

UC Davis

IDAV Publications

Title

An Adaptable Ellipsoidal Head Model for the Interaural Time Difference

Permalink

<https://escholarship.org/uc/item/8jk3n8tc>

Authors

Duda, Richard O
Avendano, Carlos
Algazi, Ralph

Publication Date

1999

Peer reviewed

AN ADAPTABLE ELLIPSOIDAL HEAD MODEL FOR THE INTERAURAL TIME DIFFERENCE

Richard O. Duda

Department of Electrical Engineering
San Jose State University
San Jose, CA 95192
rod@duda.org

Carlos Avendano and V. R. Algazi

CIPIC
University of California at Davis
Davis, CA 95616
carlos@ece.ucdavis.edu, algazi@ece.ucdavis.edu

ABSTRACT

Experimentally measured head-related transfer functions reveal that the interaural time delay varies from person to person. Furthermore, it is not constant around a cone of confusion, but can vary by as much as 18% of the maximum interaural delay. The major sources for this variation are shown to be the shape of the head and the displacement of the ears from the center of the head. A simple ellipsoidal head model is presented that can accurately account for this ITD variation and can be adapted to individual listeners.

1. INTRODUCTION

Models of head-related transfer functions (HRTF's) are becoming widely used in spatial audio systems [1, 2]. We are currently developing structural HRTF models, whose parameters are directly related to the size and shape of the listener's torso, head and outer ears [3]. This enables the customization of a generic model to an individualized HRTF through simple anthropometric measurements.

The Interaural Time Difference (ITD) provides the major cue for azimuth localization, and is a core part of any HRTF model. If the head is approximated by a sphere of radius a , the ITD for an infinitely distant source can be computed by a simple formula due to Woodworth [4]:

$$\text{ITD} = \frac{a}{c}(\theta + \sin \theta) \quad , \quad (1)$$

where θ is the azimuth angle and c is the speed of sound. Derived from a simple ray-tracing argument, this formula is restricted to angular frequencies greater than a/c , and corresponds to the difference in first arrival times. However, it is remarkably close to the exact theoretical solution,

This work was supported by the National Science Foundation under Grant No. IRI-9619339. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation.

even for a source that is quite near the sphere [5]. Furthermore, customization requires the measurement of only one quantity, the average head radius.

It follows from (1) that a surface of constant ITD is a cone of revolution about the interaural axis. Woodworth called this surface the "cone of confusion." For an ideal spherical head, neither the ITD nor the ILD (Interaural Level Difference) changes as a point is moved around a cone of confusion. Since the ITD and the ILD are primary cues for source localization, this provides a simple explanation of the fact that many localization errors (including front/back errors and up/down errors) result in mislocation on the cone of confusion.

However, a sphere provides only a first approximation to a real human head. We will present experimental results that show that in reality the ITD varies around a cone of confusion, so that the ITD is a function of elevation as well as azimuth.¹ This elevation dependence arises because, although the distance from the source to the ipsilateral ear (the ear that is visible from the source) is constant, the length of the shortest path from the source to the contralateral ear (the ear that is not visible from the source) changes with elevation. These changes are due both to the non-spherical shape of the head and to the fact that the ears are not positioned across a diameter, but are displaced behind and below the center of the head. We will show how a simple ellipsoidal model with offset ears yields the proper variation of ITD with both azimuth and elevation.

2. HRTF DATA

The ITD is directly revealed in experimentally measured impulse response data. Fig. 1 shows an image representa-

¹We employ an interaural-polar coordinate system in which the azimuth angle is measured from the source to the vertical median plane, and the elevation angle is measured up from the horizontal plane to the projection of the source into the median plane. With this coordinate system, a cone of confusion is a cone of constant azimuth.

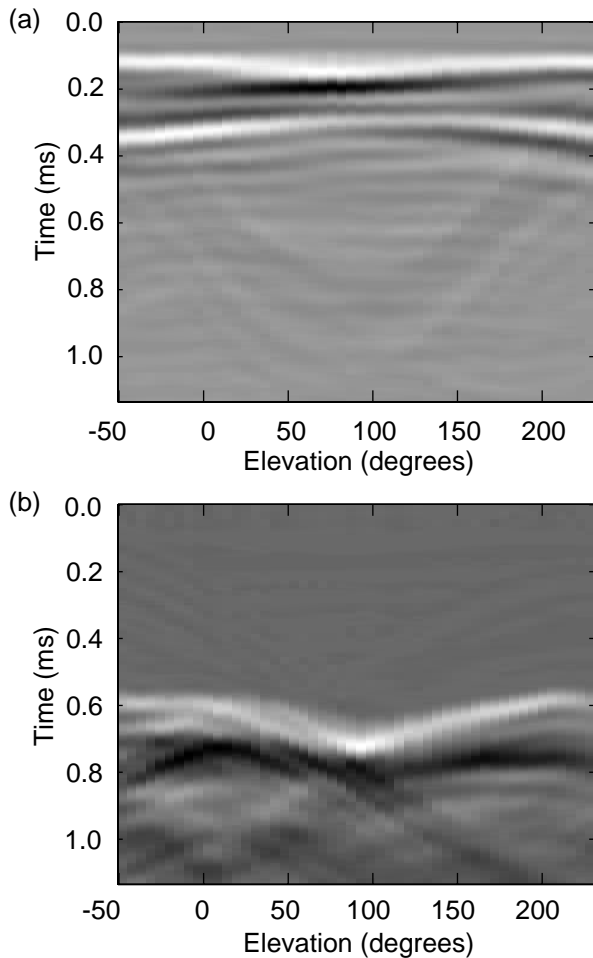


Figure 1: Elevation dependence of an experimentally measured head-related impulse response for (a) the ipsilateral ear and (b) the contralateral ear. In both cases, the azimuth is -55° . Notice the variation in the arrival time for the contralateral ear, and the bright spot around 90° elevation.

tion of the elevation dependence of the ipsilateral and contralateral impulse responses for a subject (CA) measured at an azimuth of -55° . Each column in these images corresponds to an impulse response at a fixed elevation, where bright areas are positive and dark areas are negative. The bright “ridge” across the top of each image identifies the arrival of the pulse.

Because the distance from the source to the ipsilateral ear does not change with elevation, this ridge is basically horizontal in the upper image. The downward displacement of the initial ridge in the lower image is the ITD. Clearly, the ITD is not constant, but varies by about $120 \mu\text{s}$, which is almost 18% of the maximum ITD and corresponds to a shift of approximately 15° in azimuth. When the subject for these measurements listens to the results of convolving

a sound source with these impulse responses, the sound image appears to move smoothly in a vertical plane parallel to the median plane as the elevation is slowly changed from -50° to 230° . However, if the responses of the contralateral ear are time aligned and then delayed by a constant ITD, the sound image is not confined to a vertical plane, but appears to be significantly displaced to the right and left as the source moves around the cone. In short, for accurate placement of the virtual source, Woodworth’s formula is inadequate, and the formula used to compute the ITD must be a function of elevation as well as azimuth.

A closer examination of the contralateral image reveals a “bright spot” near 90° elevation. This can be explained using an elementary multipath interpretation. At -55° azimuth, the shortest path from the source to the contralateral ear is around the front of the head if the source is in front, and around the back of the head if the source is in back. Normally, one of these two paths dominates. However, near 90° elevation, these paths have the same length, and the signals arrive in phase to produce the bright spot. Indeed, with just a little imagination, one can discern the separate components around the front and around the back in the contralateral image (Fig. 1b).

3. THE ELLIPSOIDAL HEAD MODEL

Although this shortest-path argument is an oversimplification and cannot explain all the behavior of what is clearly a complex wave propagation phenomenon, it is similar to Woodworth’s original argument, and it leads to a surprisingly accurate model for computing the ITD. In particular, the variation in the ITD with elevation is easily explained by observing that if the shape of the head is not spherical, and/or if the ears are not located across a diameter, then the length of the shortest path from the source to the contralateral ear will vary with elevation.

An ellipsoid is an obvious choice for a head model that exhibits elevation as well as azimuth dependence.² To express the model analytically, let $\vec{s} = [s_1, s_2, s_3]^T$ be a vector from the origin to the sound source, and let $\vec{e} = [e_1, e_2, e_3]^T$ be a similar vector to the contralateral ear.³ Define the ellipsoid by the equation

$$(\vec{x} - \vec{x}_0)^T A (\vec{x} - \vec{x}_0) = 1 \quad , \quad (2)$$

where $\vec{x} = [x_1, x_2, x_3]^T$ is a vector to a point on the ellipsoid, \vec{x}_0 is the center of the ellipsoid, and A is a 3×3 positive definite, symmetric matrix. Typically, $A =$

²To the best of our knowledge, the only analytic, non-spherical head models that have been investigated are prolate spheroids, and this prior work did not address the ITD [6, 7].

³We assume that all vectors are column vectors, and we denote the transpose of \vec{x} by \vec{x}^T .

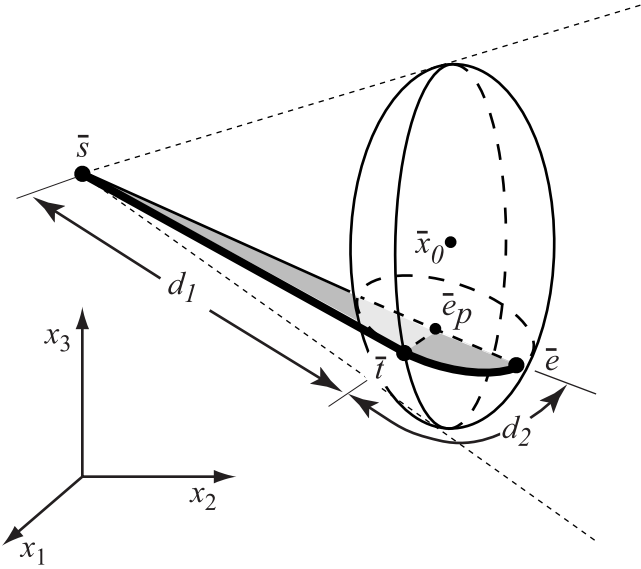


Figure 2: The ellipsoidal head model. Sound travels from the source \bar{s} a distance d_1 to a tangent point \bar{t} and around the ellipsoid a distance d_2 to the ear at \bar{e} .

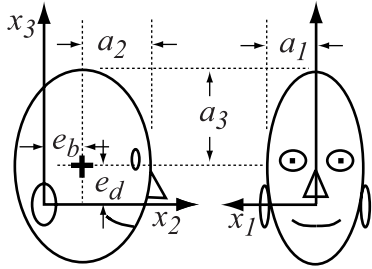


Figure 3: Head dimensions and ear location.

$\text{diag}[1/a_1^2, 1/a_2^2, 1/a_3^2]$, where a_k is the length of the principal axis along the k^{th} coordinate direction. Because the ears are on the interaural axis, $e_2 = e_3 = 0$. If e_b is the displacement of the ear toward the back of the head and e_d is the displacement down, $\bar{x}_0 = [0, e_b, e_d]^T$, and $e_1 = \pm a_1 \sqrt{1 - (e_b/a_2)^2 - (e_d/a_3)^2}$. Thus, as Fig. 3 illustrates, the five parameters a_1, a_2, a_3, e_b and e_d define the head model.

There is no closed form, analytic expression for the shortest distance from \bar{s} to \bar{e} . Dynamic programming can be used to compute the solution, but this is computationally intensive. The following heuristic procedure is relatively fast and yields a close approximation:

1. Find the plane P_t defined by the cone of rays from \bar{s} that are tangent to the ellipsoid. The intersection of P_t with the ellipsoid defines the tangent ellipse.

2. Find the point \bar{e}_p where the line from \bar{s} to \bar{e} pierces P_t .
3. Find \bar{t} by assuming that it is the point on the tangent ellipse that is closest to \bar{e}_p .
4. Find the plane P_s passing through \bar{s} , \bar{t} and \bar{e} .
5. Compute d_1 as the straight-line distance from \bar{s} to \bar{t} , and d_2 as the arc length from \bar{t} to \bar{e} of the ellipse that results from intersecting the ellipsoid with P_s .

It turns out that the equation for P_t is $\bar{n}^T(\bar{x} - \bar{x}_0) = 1$, where $\bar{n} = A(\bar{s} - \bar{x}_0)$, and that $\bar{e}_p = \bar{e} - \lambda(\bar{s} - \bar{e})$, where $\lambda = [1 - \bar{n}^T(\bar{e} - \bar{x}_0)]\bar{n}^T(\bar{s} - \bar{e})$. Finding \bar{t} requires minimizing $\|\bar{x} - \bar{e}_p\|^2$ subject to the constraints $(\bar{x} - \bar{x}_0)^T A(\bar{x} - \bar{x}_0) = 1$ and $\bar{n}^T(\bar{x} - \bar{x}_0) = 1$. Although an analytical solution exists, it is sufficiently complicated that we solved this problem by numerical computation. Once \bar{t} is found, the remaining steps are straightforward.

4. EXPERIMENTAL RESULTS

To apply this procedure, we need to have numerical values for the five parameters. One approach is to imagine enclosing the head in a rectangular box, to use the dimensions of the box for a_1, a_2 and a_3 , and to measure e_b and e_d from the center of the box. While this produces results that are qualitatively better than the spherical-model formula (1), we discovered that the resulting calculated ITD values tended to underestimate the measured ITD. Significantly better results were obtained when the head dimensions were increased by 5 to 10%.

An LMS procedure was used to optimize the model parameter values. The ITD computed by this method is shown in Fig. 4 as the curved white line superimposed on the image for the contralateral ear. Since the azimuth is constant, the spherical model yields the straight black line. The dotted points in Fig. 4 show the empirical ITD, defined as the difference in first arrival times for the two ears, where the first arrival time is the time at which the impulse response first exceeds 15% of the maximum absolute value of the response. Clearly, the ellipsoidal model, which has an average absolute error of 10 μs , fits the data better than the spherical model, which has an average absolute error of 30 μs .⁴

This is typical of the results obtained at different azimuth angles and with different subjects. Fig. 5 shows the results obtained for the same subject at 25 azimuths and 50 elevations. In this case, the average absolute ITD error is 15 μs for the ellipsoidal model versus 22 μs for the spherical model, which is a relatively small difference. However,

⁴The time between samples at the standard 44.1-kHz sampling rate is 22.7 μs . While a 1.3-sample error might seem small, it introduces a clearly audible shift in location.

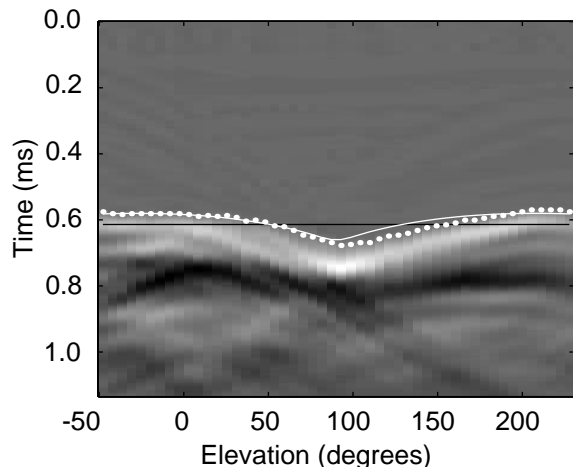


Figure 4: ITD values for -55° azimuth. The dotted curve shows the 15%-rise-time ITD. The curved line is computed from the ellipsoidal model, and the straight line is computed from the spherical model.

because the ITD for the spherical model does not depend on elevation, the spherical model introduces a systematic error that is revealed by the horizontal streaks in the data. Clearly, the ellipsoidal model removes this systematic error, and provides generally superior results at large azimuths.

5. CONCLUSIONS

A simple ellipsoidal model of the head that includes downward and backward displacements of the ears can provide accurate values for the ITD for an HRTF model. By contrast with a spherical head model, the resulting ITD's vary around a cone of confusion, and provide the proper time delays needed for precise localization. The correction is most important at larger azimuths, where the elevation dependence of the ITD is strongest.

The ellipsoidal head model requires five parameter values, all of which are related to anthropometric measurements of the listener's head. Although experiments with more subjects will be needed to determine the exact nature of this relationship, our results demonstrate the feasibility of using geometrical models in an important part of a customizable HRTF model.

6. REFERENCES

[1] Begault, D. R., *3-D Sound for Virtual Reality and Multimedia*. AP Professional, Chestnut Hill, MA, 1994.

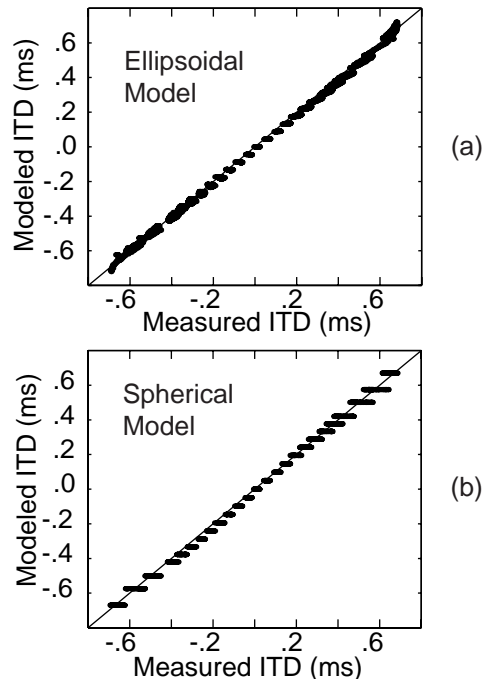


Figure 5: Comparison of measured and computed ITD values for 1250 points distributed over the sphere for (a) the ellipsoidal model and (b) the spherical model.

[2] Carlile, S., *Virtual Auditory Space: Generation and Applications*. R. G. Landes, Austin, TX, 1996.

[3] Brown, C. P. and Duda, R. O., "A Structural Model for Binaural Sound Synthesis," *IEEE Trans. on Speech and Audio Processing*, vol. 6, pp. 476-488, Sept. 1998.

[4] Woodworth, R. S. and Schlosberg, G., *Experimental Psychology*, pp. 349-361. Holt, Rinehard and Winston, NY, 1962.

[5] Duda, R. O. and Martens, W. L., "Range Dependence of the Response of a Spherical Head Model," to be published in the *J. Acoust. Soc. Am.*

[6] Sugiyama, K., "Diffraction Coefficients of Prolate Spheroids," *J. Acoust. Soc. Japan*, vol. 45, pp. 527-533, July 1989.

[7] Novy, R. W., "Characterizing Elevation Effects of a Prolate Spheroidal HRTF Model," Tech. Report 20, NSF Grant IRI-9619339, Dept. of E.E., San Jose State University, San Jose, CA, May, 1998.