Subject Dependent Transfer Functions in Spatial Hearing[1]

V. Ralph Algazi, University of California, Davis
Pierre L. Divenyi, Speech and Hearing Research Facility, V.A. Martinez
Richard O. Duda, San Jose State University

Abstract—Head Related Transfer Functions (HRTF's) characterize the transformation of pressure waves from the sound source to the listener's eardrums. They are of central interest to binaural hearing and to sound localization. The HRTF is a function of the azimuth and elevation of the sound source, and may vary from subject to subject. The functional dependence of the HRTF on azimuth and elevation is first described, and models that provide good spatial localization in azimuth are reviewed. We then discuss recent work on the more complete subject dependent models that are needed for good localization and discrimination in elevation.

1. Introduction

There has been a recent increase in interest in the synthesis of three-dimensional spatial sound. Synthesized spatial sound is important to virtual reality systems, and to the entertainment industry. Better understanding of spatial hearing is also important to advances in hearing aids, to extending the workplace by teleconferencing, or to new human/computer interfaces.

The key to generating spatial sound is the so-called *Head-Related Transfer Function* (HRTF). The HRTF captures the position-dependent changes that occur when a sound wave propagates from a sound source to the listener's ear drum. These spectral changes are due to diffraction of the sound wave by the torso, head, and outer ears or pinnae, and their character depends on the azimuth, elevation, and range from the listener to the source(1). The HRTF, which completely characterizes the acoustic information available for sound localization, will vary from person to person. Inter-subject variations may result in significant localization errors (primarily front/back reversals and large elevation errors) when one person hears the source through another person's HRTF's (2). Thus, individualized HRTF's are needed to obtain a faithful perception of spatial location. Individual HRTF's can be measured experimentally. The resulting HRTF's are stored as finite-impulse-response tables, indexed by a large number of different azimuth and elevation values (3). If head motion is taken into account, playback requires rapid interpolation between the entries in these tables. This "brute-force" approach has two serious drawbacks: (a) it provides no insight into the factors that control spatial hearing, and (b) it requires complex and expensive equipment for applications. Thus, there is a great current interest in devising mathematical parametric models of the HRTF that can be individually customized. We present, first, the mathematical and physical basis for such models.

II. Background on Measured Head-Related Transfer Functions

The HRTF relates the time-varying free-field sound-pressure $x(t)$ of a sound source to the sound-pressures $x_L(t)$ and $x_R(t)$ at the left and right ear drum (1). The time domain convolution of the source signal with the head-related impulse responses $h_L(t)$ and $h_R(t)$ for the left and right ear, respectively, becomes in the frequency domain the product of their Fourier transforms: $X_L(\omega) = H_L(\omega)X(\omega)$, $X_R(\omega) = H_R(\omega)X(\omega)$. If the source location is specified with a head-centered spherical coordinate system, H_L and H_R vary with the angular frequency ω, and with the azimuth θ, elevation ϕ, and range r. Modeling a HRTF requires finding a natural and simple representation of the two functions $H_L(\omega,\theta,\phi,r)$ and $H_R(\omega,\theta,\phi,r)$. Fig. 1 shows some of the head-related impulse responses measured for KEMAR, an acoustic manikin widely used in hearing-aid research(4).

The left half of the figure shows how the right-ear impulse response $h_R(t)$ varies as the sound source is moved around the head in a circle in the horizontal plane; in the plot, time increases along the radial direction. This diagram clearly shows that the response is both strongest and occurs earliest when the source is incident on the right ear. The right half of the figure shows how $h_R(t)$ varies as the source is moved around the head in a circle in the mid-sagittal plane. In the frequency domain the log-magnitude response of one ear in the horizontal plane displays a basically sinusoidal variation with azimuth. Such a spectral profile exhibits a peak around the 3-kHz ear-canal resonant frequency, and various dips or notches at higher frequencies that are attributed to pinna diffraction. The response falls off at frequencies above the ear-canal resonance as the result of the "head shadow."

The Interaural Level Difference (ILD) and Interaural Time Difference (ITD) are the primary cues for azimuth. If we denote the speed of sound by c and we approximate the head by a sphere of radius a, a simple ray-tracing argument (6) leads to the experimentally verified formula interaural time delay

$$\Delta T \approx (a/c)(\theta + \sin\theta)$$

A study of the ILD shows that, to a first degree of approximation, its shape as a function of elevation is independent of azimuth, and that the overall scale of the ILD being proportional to the sine of the azimuth (10). Thus, one can factor the ILD (basically the log-magnitude of the ratio H_R / H_L) as a function of frequency and elevation times a function of azimuth.

[1] Work supported by NSF Grant IRI-96-19339

0-7803-3694-1/97/$10.00 © 1997 IEEE

The response as a function of elevation reveal that the frequencies of these "pinna notches" are strongly dependent on elevation. These spectral shape features provide the primary monaural cues for elevation.

III. HRTF Models

Two basically different methods have been employed to model HRTF's — Principal components analysis (PCA) and structural modeling.

A. PCA and Series Expansions

PCA is a classical statistical technique for finding a low-dimensional representation of high-dimensional data. In HRTF applications, PCA is usually applied to the log-magnitude response in the frequency domain (7,8,9). Since the HRTF is essentially a minimum-phase function, the Hilbert transformation can be used to recover the HRTF from the magnitude alone — provided that the proper interaural time difference is maintained(7). The Fourier series expansions, that exploit the fact that the HRTF's are periodic in θ and ϕ also provide accurate and efficient representations of the ILD (10). Since PCA and Fourier expansions are additive, their use for the log-magnitude response leads to a HRTF representation as a cascade or product filter structure. Applying principal components analysis directly to the complex HRTF transfer function leads to a parallel filter structure(11). Physically, a cascade representation may be appropriate for the ear-canal resonance, and a parallel representation for pinna and shoulder "echoes" (12). Using system identification methods to fit a rational function to the HRTF the azimuth dependence of the interaural transfer function $H(\omega,\theta,\phi) = H_L(\omega,\theta,\phi) / H_R(\omega,\theta,\phi)$ can be modeled well by a fourth-order all-pole model (13, 14). However, many zeros are needed to account for "pinna echoes," and this all-pole model does not accurately represent elevation dependence.

B. Structural Models

They are based on the physical processes of wave propagation and diffraction, and were pioneered by Shaw (15) and Genuit (16). Genuit's basic HRTF model, shown in Fig. 2, uses a cascade structure to separate position dependent and position independent components. The azimuth controls the time delay and the head-shadow effects. The elevation is used to control pinna echoes. This approach is attractive in that there is a physical function for each component of the model. The components contain a relatively small number of parameters, that may be related to anthropomorphic measurements and adjusted to match the behavior of the HRTF.

IV. Development of Customized HRTF Models

In order to develop physically-based parametric models customized to particular individuals, we need to relate model parameters and experimental HRTF data.

A. Experimental data acquisition

The experimental acquisition of HRTF data is complex and time consuming. For each subject, the impulse response is measured for a dense grid of azimuths and elevations. In our time domain Snapshot(TM) system, manufactured by Crystal River Engineering, Golay codes are used to suppress noise, and the impulse response duration is limited to 3-ms. This short duration allows the exclusion of echoes, so that measurements can be performed in normal, non-anechoic rooms. We are assembling an HRTF database from a representative set of human subjects that will contain, in addition to the HRTF acoustic measurements, anthropometric measurements of appropriate torso, head and pinna dimensions of the subjects, as well as photographs of the outer ears.

B Model Development

Our structural model of the HRTF omits the effects of the shoulders and upper body in the basic model (Fig. 2). Our model has four components: a time delay, and filters for the head , the pinna filter, and the ear-canal.

The ear-canal filter models the resonances of a tube that is terminated with the ear-drum impedance. This filter is independent of the location of the sound source. A basic model of the ear canal accounts for the lowest resonant frequency. Such a model is specified by three parameters that control the transfer function: a resonant frequency ω_c (which, for KEMAR, lies between 3 and 4 kHz), a damping coefficient ζ, and a gain. The model can be refined by adding one more band-pass section to account for the next resonant frequency.

For the time delay, T_d, the first-order dependence is on the azimuth angle θ (see Fig. 1). If we select the time origin so that there is no time delay for 90° incidence, a ray-tracing argument suggests the following formula for the right ear:

$$T_d = \begin{cases} T_0(1-\sin\theta), & 0 \le \theta < \frac{\pi}{2} \\ T_0(1-\theta), & -\frac{\pi}{2} \le \theta < 0 \end{cases}$$

This formula is based on a spherical-head approximation in which $T_0 = a / c$, where a is the head radius and c is the speed of sound. T_0 can also be viewed as a free parameter adjusted to fit the data. This first-order approximation can be refined to match experimental data showing that T_0 is frequency dependent, with the group delay at low frequencies being approximately 50% greater than the group delay at high frequencies(5).

The head filter accounts for the "head shadow" resulting from sound waves diffracting around the head. For a spherical-head approximation, there is a well-known theoretical solution due to Lord Rayleigh for the diffraction of a plane wave by a sphere (20). This solution is well approximated by a one-pole, one-zero filter model in which the location of the zero is azimuth dependent, and the cutoff frequency ω_h is the only parameter.

In the pinna model, elevation dependence is a first-order concern. Pinna effects are complex, and occur at frequencies that are sufficiently high to make experimental measurements difficult (16). However, Watkins (17) has shown that a simple "two-echo" model captures much of the vertical localization information. This model contains four parameters, an elevation-dependent reflection coefficient $\rho_v(\phi)$, an elevation-dependent time delay $\tau_v(\phi)$, an azimuth-dependent reflection coefficient $\rho_a(\phi)$, and an azimuth-dependent time delay $\tau_a(\phi)$. Students at the SJSU DSP Laboratory have made a DSP implementation of such a model (18). Their implementation uses an azimuth-dependent, single-pole head-shadow filter, an azimuth-dependent time delay, two pinna reflections, and an azimuth and elevation dependent shoulder reflection component. Values for the parameters were heuristically estimated. Informal listening tests confirmed that the azimuth effects were quite good, but that elevation effects were unconvincing. Recently, one of us and another SJSU student (19) obtained better elevation results with a high resolution time-domain estimation of delays due to the pinna.

Thus, HRTFs can be modeled in terms of modules that contain parameters with values determined by fitting the models to data. We also plan to extract these model parameters directly from anthropometric measurements obtained by using a 3D digitization system, or by stereo analysis of digitized photographs.

Once a complete set of model parameters has been determined for a large number of HRTF's acquired experimentally, a psychophysical validation will test the model HRTFs in localization-identification experiments.

V. Examples of Elevation HRTF Data

We illustrate features of the elevation data and differences from subject to subject. The data was collected at a 50° azimuth, for elevations 5 degrees apart, and 72 impulse responses were collected for each ear. Fig. 3 shows two elevation data sets as images, with amplitude mapped to gray. Time (0.5 ms total duration) runs vertically . Observe the delays that become preeminent or fade as the elevation varies. Note that the response at the distant ear shows little elevation dependence, except for a broad bright spot at elevations where all sound propagating around the head may arrive simultaneously. Note also significant differences between subjects in the number, location and amplitude of the delays. The importance of the individual differences to effective models of spatial hearing remain to be determined.

VI. Discussion and Conclusions

While the technology exists to measure the HRTF for any individual, we currently lack effective ways to model and implement these complicated response functions. Simple, physically-based models that expose aspects of the directional response critical to spatial hearing are needed for a basic physical understanding . Further, a simple, easily implemented model of the HRTF opens opportunities for applications. By implementing models for the directional differences in the sound reaching the two ears, synthetic binaural sounds can be generated efficiently. This would also provide simpler procedure for customizing sound sources to individual users. Adaptation of the man-machine environment to individual characteristics is important for the increasing use of hearing in man-machine systems.

References

1. Blauert, J. (1983). *Spatial Hearing* (MIT Press, Cambridge, MA).
2. Wenzel, E. M., M. Arruda, D. J. Kistler and F. L. Wightman (1993). "Localization using non individualized head-related transfer functions," *J. Acoust. Soc. Am.*, Vol. 94, pp. 111-123.
3. Foster, S. H. and Wenzel, E. M. (1991). "Virtual acoustic environments: The Convolvotron," presented at "Tomorrow's Realities Gallery," *SIGGRAPH 91* (18th ACM Conference on Computer Graphics and Interactive Techniques, Las Vegas, NV).
4. Duda, R. O. (1991). "Short-Time Measurement of the KEMAR Head-Related Transfer Function," Technical Report, Advanced Technology Group, Apple Computer, Inc.
5. uhn, G. F. (1987). "Physical acoustics and measurements pertaining to directional hearing," in W. A. Yost and G. Gourevitch, Eds., *Directional Hearing*, pp. 3-25 (Springer Verlag, NY).
6. Mills, A. W. (1972). "Auditory localization," in J. V. Tobias, Ed., *Foundations of Modern Auditory Theory, Vol. II*, pp. 303-348 (Academic Press, NY).
7. Martens, W. L. (1987). "Principal components analysis and resynthesis of spectral cues to perceived direction," *Proceedings of the International Computer Music Conference*, J. Beauchamp, Ed. (International Computer Music Association, San Francisco, CA), pp. 274-281.
8. Kistler, D. J. and F. L. Wightman, (1992). "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," *J. Acous. Soc. Am.*, Vol. 91, pp. 1637-1647.
9. Middlebrooks, J. C., and D. M. Green (1992). "Observations on a principal components analysis of head-related transfer functions", *J. Acoust. Soc. Am.*, Vol. 92, pp. 597-599.
10. Duda, R. O. (1993). "Estimating azimuth and elevation from the interaural head-related transfer function," presented at the Conference on Binaural and Spatial Hearing (Dayton, OH). To appear in R. Gilkey, and T. Anderson, Eds., *Binaural and Spatial Hearing* (Lawrence Earlbaum Associates, Hillsdale, NJ)..
11. Chen, J., B. D. Van Veen and K. E. Hecox (1993). "Synthesis of 3D virtual auditory space via a spatial feature extraction and regularization model," in *VRAIS 93* (Proc. IEEE Virtual Reality Annual International Symposium) (Seattle, WA), pp. 188-193.
12. Duda, R. O. (1993). "Modeling head related transfer functions," *Proc. Twenty-Seventh Asilomar Conference on Signals, Systems & Computers* (Asilomar, CA).
13. Ljeung, R. (1992). *System Identification for the User* (Addison-Wesley, Reading, MA).
14. Nguyen, H. T., "Use of System Identification Methods to Estimate the Azimuth and Elevation of a Sound Source," Technical Report No. 7, NSF Grant No. IRI 92-14233 (December 1993).
15. Shaw, E. A. G. (1974b) "Wave properties of the human ear and various physical models," *J. Acoust. Soc. Am., Vol.* 56, S3(A).
16. Genuit, K. (1986). "A description of the human outer-ear transfer function by elements of communication theory," paper B6-8, *Proc. 12th International Congress on Acoustics* (Toronto, Canada).
17. Watkins, A. J. (1978). "Psychoacoustical Aspects of Synthesized Vertical Locale Cues," *J. Acoust. Soc. Am.*, Vol. 63: pp. 1152-1165.
18. Cassaro, T., and M. J. Van Belleghem (1993). "Implementing Time-Variable DSP Filters to Synthesize Binaural Sounds," Tech. Rpt. No. 2, NSF Grant No. IRI-9214233, Dept. of Elec. Engr., SJSU San Jose, CA.
19. Brown, C. P. And Duda, R. O. (1997) "an Efficient HRTF Model for 3-D Sound, " Proc. 1997 Workshop on Applications of Digital signal Processing to Audio and Acoustics (Mohonk, NY)
20. Morse, P. M., and K. U. Ingard (1968). *Theoretical Acoustics* (McGraw-Hill, New York).
21. Bauck, J. L., and D. H. Cooper (1980). "On acoustical specification of natural stereo imaging," Proc. *66th Convention of the Audio Engineering Society* (Los Angeles, CA).

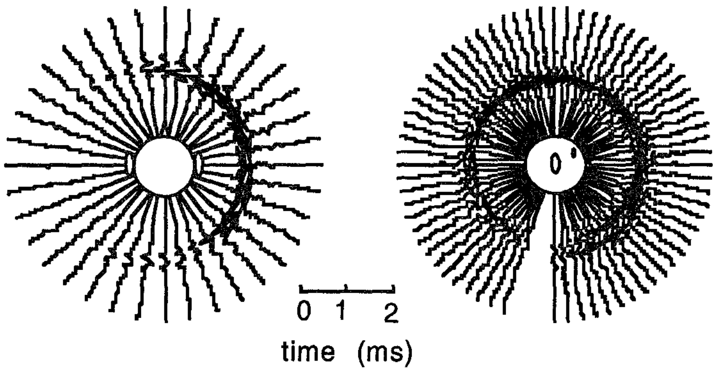

Fig. 1. The KEMAR Right-Ear Head-Related Impulse Response in the Horizontal and the Mid-Sagittal Planes

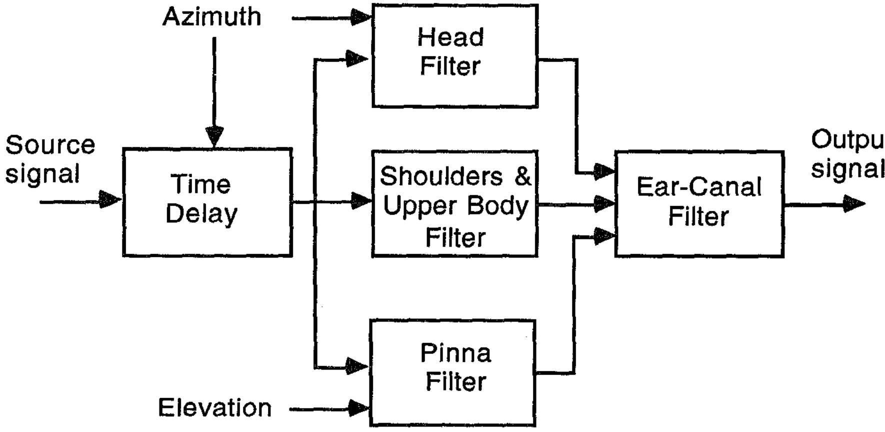

Fig. 2. A Basic Structural HRTF Model

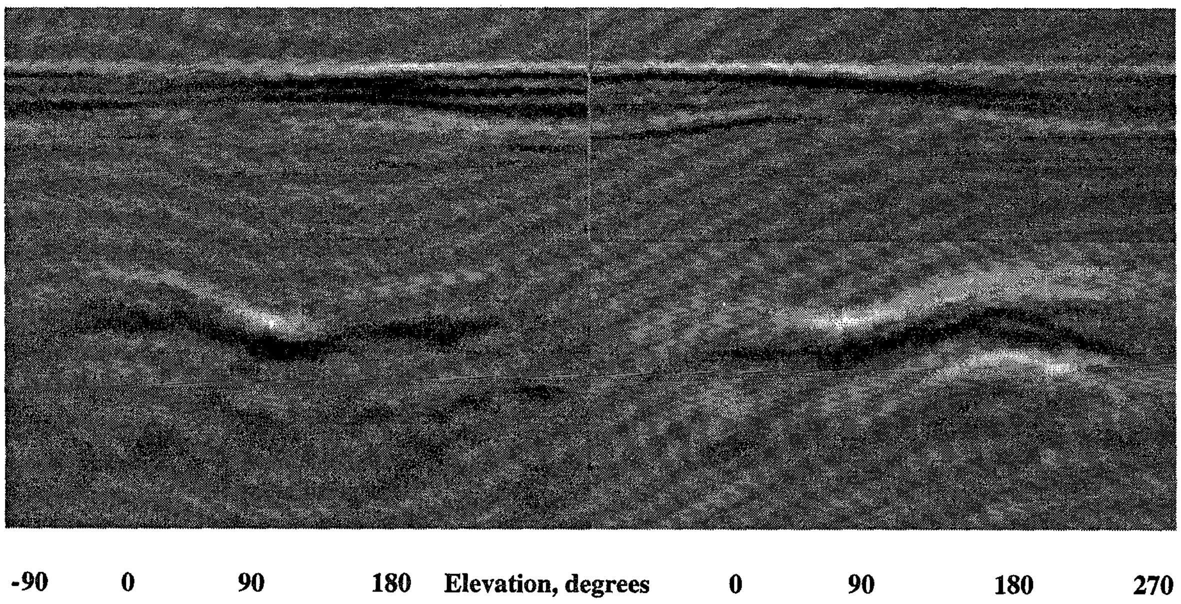

Fig. 3. Two individual Head-Related Impulse Responses : Top: Proximal Ear, Bottom: Distant Ear.