a listener, the pressure waveform is altered by diffraction caused by the torso, shoulders, head and pinnae. In engineering terms, these propagation effects can be captured by two transfer functions, H_L and H_R, that specify the relation between the sound pressure of the source and the sound pressures at the left and right ear drums of the listener. These so-called Head Related Transfer Functions (HRTF's) are acoustic filters that vary both with frequency and with the azimuth θ, elevation ϕ and range r to the source[3]. If a monaural sound signal representing the source is passed through these filters and heard through headphones, the listener will hear a sound that seems to come from a particular location in space. Appropriate variation of the filter characteristics will cause the sound to appear to come from any desired spatial location [23, 24]. This is the basis of the Convolvotron, a virtual acoustic display system that generates convincing 3D sounds through headphones by real-time convolution with experimentally measured HRTF's [21]. Such systems have a number of possible applications, including providing more effective acoustic information to pilots [14], supporting teleconferencing, and enhancing virtual reality systems [21, 8].

Several attempts have been made to model HRTF's, both to understand their behavior and to simplify the binaural synthesis process. This system identification task has been complicated by four major problems: (a) the difficulty of approximating the effects of wave propagation and diffraction by simple, low-order parameterized filters, (b) the complicated joint dependence of the HRTF's on azimuth, elevation and range, (c) the lack of a quantitative criterion for measuring the accuracy of an approximation, and (d) the great person-to-person variability of HRTF's. In this paper, we present the requirements for an effective model and review the various models that have been proposed.

2 The HRTF and Localization Cues

Psychoacousticians have extensively studied the various cues that people apparently use to determine the location of sound sources [3, 15]. It is well known that the primary cues for azimuth are binaural, and include the interaural time difference (ITD) and the interaural intensity difference (IID). It is widely believed that the most important elevation cues are monaural, and are derived from changes that occur in the spectrum as the sound source moves up or down, although we have shown that outside of the median plane the elevation can be accurately recovered from the IID alone [7]. The cues for range and the qualities that make sounds heard over headphones seem externalized are only partially understood [15]; since people are not particularly good at range estimation, we limit our attention to azimuth and elevation.

Fig. 1 shows experimentally measured head related impulse responses $h_R(t, \theta, \phi)$ for the right ear, both in the horizontal plane and in the median sagittal plane. The Fourier transform $H_R(\omega, \theta, \phi)$ captures the same information in the frequency domain, with the amplitude $A_R = 20 \log_{10} |H_R|$ providing intensity information and the phase $\Phi_R = \angle H_R$ providing timing

Preprint for the Twenty-Seventh Asilomar Conference on Signals, Systems & Computers (Asilomar, CA, October 31-November 3, 1993)

information. The binaural IID and ITD differences are easiest to understand in the frequency domain. In particular, if $H_L(\omega, \theta, \phi)$ is the HRTF for the left ear, and if $H = H_R/H_L$ is the interaural transfer function, then $A = 20 \log_{10} |H| = A_R - A_L$ gives the spectral IID and the group delay $\partial\Phi/\partial\omega = \partial(\Phi_R - \Phi_L)/\partial\omega$ gives the spectral ITD.

Fig. 2 shows how the IID varies with azimuth in the horizontal plane. Although this frequency response is rather complex, at any particular frequency it varies roughly sinusoidally with azimuth. Measurement of the group delay shows that the ITD also varies roughly sinusoidally with azimuth, although it is about 50% higher at low frequencies than at high frequencies [12]. For a symmetric head, both the IID and the ITD are zero in the median plane, and the cues for localization in elevation are monaural. Fig. 3 shows how A_R varies with elevation in the median plane. The major response peak around 3.5 kHz is due to ear-canal resonance, and the moving features are due to refraction by the outer ears or pinnae.

Unfortunately, the full behavior of the HRTF can not be determined from behavior in the orthogonal horizontal and median planes, and the HRTF is a rather complicated joint function of azimuth and elevation. Significant simplification can be achieved by using a spherical coordinate system in which the polar axis coincides with the interaural axis. In these coordinates, to a first approximation the ITD varies only with azimuth, and can be obtained from the results in the horizontal plane. Furthermore, the IID can be expanded in a Fourier series in azimuth, the first term of which has the form $b_1(\omega, \phi) \sin\theta$, where the Fourier coefficient b_1 is shown in Fig. 4 [7]. Although this is a rough approximation, it greatly simplifies the problem by factoring the azimuth and the elevation dependence.

If one wants to model such complicated functions, a key problem is to identify their essential characteristics. Unfortunately, we have no meaningful quantitative error criterion for determining the quality of an approximation, and we must turn to listening tests. When the subject whose HRTF has been measured listens through headphones to the results of filtering a monaural sound source by such impulse responses, the perceived sound image is usually faithfully localized in space [24]. It does not follow that other listeners will experience a well localized image. While the gross features of the HRIR's are the same for everybody, there is great person-to-person variation in the detailed features. Although some researchers think that the seriousness of the interperson variability has been exaggerated [19], the anecdotal experience of many people having difficulty hearing spatial effects with binaural recordings has been confirmed by psychoacoustic testing [22], and it is generally accepted that customized HRIR's are required for effective localization. This make parameterized HRTF models all the more desirable.

3 The Spherical Head Model

In theory, it should be possible to calculate the HRTF's by solving the wave equation, subject to the boundary conditions presented by the torso, head and pinnae. Needless to say, this is analytically beyond reach and computationally formidable. Over 100 years ago, Lord Rayleigh obtained a remarkably good low-frequency approximation by obtaining an exact solution to the simpler problem of the diffraction of a acoustic plane wave by a rigid sphere [17]. Among other things, his solution showed that (a) the "head-shadow" IID effects begin to appear around 1 kHz, and (b) the ITD varies sinusoidally with azimuth and gradually though rather complexly with frequency[12].

Even though Rayleigh's model was simple, his solution was not, and various approximations have been proposed. A typical approximate model is a cascade of an azimuth dependent "head-shadow" filter and an azimuth-dependent propagation delay. A simple ray-tracing model leads to the ITD formula $\Delta T = 2(a/c)\sin\theta$, where a is the head radius and c is the speed of sound, although the formula $\Delta T = (a/c)(\theta + \sin\theta)$ gives a better fit to experimental data [16]. A computational solution has been developed for Rayleigh's "head shadow" results [2]. However, we have been able to get a good low-frequency fit to Rayleigh's solution with the following simpler model:

$$H_R(\omega, \theta) = \frac{1 + j2\alpha\omega\tau}{1 + j\omega\tau} e^{-j\omega T_R}$$

$$H_L(\omega, \theta) = \frac{1 + j2(1-\alpha)\omega\tau}{1 + j\omega\tau} e^{-j\omega T_L}$$

where $\alpha = \frac{1}{2}(1 + \sin\theta), \tau = \frac{1}{2}(a/c), T_R = (1-\alpha)\tau$ and $T_L = \alpha\tau$. This model fits Rayleigh's solution well for frequencies below 2 kHz. When one listens to synthetic binaural sounds produced by this filter, the apparent location moves smoothly from the left ear to the right ear as θ is varied from $-90°$ to $90°$ [5]. However, this model does not provide any elevation dependence, and the apparent location is not externalized, but appears to be inside the head.

4 Pinna-Echo Models

About 25 years ago, Batteau showed that the pinnae play a critical role in determining elevation, and he conjectured that two major ridges in the outer ear act like reflecting surfaces, producing multipath echoes whose timing gave the cues for elevation [1]. This leads to a pinna transfer function of the form $H_p = (1 + \rho_1 e^{-j\omega\tau_1} + \rho_2 e^{-j\omega\tau_2})/(1 + \rho_1 + \rho_2)$, where both the reflection coefficients ρ_i and the echo delays τ_i can vary with azimuth and elevation. Although pinna-echo models have been criticized as oversimplifying the complicated diffraction process [18, 3], their frequency response curves exhibit comb-filter notches that resemble experimental data [4], and psychoacoustic tests indicate a strong correlation between the notch frequencies and the perception of elevation [20].

5 General Structural Models

Perhaps the most ambitious HRTF model to date is due to Genuit, who has proposed a combination of cascade and multipath filter sections to account for static effects (ear-canal resonance, eardrum impedance), azimuth-dependent effects (interaural delay, head diffraction), and elevation-dependent effects (a second pinna reflection, shoulder reflection) [9, 10]. The attractive feature of Genuit's model is that it not only separates the azimuth-dependent and elevation-dependent components, but it also promises to relate the filter parameters to easily measurable body dimensions, such as the width of the shoulders and the depth of the cavum conchae.

While structural models are conceptually very appealing, they have not yet been shown to be effective in creating synthetic binaural sounds. An informal evaluation of a real-time implementation of a simplified structural model gave azimuth effects that were no better than the spherical head model, elevation effects that were weak at best, and no sense of externalization [5]. It is entirely possible that much better results can be obtained if the model parameters are carefully chosen, but no effective procedure for doing this is known.

6 Models Based on Series Expansions

Since experimentally measured HRTF's can produce excellent azimuth, elevation, and range ("out-of-head") results, an alternative approach is to obtain models by applying general system identification methods to experimental data. The most popular procedure to date has been to apply a principal components or Karhunen-Loéve expansion to the log-amplitude response [13, 11], although the periodic variation of the HRTF with azimuth and elevation makes a Fourier series expansion of the log-IID more natural [7]. The terms in a series expansion of a log-amplitude spectrum correspond to a cascade model, which is appropriate for head diffraction and ear-canal resonance, but less suitable for echoes and similar multipath phenomena. An alternative is to apply a KL expansion to the complex HRTF itself [6]; although mathematically valid, such an additive expansion seems unnatural for a transfer function.

Given enough terms in the series, all of these expansions can produce excellent approximations and good spatial effects. Unfortunately, they cannot be used without having experimental HRTF data. Thus, they do not provide a direct method for obtaining customized HRTF's.

7 Conclusions

The modeling of HRTF's is a challenging problem in system identification. While azimuth effects are readily modeled, elevation is more difficult, and range or externalization is largely a mystery. Structural models offer the promise of easy customization, but no effective procedure for determining the parameters has been demonstrated. Mathematical expansions produce excellent performance, but no effective procedure for obtaining expansion from body dimensions has been found. A combination of the two approaches seems to offer the greatest promise for an effective solution.

Acknowledgements

This work was supported by the National Science Foundation under NSF Grant No. IRI-9214233. I greatly appreciate the experimental HRTF data that was generously provided to me by Dr. Frederic Wightman of the University of Wisconsin. I also want to express my appreciation to Dick Lyon and Malcolm of Apple Computer, Inc., for their support and encouragement.

References

[1] Batteau, D. W., "The Role of the Pinna in Hu-

man Localization," *Proc. Royal Society London*, Vol. 168 (series B), pp. 158-180 (1967).

[2] Bauck, J. L. and D. H. Cooper, "On Acoustical Specification of Natural Stereo Imaging," in *Proc. 66th Convention Audio Engineering Society* (Los Angeles, CA), May, 1980.

[3] Blauert, J. P., *Spatial Hearing* (MIT Press, Cambridge, MA, 1983).

[4] Butler, R. A. and K. Balendiuk, "Spectral Cues Utilized in the Localization of Sound in the Median Sagittal Plane," *J. Acoust. Soc. Am.*, Vol. 61, pp. 1264-1269, 1977.

[5] Cassaro, T. and M. J. Van Belleghem, "Implementing Time-Variable DSP Filters to Synthesize Binaural Sounds," Technical Report No. 2, NSF Grant No. IRI-9214233, Dept. of Elec. Engr., San Jose State Univ., San Jose, CA (May, 1993).

[6] Chen, J., B. D. Van Veen and K. E. Hecox, "Synthesis of 3D Virtual Auditory Space via a Spatial Feature Extraction and Regularization Model," in *Proc. IEEE Virtual Reality Annual International Symposium (VRAIS93)* (Seattle, WA), 1993, pp. 188-193.

[7] Duda, R. O., "Estimating Azimuth and Elevation from the Interaural Head Related Transfer Function," in *Conference on Binaural and Spatial Hearing* (Dayton, OH, September 9-12, 1993).

[8] Durlach, N. I. et al., "On the Externalization of Auditory Images," *Presence,* Vol. 1, pp. 251-257 (Spring 1992).

[9] Genuit, K., "A Description of the Human Outer Ear Transfer Function by Elements of Communication Theory," *Proc. 12th ICA* (Toronto, 1986).

[10] Genuit, K., "Method and Apparatus for Simulating Outer Ear Free Field Transfer Function," U.S. Patent No. 4672569 (9 June 1987).

[11] Kistler, D. J. and F. L. Wightman, "A Model of Head-Related Transfer Functions Based on Principal Components Analysis and Minimum-Phase Reconstruction," *J. Acoust. Soc. Am.,* Vol. 91, pp. 1637-1647 (March 1992).

[12] Kuhn, G. F., "Acoustics and Measurements Pertaining to Directional Hearing," in *Directional Hearing*, W. A. Yost and G. Gourevitch, Eds., pp. 3-25 (New York, NY: Springer Verlag, 1987).

[13] Martens, W. L., "Principal Components Analysis and Resynthesis of Spectral Cues to Perceived Direction," in *Proc. International Computer Music Conference* pp. 274-281 (1987).

[14] McKinley, R., "Flight Demonstration of a 3-D Audio Display," in *Conference on Binaural and Spatial Hearing* (Dayton, OH, Sept. 9-12, 1993).

[15] Middlebrooks, J. C. and D. M. Green, "Sound Localization by Human Listeners," *Annu. Rev. Psychol.,* Vol. 42, pp. 135-159 (1991).

[16] Mills, A. W., "Auditory Localization," in *Foundations of Modern Auditory Theory, Vol. II* (J. V. Tobias, Ed.), pp. 303-348 (Academic Press, NY, 1972)

[17] Rayleigh, J. W. S., *The Theory of Sound* (Macmillan, London, 1877); second edition republished by Dover Publications, NY, 1945).

[18] Shaw, E. A. G., "Transformations of Sound Pressure Level from the Free Field to the Eardrum in the Horizontal Plane," *J. Acoust. Soc. Am.*, Vol. 56, pp. 1848-1861 (September 1974).

[19] Walczak, N. B., "The Perceptual Significance of Inter-subject Differences in the Directional Transfer Function (DTF)," in *ICASSP 88 (Proc. IEEE Int. Conf. Acoustics Speech and Signal Processing)* (New York, NY), pp. 2520-2523 (April 1988).

[20] Watkins, A. J., "Psychoacoustical Aspects of Synthesized Vertical Locale Cues," *J. Acoust. Soc. Am.,* Vol. 63: pp. 1152-1165 (April, 1978).

[21] Wenzel, E. M., "Localization on Virtual Acoustic Displays," *Presence,* Vol. 1, pp. 80-107 (Winter 1992).

[22] Wenzel, E. M. et al., "Localization Using Nonindividualized Head-Related Transfer Functions," J. Acoust. Soc. Am., Vol. 94, pp. 111-123 (July 1993).

[23] Wightman, F. L., D. J. Kistler and M. E. Perkins, "A New Approach to the Study of Human Sound Localization," in W. A. Yost and G. Gourevitch, eds., *Directional Hearing*, pp. 26-48 (Springer Verlag, New York, 1987).

[24] Wightman, F. L. and D. J. Kistler, "Headphone Simulation of Free-Field Listening. II: Psychophysical Validation," *J. Acoust. Soc. Am.,* Vol. 85, pp. 868-878 (February 1989).

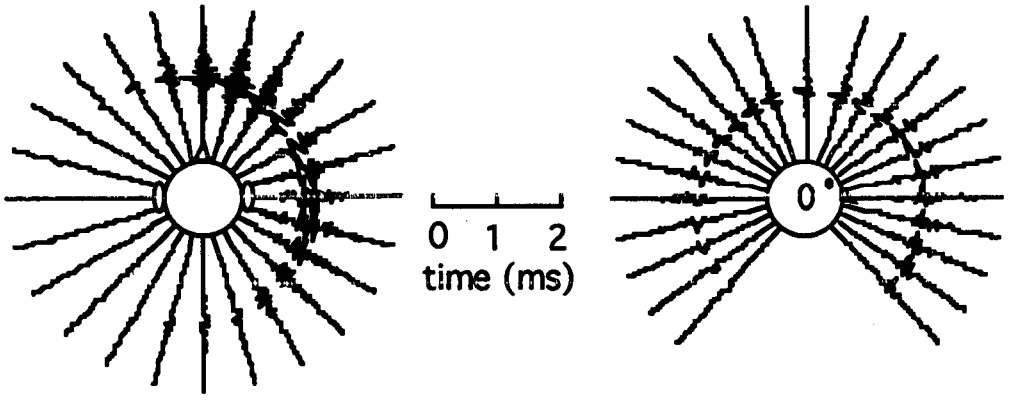

Fig. 1 Head related impulse response for right ear for SLV data from U. Wisconsin.

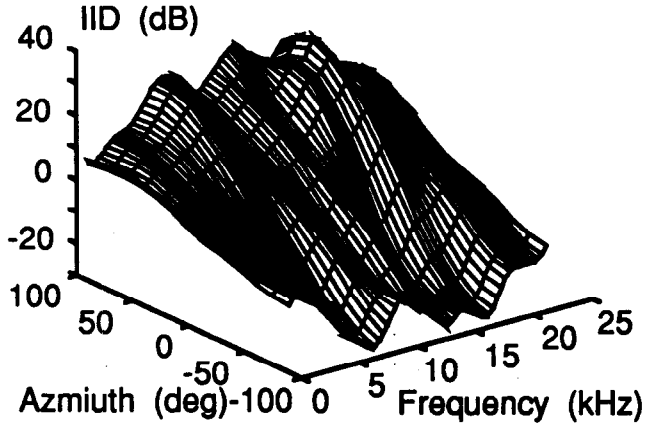

Fig. 2 Smoothed IID in the horizontal plane

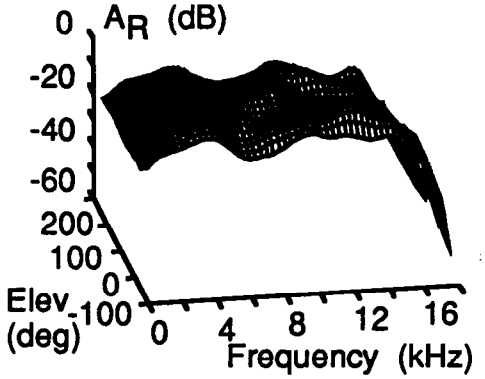

Fig. 3 Smoothed median plane amplitude response

Fig. 4 First Fourier sine coefficient in the expansion of the IID