ANTHROPOMETRIC PARAMETERISATION OF A SPHERICAL SCATTERER ITD MODEL WITH ARBITRARY EAR ANGLES

Hannes Gamper, Mark R. P. Thomas, Ivan J. Tashev

Microsoft Research
Redmond, WA 98052, USA
{hagamper, markth, ivantash}@microsoft.com

ABSTRACT

Accurate modelling of the interaural time difference (ITD) is crucial for rendering localised sound. Parametric models allow personalising ITDs using anthropometrics. However, the mapping between anthropometric features and model parameters is not straightforward. Here, we propose deriving personalised ITD model parameters from a sphere fitted to a 3-D head scan. The proposed ITD personalisation is evaluated on an HRTF database containing 181 subjects, for a simple spherical ITD model as well as a frequency and elevation-dependent spherical scatterer model with arbitrary ear angles.

Index Terms— Head-related transfer function, interaural time delay, spherical scatterer, 3-D head scan

1. INTRODUCTION

Head-related transfer functions (HRTFs) describe the acoustic path between a sound source and a listener's ear entrances. HRTFs are highly individual, due to their dependence on the anthropometry of the listener. Accurate measurement or modelling of the HRTFs is necessary in spatial sound rendering to correctly reproduce the position of virtual sound sources. Recently, the mapping between anthropometric to acoustic features [1], anthropometry-based HRTF modelling and synthesis [2, 3, 4, 5, 6], and "best match" HRTF selection [7, 8] have been studied.

The interaural time difference (ITD) is defined as the propagation delay difference between the right and left ear entrance. Due to the spatial separation of the ear entrances, the ITD varies as a function of the source direction. The direction dependence of the ITD is an important HRTF cue that allows the human auditory system to localise acoustic sources [9, 10]. To render the location of a virtual sound source correctly, accurate reproduction of ITDs is crucial [11]. As ITD measurements are seldom available for a particular user, models have been proposed previously to estimate the ITDs based on head geometry. Woodworth proposed a model relating the ITD to the radius of a sphere approximating the head and the angle of sound incidence [12]. Due to its simplicity and computational efficiency, the Woodworth model is commonly used in spatial sound rendering systems [13, 14]. Kuhn performed ITD measurements on a manikin and compared the results to ITDs derived by modelling the scattering on a rigid sphere approximating the head [15]. He showed both measured and modelled ITDs to be frequency dependent, decreasing from a low-frequency limit asymptotically to a high-frequency limit at around 2 kHz. Experiments by Kulkarni et al. indicate that human listeners are not sensitive to the frequency dependence of ITDs [16]. Although these results were confirmed by Constan and Hartmann, they caution that the effective ITD of a broadband stimulus is a weighted average of the actual, frequency-dependent ITD [17]. Algazi et al. showed that the radius of the Woodworth model can be derived as a linear combination of the head width, length, and height [11]. Aaronson and Hartmann extended the Woodworth model to account for source distance and ears moved forward or backward along the surface of the sphere [14]. Duda et al. introduced an ellipsoidal head model accounting for arbitrary ear angles [18]. The authors state that the measurements of a bounding box enclosing the head could be used to parameterise the model. However, the exact relationship between model parameters and anthropometric features was not established. We previously proposed a method for estimating ITDs directly from a high-resolution 3-D head scan using ray tracing [19].

Here, we study the relation between the parameters of a sphere fitting 3-D head scans and the parameters of Woodworth's ITD model and Kuhn's spherical scatterer model with arbitrary ear angles. Experiments are carried out on a large database containing the HRTF measurements and 3-D head scans of 181 subjects.

2. PROBLEM FORMULATION

Throughout this paper we assume a right-handed coordinate system. Let $\Omega \triangleq (\theta, \phi)$, where $\theta \in [0, \pi]$ is colatitude angle, measured down from the positive z (up) axis, and $\phi \in [0, 2\pi)$ is azimuth angle, measured counterclockwise from the positive x (forward) axis about the z axis. Additionally, let $\psi \in [-\pi/2, \pi/2]$ be lateral angle in interaural coordinates [20]. For every subject there exists an impulse response $h(\Omega, t)$ with corresponding transfer function $H(\Omega, \omega)$, and a set of vertices $\mathcal{X}$ sampling the 3D head scan, centred on the interaural axis.

Let $\tau(\Omega, \omega)$ be a reference ITD for plane waves incident on the centre of the coordinate system, where ω is the centre frequency of a spectral sub-band with a finite bandwidth. The PHAT-weighted cross-correlation function is calculated as,

$$R(\Omega, \omega, \tau) = \int_{\omega-\Delta^-}^{\omega_\mathrm{c}+\Delta^+} \frac{H^\mathrm{R}(\Omega,\omega)H^\mathrm{L}(\Omega,\omega)^*}{|H^\mathrm{R}(\Omega,\omega)H^\mathrm{L}(\Omega,\omega)^*|} e^{j\omega\tau}, \quad (1)$$

where Δ^- and Δ^+ are arbitrary spectral band limits, and $\{\mathrm{L}, \mathrm{R}\}$ denote left and right ears respectively. The frequency-dependent reference ITDs can then be estimated:

$$\tau(\Omega, \omega) = \arg\max_{\tilde{\tau}} R(\Omega, \omega, \tilde{\tau}). \quad (2)$$

The aim is to approximate $\tau(\Omega, \omega)$, using a trained parametric model ξ that estimates $\hat{\tau}(\Omega, \omega|\xi)$ from the fitted parameters.

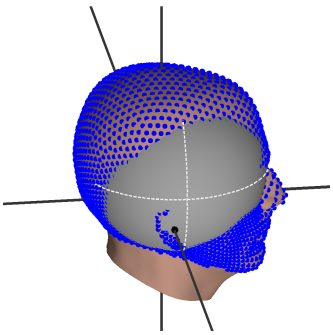

Figure 1: Sphere fitted to uniformly distributed points on the surface of a 3-D head scan; dashed lines indicate equator and meridian.

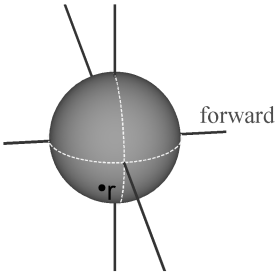

Figure 2: Proposed spherical scatterer model with arbitrary ear angles; $\mathbf{r}$ denotes ear vector, dashed lines equator and meridian.

3. PARAMETER ESTIMATION

3.1. Sphere Fitting to Point Cloud

The vertices $\mathcal{X}$ of a 3-D head scan are resampled to give $\bar{\mathbf{x}}_p = [\bar{x}_p\ \bar{y}_p\ \bar{z}_p]^T$, the points of intersection between $\mathcal{X}$ and the normals of a cloud of P points uniformly distributed on the surface of a bounding sphere [21] (see Figure 1, blue dots). A sphere with centre $\mathbf{x}_\mathrm{c} = [x_\mathrm{c}\ y_\mathrm{c}\ z_\mathrm{c}]^T$ and radius r_c is calculated to best fit $\mathcal{X}$ by solving

$$(\mathbf{x}_\mathrm{c}, r_\mathrm{c}) = \underset{\tilde{\mathbf{x}}_\mathrm{c}, \tilde{r}_\mathrm{c}}{\arg\min} \sum_{p=0}^{P-1} \left| \|\bar{\mathbf{x}}_p - \tilde{\mathbf{x}}_\mathrm{c}\|_2 - \tilde{r}_\mathrm{c} \right| \tag{3}$$

using Matlab's `fmincon`.

3.2. Woodworth Model

The Woodworth Model [12] predicts ITD by modeling the path of propagation around a sphere,

$$\hat{\tau}_\mathrm{w}(\psi | r_\mathrm{w}) = \frac{-r_\mathrm{w}(\psi + \sin\psi)}{c}, \tag{4}$$

where c is the speed of sound $(\mathrm{m} \cdot \mathrm{s}^{-1})$ and r_w is the radius of a spherical approximation of the head. The model is independent of frequency and interaural polar angle. Quantizing frequency and space into discrete points ω_k, Ω_i, with corresponding interaural lateral angles ψ_i, the optimal radius r_w can be calculated,

$$r_\mathrm{w} = \underset{\tilde{r}_\mathrm{w}}{\arg\min} \sum_{k=0}^{K-1} \sum_{i=0}^{N-1} \left(\tau(\Omega_i, \omega_k) - \hat{\tau}_\mathrm{w}(\psi_i | \tilde{r}_\mathrm{w}) \right)^2. \tag{5}$$

3.3. Spherical Scatterer Model

The Kuhn scattering model [15] derives the complex pressure on the surface of a rigid scattering sphere from a unit plane wave,

$$p(\Omega, \omega | \xi_\mathrm{s}) = \frac{1}{(kr_\mathrm{s})^2} \sum_{n=0}^{L} \frac{i^{n+1}(2n+1) P_n(\cos\Theta)}{h_n^{(2)'}(kr_\mathrm{s})}, \tag{6}$$

where $\xi_\mathrm{s} = (\theta_\mathrm{s}, \phi_\mathrm{s}, r_\mathrm{s})$, $k = \frac{\omega}{c}$, r_s is the radius of a spherical approximation of the head, L is the expansion order, $i = \sqrt{-1}$, $P_n(\cdot)$ is a Legendre polynomial, $h_n^{(2)'}(x)$ is the derivative of a spherical Hankel function of the second kind, and $\cos\Theta$ is derived from the solid angle between the vector of the incident plane wave $\mathbf{r} \triangleq (\theta, \phi, 1)$ and the ear vector $\mathbf{r}_\mathrm{s}$,

$$\cos\Theta = \mathbf{r} \cdot \frac{\mathbf{r}_\mathrm{s}}{r_\mathrm{s}}. \tag{7}$$

Nominally the ears lie at $\left(\frac{\pi}{2}, \pm\frac{\pi}{2}, r_\mathrm{s}\right)$; here we propose an extension such that $\mathbf{r}_\mathrm{s} = \left(\frac{\pi}{2} + \theta_\mathrm{s}, \pm\frac{\pi}{2} + \phi_\mathrm{s}, r_\mathrm{s}\right)$. The frequency-dependent ITD is then derived from (6),

$$\hat{\tau}_\mathrm{s}(\Omega, \omega | \xi_\mathrm{s}) = \left(\angle p^\mathrm{R}(\Omega, \omega | \xi_\mathrm{s}) - \angle p^\mathrm{L}(\Omega, \omega | \xi_\mathrm{s}) \right) / \omega, \tag{8}$$

where $\angle\cdot$ is the angle of a complex variable. The optimal spherical scatterer model parameters are finally found by solving

$$\xi_\mathrm{s} = \underset{\tilde{\xi}_\mathrm{s}}{\arg\min} \sum_{k=0}^{K-1} \sum_{i=0}^{N-1} \left(\tau(\Omega_i, \omega_k) - \hat{\tau}_\mathrm{s}(\Omega_i, \omega_k | \tilde{\xi}_\mathrm{s}) \right)^2. \tag{9}$$

Figure 2 depicts the proposed spherical scatterer model with ear angles, $(\theta_\mathrm{s}, \phi_\mathrm{s}) = (25°, 20°)$.

4. EXPERIMENTAL EVALUATION

Experiments were carried out on a database of HRTF measurements and 3-D scans containing 181 subjects. The HRTFs were measured for 400 directions at a constant radius, as described by Bilinski et al. [3]. The 3-D head scans were obtained using a Flexscan3D optical scanning setup.

For each subject, a sphere was fitted to the 3-D scan by solving (3). Figure 1 illustrates the result of the fitting process for one subject. The offset of the sphere's origin can be expressed in terms of a colatitude and azimuth offset, θ_c and ϕ_c, with respect to the interaural axis:

$$\theta_\mathrm{c} = \arctan\left(\frac{z_\mathrm{c}}{\sqrt{r_\mathrm{c}^2 - z_\mathrm{c}^2}} \right), \tag{10}$$

$$\phi_\mathrm{c} = \arctan\left(\frac{y_\mathrm{c}}{\sqrt{r_\mathrm{c}^2 - y_\mathrm{c}^2}} \right). \tag{11}$$

Using (2), the reference ITDs were calculated for each subject in the database as the interaural cross correlation, a method commonly used in this context [22]. The optimal parameters for the Woodworth model and the spherical scatterer were obtained by solving (5) and (9) for the spectral sub-band centre frequencies

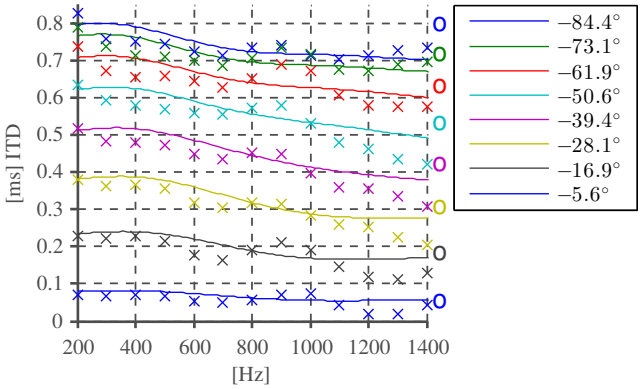

Figure 3: Frequency dependence of measured (reference) ITDs (×) on the horizontal plane at various lateral angles for one subject, and ITDs estimated by solving (5) for the Woodworth model (○), and (9) for the proposed spherical scatterer model (solid lines). The x-axis labels indicate the centre frequencies, ω, used in the parameter optimisation.

Table 1: Optimisation results (181 subjects).

Eq.	radius [cm]			θ [°]			ϕ [°]		
	min	mean	max	min	mean	max	min	mean	max
(3)	8.4	9.9	10.7	14.3	25.4	30.7	-2.9	4.3	14.7
(5)	10.1	11.5	12.6	-	-	-	-	-	-
(9)	8.5	9.6	10.4	-0.3	11.7	21.5	-15.9	-3.8	8.8

$\omega \in \{200, 400, 600, 800, 1000, 1200, 1400\}$. Figure 3 illustrates the measured ITDs of one subject (×) and the ITDs estimated using the spherical scatterer with optimised parameters. Both estimated and measured (reference) ITDs decrease with frequency, whereas the Woodworth model (○) is frequency independent. The spherical scatterer seems to provide fairly accurate ITD estimates across lateral angle and frequency, whereas the best-fitting Woodworth model overestimates ITDs at large lateral angles. Table 1 summarises the optimisation results. For a sphere fitted to the 3-D head scans, the average optimal ear colatitude and azimuth offsets, θ_c and ϕ_c, are positive, indicating that the ears typically lie below and behind the centre of the head, as suggested by Duda et al. [18]. The same is true for the average optimal ear colatitude offset of the spherical scatterer. The average optimal radii of all three models are larger than the 8.75 cm suggested by Algazi et al. [11]. For the spherical scatterer and the sphere fitted to the 3-D scans, the larger radius is a result of the models allowing for arbitrary ear angles: If the ears lie below and behind the centre of the sphere or scatterer, the distance between the ears is less than the diameter, which reduces the ITD. The relatively large optimal radii of the Woodworth model are a result of the Woodworth model underestimating ITDs below about 1.5 kHz [14].

Figure 5a–c depicts the correlations between the radii of the three models. The radii are fairly well correlated, with Pearson correlation coefficients above 0.8. This indicates that the radius of a sphere, r_c, fitted to a 3-D head scan can be used to estimate the radius of the Woodworth model, r_w, or the proposed spherical scatterer model, r_s. Using a line fit, the following mapping functions are derived:

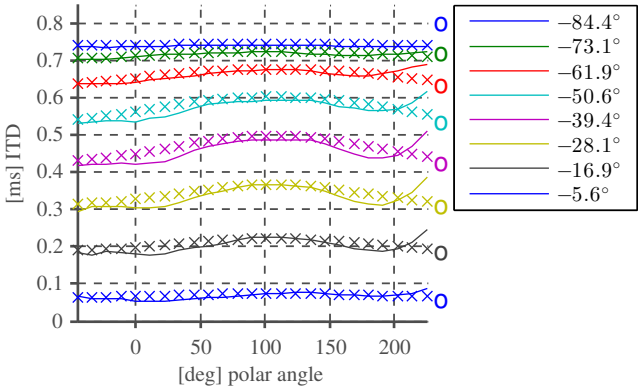

Figure 4: Elevation dependence of measured (reference) ITDs (×) at various lateral angles for one subject, and ITDs estimated by solving (5) for the Woodworth model (○) and (9) for the proposed spherical scatterer model (solid lines).

$$r_w = 1.0253\, r_c + 1.3947 \text{ cm}, \quad (12)$$

$$r_s = 0.8390\, r_c + 1.2703 \text{ cm}, \quad (13)$$

and

$$r_s = 0.7553\, r_w + 0.8543 \text{ cm}. \quad (14)$$

Similarly, the ear colatitude and azimuth offsets estimated from the sphere fitted to the head scans, θ_c and ϕ_c, can be used to predict the colatitude and azimuth offsets of the spherical scatterer:

$$\theta_s = 0.6981\, \theta_c - 6.0466^\circ \quad (15)$$

and

$$\phi_s = 1.0028\, \phi_c - 8.1074^\circ. \quad (16)$$

Figure 5d shows the root-mean-squared error (RMSE) of the Woodworth and the spherical scatterer model with optimal parameters. The spherical scatterer model clearly outperforms the Woodworth model, as it has two more free parameters, i.e., ear colatitude and azimuth offset, and it models both the frequency dependence of ITDs at low frequencies and the elevation dependence. Figure 4 illustrates the elevation dependence of the reference ITDs (×) and ITDs estimated using the spherical scatterer with optimal parameters (solid lines). Both the measured and modelled ITDs increase towards higher elevations, a result that is in line with findings from earlier studies [18, 19].

Using (12), (13), (15), and (16), the parameters for the Woodworth model and the spherical scatterer are estimated from the sphere fitted to the 3-D head scan. ITDs are estimated using both models and compared to the measured ITDs of the 181 subjects in the test database. The RMSEs with optimal and estimated parameters are summarised in Table 2. For both models, the RMSE increases only marginally when using parameters estimated directly from the 3-D scans. This is presumably the result of relatively low errors in the estimation of the model radii from the radii of the fitted spheres. Note that the errors reported here for either model are larger than errors reported previously, e.g., by Duda et al. and Algazi et al. [18, 11]. The errors reported here are for low frequencies, and averaged across frequency bands. Therefore, mismatches in the frequency dependence of the tested models increase the RMSE.

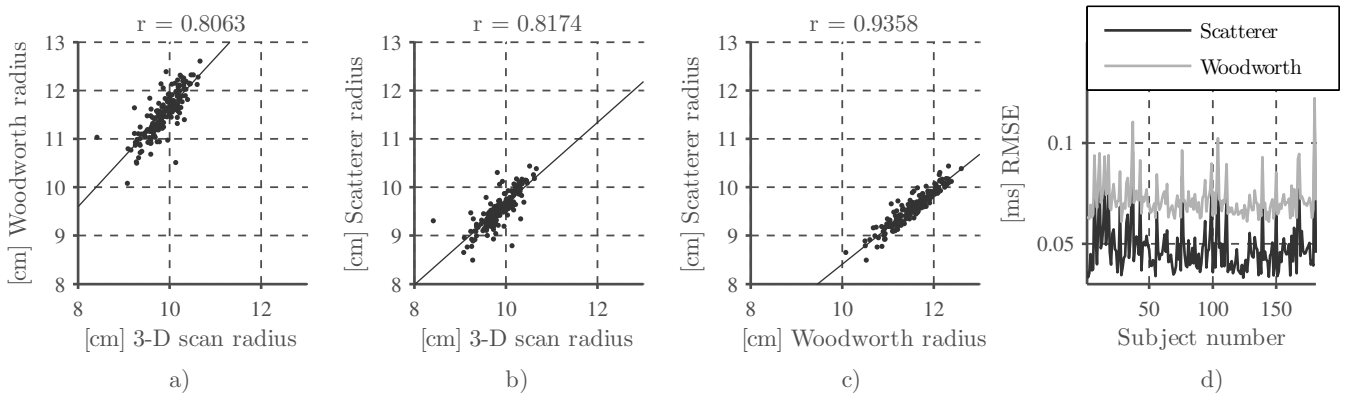

Figure 5: (a–c) Correlations between optimal radii of the three tested models and linear fits (dark line); r denotes the Pearson correlation coefficient; (d) root-mean-squared errors (RMSEs) of the tested ITD estimation models for 181 subjects.

Table 2: ITD estimation errors with optimal parameters and parameters estimated from a sphere fitted to the 3-D head scan of each subject.

Parameters	Woodworth (4) RMSE [ms]	spherical scatterer (6) RMSE [ms]
optimal	0.0725	0.0495
estimated from 3-D scan	0.0735	0.0513

Results reported by Constan and Hartmann suggest that this mismatch may not be perceptually relevant as long as the perceived, frequency-independent ITD is estimated accurately from the modelled ITD [17]. It should also be noted that no smoothing was performed on the raw reference ITD estimates.

5. CONCLUSIONS

We studied the estimation of low-frequency interaural time differences (ITDs) from anthropometric features using ITD models commonly used in spatial sound rendering applications: The Woodworth model and a spherical scatterer model extended to arbitrary ear angles to model the ITDs' frequency and elevation dependence. A comparison between measured ITDs and ITDs estimated using the spherical scatterer indicated good agreement across the tested frequency and elevation range. To evaluate the performance of the Woodworth model and the spherical scatterer model, ITDs of 181 human subjects were estimated by optimising the model parameters using ground-truth ITD measurements. The RMSEs were 73 μs and 50 μs, for the Woodworth model and the spherical scatterer model, respectively. These results indicate that the spherical scatterer model with arbitrary ear angles may provide more accurate ITD estimates than the Woodworth model, while also modelling frequency and elevation dependence of the ITDs. To investigate the mapping between anthropometric features and the estimated model parameters, we fitted the origin and radius of a sphere to the 3-D head scan of each test subject. For both tested ITD models, the radii of the spheres fitted to the 3-D head scans were highly correlated with the ITD model radii optimised using the ground-truth ITD measurements. In a second evaluation step, the performance of both ITD models was tested using parameters estimated directly from the spheres fitted to the 3-D head scans. The performance of both models decreased only slightly compared to using the optimal parameter set. This result indicates that a person's ITDs can be estimated using the Woodworth model or a spherical scatterer model with arbitrary ear angles using parameters derived directly from a 3-D head scan.

Future work includes the validation of the proposed anthropometric mapping functions using listening tests and a study on the robustness against noise or missing data in the head scans.

6. REFERENCES

[1] C. Jin, P. Leong, J. Leung, A. Corderoy, and S. Carlile, "Enabling individualized virtual auditory space using morphological measurements," in *Proc. IEEE Pacific-Rim Conf. on Multimedia*, Sydney, Australia, 2000, pp. 235–238.

[2] P. Satarzadeh, V. R. Algazi, and R. O. Duda, "Physical and filter pinna models based on anthropometry," in *AES 122nd Convention Preprints*, Vienna, Austria, 2007, paper number 7098.

[3] P. Bilinski, J. Ahrens, M. R. P. Thomas, I. J. Tashev, and J. C. Platt, "HRTF magnitude synthesis via sparse representation of anthropometric features," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process. (ICASSP)*, Florence, Italy, 2014, pp. 4501–4505.

[4] F. Grijalva, L. Martini, S. Goldenstein, and D. Florencio, "Anthropometric-based customization of head-related transfer functions using Isomap in the horizontal plane," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process. (ICASSP)*, Florence, Italy, 2014, pp. 4473–4477.

[5] I. Tashev, "HRTF phase synthesis via sparse representation of anthropometric features," in *Proc. Information Theory and Applications Workshop (ITA)*, San Diego, CA, USA, 2014.

[6] J. He, W.-S. Gan, and E.-L. Tan, "On the preprocessing and postprocessing of hrtf individualization based on sparse representation of anthropometric features," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process. (ICASSP)*, Brisbane, Australia, April 2015.

[7] D. N. Zotkin, J. Hwang, R. Duraiswami, and L. S. Davis, "HRTF personalization using anthropometric measurements," in *Proc. IEEE Workshop Applicat. of Signal Process. to Audio and Acoust. (WASPAA)*, New Paltz, NY, USA, 2003, pp. 157–160.

[8] D. Schönstein and B. F. G. Katz, "HRTF selection for binaural synthesis from a database using morphological parameters," in *Proc. Int. Congress on Acoustics*, Sydney, Australia, 2010.

[9] J. Blauert, *Spatial Hearing - Revised Edition: The Psychophysics of Human Sound Localization*. Cambridge, MA, USA: The MIT Press, 1996.

[10] F. L. Wightman and D. J. Kistler, "Factors affecting the relative salience of sound localization cues," in *Binaural and Spatial Hearing in Real and Virtual Environments*, R. H. Gilkey and T. R. Anderson, Eds. Lawrence Erlbaum Associates, 1997, pp. 1–23.

[11] V. R. Algazi, C. Avendano, and R. O. Duda, "Estimation of a spherical-head model from anthropometry," *Journal Audio Eng. Soc.*, vol. 49, no. 6, pp. 472–479, 2001.

[12] R. S. Woodworth and G. Schlosberg, *Experimental Psychology*. NY: Holt, Rinehard and Winston, 1962.

[13] D. Zotkin, R. Duraiswami, and L. Davis, "Rendering localized spatial audio in a virtual auditory space," *IEEE Trans. Multimedia*, vol. 6, no. 4, pp. 553–564, 2004.

[14] N. L. Aaronson and W. M. Hartmann, "Testing, correcting, and extending the Woodworth model for interaural time difference," *J. Acoust. Soc. Am.*, vol. 135, no. 2, pp. 817–823, 2014.

[15] G. F. Kuhn, "Model for the interaural time differences in the azimuthal plane," *J. Acoust. Soc. Am.*, vol. 62, no. 1, pp. 157–167, 1977.

[16] A. Kulkarni, S. K. Isabelle, and H. S. Colburn, "Sensitivity of human subjects to head-related transfer-function phase spectra," *J. Acoust. Soc. Am.*, vol. 105, no. 5, pp. 2821–2840, 1999.

[17] Z. A. Constan and W. M. Hartmann, "On the detection of dispersion in the head-related transfer function," *J. Acoust. Soc. Am.*, vol. 114, no. 2, pp. 998–1008, 2003.

[18] R. O. Duda, C. Avendano, and V. R. Algazi, "An adaptable ellipsoidal head model for the interaural time difference," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process. (ICASSP)*, Washington, DC, USA, 1999, pp. 965–968.

[19] H. Gamper, M. R. P. Thomas, and I. J. Tashev, "Estimation of multipath propagation delays and interaural time differences from 3-d head scans," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, Australia, Apr. 2015.

[20] E. A. Macpherson and J. C. Middlebrooks, "Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited," *J. Acoust. Soc. Am.*, vol. 111, no. 5, pp. 2219–2236, 2002.

[21] J. Fliege and U. Maier, "A two-stage approach for computing cubature formulae for the sphere," in *Mathematik 139T, Universität Dortmund, Fachbereich Mathematik, 44221*, 1996.

[22] B. F. G. Katz and M. Noisternig, "A comparative study of interaural time delay estimation methods," *J. Acoust. Soc. Am.*, vol. 135, no. 6, pp. 3530–3540, 2014.