Estimation of a Spherical-Head Model from Anthropometry

V. Ralph Algazi,
Center for Image Processing and Integrated Computing,
University of California,
Davis, California 95616
algazi@ece.ucdavis.edu

Carlos Avendano
Creative Advanced Technology Center,
Scotts Valley, CA 95067
carlos_avendano@atc.creative.com

Richard O. Duda
Department of Electrical Engineering,
San Jose State University,
San Jose, California 95192,
rod@duda.org

March 2, 2001

Abstract

An optimal spherical-head model is derived from the high-frequency interaural time difference (ITD) of a population of 25 subjects. Analysis demonstrates that the optimal sphere results in very small lateral angle errors, except near the interaural poles. The spherical-head model is then estimated using the anthropometry of the subjects, based on simple and robust empirical predictive equations. Such a customization greatly decreases objective angular errors that occur when a generic model is used.

1 Introduction

The spherical-head model has been widely used in the spatial hearing field [1], [2], [3], [4], [5],[6] . The sphere is a very good first approximation to the human head and its analytical simplicity facilitates the design of signal processing algorithms for sound

spatialization. The spherical-head model is generally used to predict or to synthesize spatial sound having both the interaural time difference (ITD) and the interaural level difference (ILD) of the human head-related transfer function (HRTF) [2]. However, the radius of the spherical-head model needs to match the dimensions of the head of the listener. The use of an incorrect head radius introduces corresponding errors in the apparent location of the virtual auditory source. The localization of sound sources in azimuth, determined by the ITD, is quite accurate, with a localization sensitivity or blur of 1° or a sensitivity to interaural time differences as low as $10\mu s$[7]. Such a differential sensitivity corresponds approximately to an error as low as 1% in head radius. Equally important, when a head-tracker is used with a mismatched head size, it is a common experience that a virtual source at a fixed location will appear to move in the direction of the listener's head motion if the model is smaller than the listener's head and in the opposite direction if the model head is larger. It is also a common experience that the range of perceived azimuths will be reduced if the head size is too small and that the subjects will be unable to localize sources if the ITD exceed substantially the maximum they experience naturally [8].

In 1921, Hartley and Fry [9] measured the average head radius for a "number of individuals" and reported a value of 8.75 cm. This is the value frequently used in the literature and is also very close to the average value we observed in our study. However, head size varies substantially within the population. For the adult population in the US, the distribution of sizes is different for men and women. The mean and the first and ninety-ninth percentiles of head length and head width are tabulated in Table 1 [10].

We have used the terminology of [10], where the "head length" refers to the front/back dimension (Fig. 1). Note that there is a 20% variation in size for both male and female populations, for either head length or head width. Considering the population as a whole, the size variation is close to 30%. Furthermore, the typical head is 25% longer

Table 1: Head dimension statistics of US population [cm], L=length, W=width.

	1% L	Mean L	99% L	1% W	Mean W	99% W
Women	15.0	18.0	19.8	13.2	14.5	15.9
Men	18.0	19.6	21.4	14.2	15.5	16.9

than it is wide. Therefore it is important to determine the radius of the spherical-head approximation for each person, but it is also not clear how to do so.

Both the ITD and ILD are functions of spatial location and frequency.[1] The low-frequency ITD (below approximately 1.5 kHz) is the principal contributor to the lateral localization of sound [7]. In an experimental study using the KEMAR mannequin, Kuhn found that the ITD was approximately 50% greater at low frequencies than at high frequencies, with a transition occurring between 500 Hz and 3 kHz [3]. He also showed that the average high-frequency ITD is predicted well by Woodworth's ray-tracing formula, which links the ITD directly to the head radius [1]. Kuhn observed that the theoretical low-frequency ITD for a spherical head model is exactly 1.5 times Woodworth's high-frequency ITD. Thus, the psycho-acoustically significant low-frequency ITD can be easily obtained from the high-frequency ITD, which is precisely known for the spherical head model.

This is important, because problems arise when one attempts to accurately estimate the ITD from experimentally measured HRTFs. Minnaar et al. [12] analyzed and compared several frequency-domain and time-domain methods for directly estimating the low-frequency ITD. Because the loudspeakers used to measure HRTFs are relatively small, the experimental results are often untrustworthy below 1 kHz, and never extend all the way down to DC. Thus, the asymptotic low-frequency ITD — the limit of the

[1]Mathematically, the ITD at a particular frequency is defined as the difference in the group delays for impulse responses for the two ears, where the group delay is the negative rate of change of phase with frequency.

group delay as the frequency approaches zero — has to be indirectly inferred from experimental data. Estimating the high-frequency ITD from the phase spectrum is also problematic, because torso reflections and pinna reflections and resonances will affect the phase response for HRTF data for human subjects. However, the average high-frequency ITD can be reliably estimated by the difference between the leading edges of the left-ear and right-ear impulse responses [3]. In particular, when this procedure is applied to the exact impulse responses for the spherical-head model, the results have been shown to be in close agreement with Woodworth's formula [6]. Thus, we chose to use the rise-time method to estimate the high-frequency ITD, using Woodworth's formula to link the ITD to the sphere radius, and multiplying the results by 1.5 when we wanted the low-frequency ITD.

For these reasons, this work will focus on the high-frequency ITD (called only ITD from now on) and on Woodworth's formula as its predictor. This study has two goals. The first goal is to analyze the fit of the ITD of an optimal spherical-head approximation to the measured ITDs for a population of subjects. The second goal is to devise, for the same population, a predictive formula for the spherical-head radius based on anthropometry.

2 Optimal Spherical-Head Radius from ITD Measurements

The optimal head radius of the spherical-head model was computed using a least squares fit between the measured ITD and the ITD produced by Woodworth's formula.

2.1 Measurements

HRTF data were collected using an interaural-polar coordinate system (see Fig. 2). In these coordinates, the left/right dimension is measured by the lateral angle θ, often denoted "azimuth" and the up/down dimension is measured by the polar angle ϕ. Head-related impulse responses (HRIR's) were measured in the time domain, at a sampling rate of 44.1 kHz, with a Crystal River Snapshot^{TM} system, using a Golay code of length 2048. The recordings were made with Etymōtic ER$-$7C probe microphones (with their battery-powered pre-amplifiers) and the interface to the PC was through a Momentum 56 PC card. The ITD measurements were obtained at 25 different lateral angles ($-80^\circ \leq \theta \leq 80^\circ$) and 50 polar angles ($-45^\circ \leq \phi \leq 235^\circ$), for a total of $N = 1250$ measurements for each ear for each subject (see [14] for more details on the measurement apparatus and coordinate system). The high-frequency Head-Related Impulse Responses, HRIRs, were obtained by filtering the measured response with a sixth-order high-pass Butterworth filter with a 1500 Hz cut-off frequency. The high-frequency ITD was computed as the difference between onset times of the left and right HRIRs. The onsets were defined as the instants at which the impulse responses reached 20% of their first maximum peak amplitudes. Onset times were computed at a resolution of an eighth of a sample. Fig. 3 shows an example of onset determination. The ITDs are evaluated over all azimuths and elevations and can be mapped as an surface in two dimensions. Such a surface is shown on the left of Fig. 4. We observe occasional irregularities in that surface that are due to abrupt subject motion during the measurement session. To smooth out these irregularities, a spherical harmonics expansion, equivalent to a Fourier series expansion but for spherical coordinates,[2] is computed. Truncation of the high-order coefficients in such an expansion result in the smoothed ITD surface on the right of Fig 4.

[2]See for instance [13] for a discussion of the application of spherical coordinates to HRTFs.

Fig. 5 shows a representative example of the ITD function derived by this method. For a perfect sphere and ears located at $\pm 90°$, the constant ITD contours would be vertical in the front view and concentric circles on the right view. In the experimental data, most of the deviations from the ITD of a sphere occurs for high lateral angles, as illustrated next.

2.2 Least Squares Fit

The optimal head radius of the spherical-head model for each subject was computed using a least squares minimization procedure. The measured ITD data for each subject was approximated by minimizing the error between measurements and the ITD produced by Woodworth's formula[1], which can be written as

$$\hat{\tau} = \frac{a}{c}(\sin|\theta| + |\theta|), \quad -90° \leq \theta < 90° \tag{1}$$

In Eq. 1, $\hat{\tau}$ is the modeled ITD, a is the radius of the sphere, c is the speed of sound, and θ is the lateral-angle. Suppose that we measure the actual ITD for N different data points to produce the measured ITD vector $\overline{\tau} = [\tau_1, \tau_2, \cdots, \tau_N]^T$. Let $\hat{\tau} = (a/c)r$, where $r = \sin\theta + \theta$, and let $\overline{r}$ be the corresponding vector of values of r computed for all measurement locations: $\overline{r} = [r_1, r_2, \cdots, r_N]^T$. Then the optimal radius a_0 can be obtained by minimizing the length of the error vector $\overline{e} = \overline{\tau} - (a/c)\overline{r}$, which yields

$$a_0 = c\frac{\overline{\tau}^T\overline{r}}{\overline{r}^T\overline{r}} \tag{2}$$

In this optimization, the speed of sound was taken to be 343 m/s. To validate this procedure, it was applied to ITD measurements made for a bowling ball, where the radius was known to be 10.91 cm. Except for some distortion at high lateral angles (of

about $\pm 2^{\circ}$) due to the HRTF measurement apparatus, Woodworth's formula corresponds closely to the high-frequency ITD derived from onsets. The optimal radius computed from Eq. 2 yielded a value of 11.0 cm, thus introducing an estimate error of less than 1%.

2.3 Results

The subjects used in this study were adult students, both male and female, Caucasian and Asian, with a representative range of head sizes. The distribution of optimal head radius for the 25 subjects shows a range of values from 7.9 cm to 9.5 cm, with a mean value of 8.7 cm. The RMS error of the fit ranges from 22 to 47μs, and has mean of 32μs. The angular error can be determined by mapping the measured ITD to lateral angle based on the optimal spherical-head model for each subject. Global results are shown in Fig. 6 where the average angular error of all subjects is plotted as a function of spatial coordinates. The individual errors of two representative subjects are shown in Fig. 7. For the most part the angular error is smaller than 5 degrees. The largest errors (in the order of 12 degrees) occur at high lateral angles, and for most subjects it is largest at medium polar angles (above the subject). Note that front/back and left/right asymmetries are more apparent for individual subjects. They are due, in part, to a slight tilt of the head during the HRIR measurements.

This localized angular error is primarily due to the displacement of the entrance to the ear canal with respect to the center of the head. As shown in [15], the polar angle asymmetry introduced by this displacement causes the ITD to vary on a cone of confusion, resulting in local deviations of up to 15% in ITD at medium polar angles.

It should also be mentioned that the non-spherical nature of the human head will also contribute to this type of error. It can be shown that an ellipsoidal model with

offset ears will reduce the localized error of a spherical head model at mid-polar angles [15].

3 Optimal Spherical-Head Radius Estimation from Anthropometry

In the previous section, the optimal spherical-head model was computed from the acoustically measured ITD data for each subject.Usually the individualized HRTF is not available and the ITD of the listener is unknown. A method to estimate a customized spherical-head model from anthropometry would therefore be very attractive. Without anthropometry, the head radius for the model could be taken to be the mean head radius of the population. Doing so will clearly introduce large errors for listeners whose head size is far from the mean. When the individual head dimensions are available, the task of adapting the sphere radius to each listener becomes feasible.

Anthropometric parameters were obtained from photographs of the subjects. For feature identification and calibration purposes, reference marks were attached to each subject's head (See Fig.1). Three head parameters, X_1 (head half width), X_2 (head half height) and X_3 (head half length) were computed from the side and front views of the subject's head by manually identifying head boundaries on high-resolution digital photographs.

The optimal sphere radii can now be related to anthropometric parameters by empirical regression formulas that are derived using the statistics of the population. In this section we derive such empirical formulas for the estimation of the spherical-head model radius.

A first approximation is to average the three head dimensions, length, width and

height. Comparison between the optimal radius and the average of the three head parameters for all subjects indicates that the average overestimates the optimal radius in all cases (see Fig. 8). Middlebrooks finds a high correlation between head width and maximum ITD [16]. A more general linear model that can readily be optimized is to estimate the head radius as a weighted sum of the head dimensions, i.e.

$$a_e = w_1 X_1 + w_2 X_2 + w_3 X_3 + b. \tag{3}$$

If the optimal sphere radius is the desired parameter, then a regression on the anthropometric data for all subjects can be used to compute the values of the weights w_1, w_2 and w_3, and the bias b. Performing such a regression on the data for all subjects yields the following weights: $w_1 = 0.51$, $w_2 = 0.019$, $w_3 = 0.18$, and $b = 3.2$ cm. These results show that the dimension X_2, the half height of the head, is not very significant. The height dimension is also the most difficult to determine for human heads. It is interesting to note as well the large constant term in the regression equation.

To investigate the sensitivity of the weights in the regression equation, the error in head radius caused by perturbing them was analyzed. These sensitivities are measured in radius error in centimeters per percent change in the weight. The results are: $\Delta_{w_1} = 0.36$, $\Delta_{w_2} = 0.003$, $\Delta_{w_3} = 0.18$, and $\Delta_b = 0.32$ [cm/Δ%]. As expected, the weight w_2 is not critical and could be set to zero. The weights w_1 and w_2 are both important as well as the significant constant offset b.

The estimation error from Eq. 3 is shown in Fig. 8 together with the errors introduced by using a generic model and the mean of the head dimensions (i.e. $w_1 = w_2 = w_3 = 1/3$, and $b = 0$). This generic model will introduce large errors, in the order of 0.8 cm for subjects whose head radii are away from the mean. In contrast, the anthropometry-derived radii are closer to the optimal throughout the range of head sizes, with an

average error of 0.12 cm. Notice that averaging the head dimensions to obtain the sphere radius is the worst strategy.

We can also map angular errors due to a head radius mismatch into angular errors by using Woodworth's formula. We find that angular errors are worse for high lateral angles and when the head radius is under estimated. The perceptual effects of large ITD errors is complex, and excess ITD, beyond the subject's normal experience, leads to a diffuse source which is difficult to localize[8].

4 Conclusion

It has been shown that the high-frequency ITD of human subjects can be predicted with high accuracy by a customized spherical-head model derived from HRTF measurements. Customization based on simple head measurements was also derived and tested. The anthropometry based empirical formula robustly predicts the optimal radius of the model for the population investigated.

5 Acknowledgments

This work was supported by the National Science foundation under grants IRI-96-19339 and ITR-00-97256, and the DiMI program of the University of California. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors would like to thank Dennis Thompson and Valerie Tellefson for their assistance in collecting the HRTF data. We also acknowledge with thanks the constructive comments and suggestions of the anonymous reviewers.

References

[1] R.S. Woodworth and H. Schlosberg, *Experimental Psychology*, Holt, Rinehard and Winston, NY, pp. 348-361, 1962.

[2] J. Blauert, *Spatial Hearing*, MIT press, Cambridge MA, 1974.

[3] G.F. Kuhn, "Model for the Interaural Time Difference in the Azimuthal Plane," *J. Acoust. Soc. of Am.*, Vol. 62, No. 1, pp. 157-167, July 1977.

[4] J.A. Molino, "Psychophysical Verification of Predicted Interaural Differences in Localizing Distant Sound Sources," *J. Acoust. Soc. of Am.*, Vol. 55, No. 1, pp. 139-147, January 1974.

[5] D. H. Cooper and J. L. Bauck, "Prospects for Transaural Recording," *J. Aud. Eng. Soc.*, Vol. 37, pp. 3-19, January/February 1989.

[6] R. O. Duda and W. L. Martens, "Range Dependence of the Response of a Spherical Head Model" *J. Acoust. Soc. Am.*, Vol. 104, pp. 3048-3058, November 1998.

[7] J. Blauert, *Spatial Hearing, Chapter 2*, Revised Edition, MIT press, Cambridge MA, 1997below 500 Hz

[8] B. G. Shinn-Cunningham, N.I. Durlach and R.M. Held "Adapting to Supernormal Auditory Localization Cues," *J. Acoust. Soc. of Am.*, Vol. 103, No. 6, pp. 3656-3666, 1998

[9] R. V. L. Hartley and T. C. Fry, "The Binaural Location of Pure Tones" *Phys. Rev.* Vol. 18, pp. 431-442, 1921.

[10] Henry Dreyfuss Associates, *The Measure of Man and Woman*, Whitney Library of Design, New York 1993.

[11] A. Kulkarni, S.K. Isabelle and H.S. Colburn, "Sensitivity of Human Subjects of Head-Related Transfer-Function Phase Spectra," *J. Acoust. Soc. of Am.*, Vol. 105, No. 5, pp. 2821-2840, May 1999.

[12] P. Minnaar, J. Plogsties, S. K. Olesen, F. Christensen and H. Møller "The Interaural Time Difference in Binaural Synthesis," *108th AES Convention* Paris, France, preprint 5133, February 2000.

[13] M. J. Evans, J. A. S. Angus, A. I. Tew : "Analyzing head-related transfer function measurements using surface spherical harmonics", *J. Acoust. Soc. of Am.*, vol. 104, pp. 2400-2411 (1998)

[14] V.R. Algazi, C. Avendano and D. Thompson, "Dependence of Subject and Measurement Position in Binaural Signal Acquisition," *J. Audio Eng. Soc.*, Vol. 47, No. 11, pp. 937-947, November 1999.

[15] R.O. Duda, C. Avendano and V.R. Algazi, "An Adaptable Ellipsoidal Head Model for the Interaural Time Difference," *IEEE international Conference on Acoustics Speech and Signal Processing*, Phoenix AZ, pp. II-965-968, 1999.

[16] J. C. Middlebrooks, "Individual Differences in External Ear Transfer Functions Reduced by Scaling in Frequency", *J. Acoust. Soc. of Am.*, Vol. 106, No. 3, pp. 1480-1492, September 1999.

List of Figure Captions

- Figure 1. Head parameters obtained from digital photographs, where $2X_1$=head width, $2X_2$ = head height and $2X_3$= head length. The photograph has 1024 by 1344 pixels and 0.04 cm/pixel spatial resolution.
- Figure 2. Interaural polar coordinate system.
- Figure 3. Threshold determination of onsets on the ipsilateral and contralateral HRIRs.
- Figure 4. The raw ITD shows small irregularities due to head motion during the measurement session. These are smoothed by truncation of the spherical harmonic series representation of the ITD surface.
- Figure 5. An experimental high-frequency ITD function.
- Figure 6. Average angular error, in degrees, produced by the optimal spherical-head model.
- Figure 7. Angular error produced by the optimal spherical-head model for two representative subjects.
- Figure 8. Error between the optimal head radii and a generic radius (-). the mean head dimension (...) and the anthropometry derived radii (-.). Subjects sorted by head size.

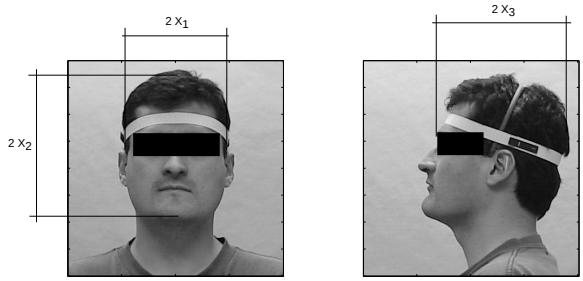

Figure 1: Head parameters obtained from digital photographs, where $2X_1$=head width, $2X_2$ = head height and $2X_3$= head length . The photograph has 1024 by 1344 pixels and 0.04 cm/pixel spatial resolution.

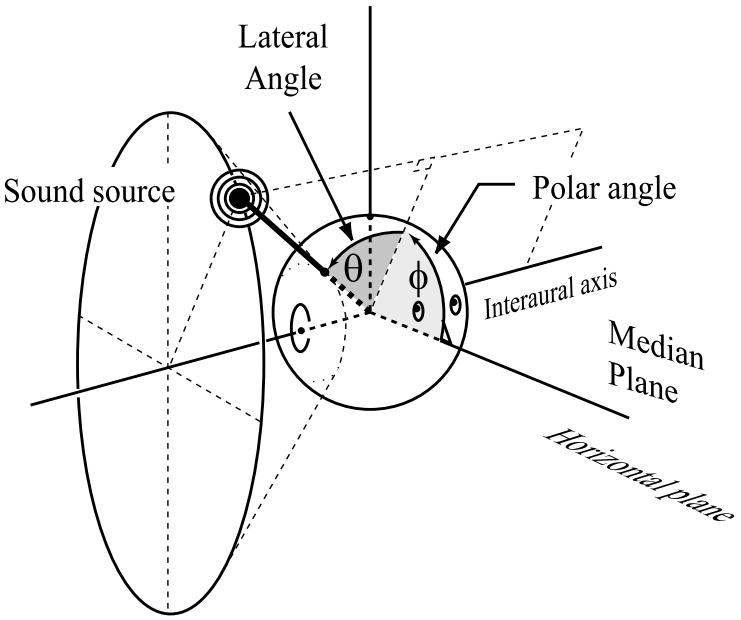

Figure 2: Interaural polar coordinate system

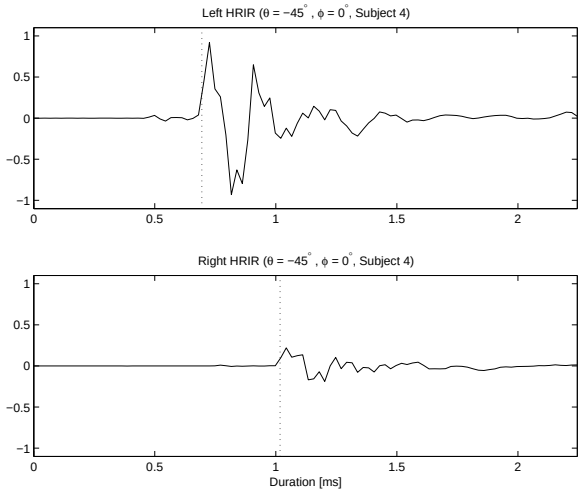

Figure 3: Threshold determination of onsets on the ipsilateral and contralateral HRIRs.

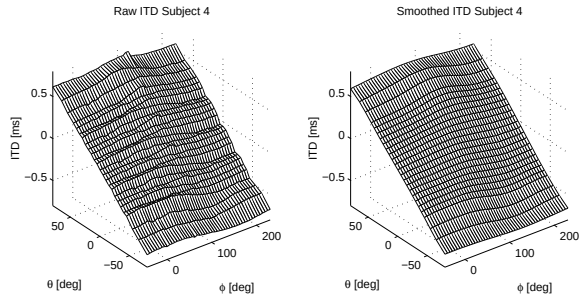

Figure 4: The raw ITD shows small irregularities due to head motion during the measurement session. These are smoothed by truncation of the spherical harmonic series representation of the ITD surface.

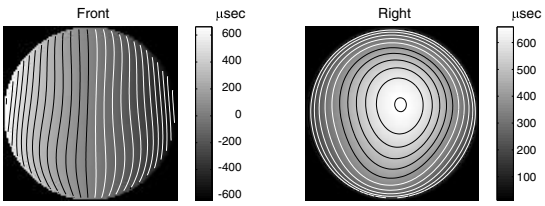

Figure 5: An experimental high-frequency ITD function

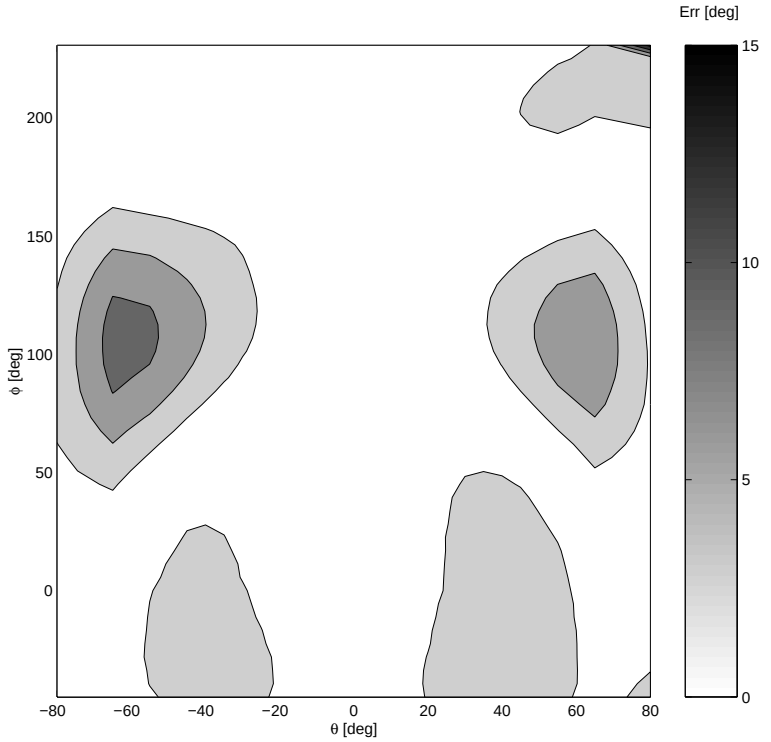

Figure 6: Average angular error, in degrees, produced by the optimal spherical-head model.

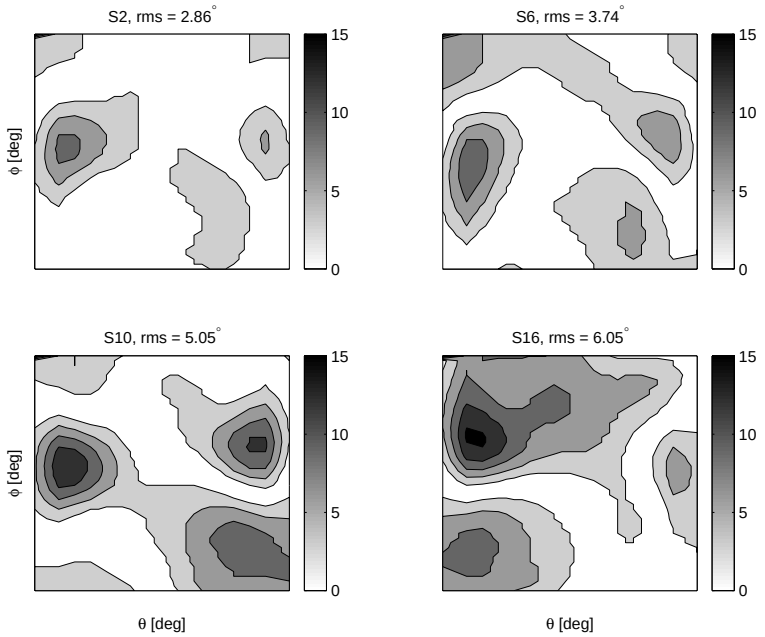

Figure 7: Angular error produced by the optimal spherical-head model for four representative subjects.

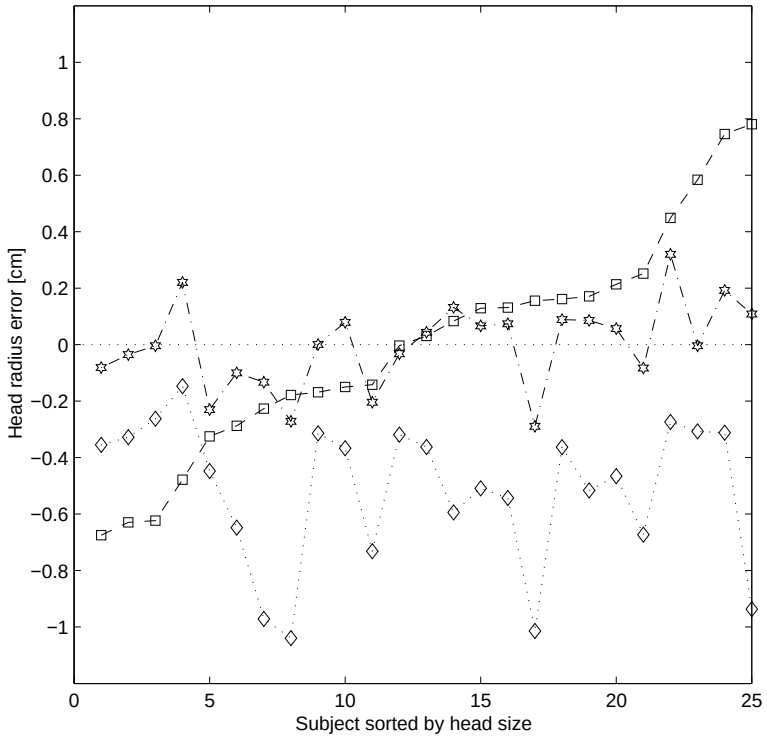

Figure 8: Error between the optimal head radii and a generic radius (-). the mean head dimension (...) and the anthropometry derived radii (-.). Subjects sorted by head size.