ANTHROPOMETRIC TUNING OF A SPHERICAL HEAD MODEL FOR BINAURAL VIRTUAL ACOUSTICS BASED ON INTERAURAL LEVEL DIFFERENCES

Simone Spagnol

University of Padova
Department of Information Engineering
Via Gradenigo 6/B, Padova, Italy
spagnols@dei.unipd.it

Federico Avanzini

University of Padova
Department of Information Engineering
Via Gradenigo 6/B, Padova, Italy
avanzini@dei.unipd.it

ABSTRACT

In this paper we propose a method to customize a spherical head model for binaural sound rendering based on the listener's anthropometry. Interaural level difference (ILD) information from a HRTF database is used to subjectively tune the radius parameter of the spherical model so as to best fit individual measures. Multiple linear regression on anthropometric data is performed, yielding a closed formula relating the three head dimensions to the ILD-optimized radius. The effectiveness of the proposed radius estimation method in predicting the correct ILD with a spherical model is compared to that of alternative methods from the literature. Results show that the average spectral distortion between experimental and predicted ILDs with our method is significantly lower than with other estimation methods for lateral source locations. The proposed customization approach provides substance towards the development and evaluation of personal auditory displays for binaural virtual acoustics.

1. INTRODUCTION

At the beginning of the last century, Lord Rayleigh's studies on the scattering of sound waves by obstacles gave birth to the field of 3-D sound. In particular, he derived an analytical formulation of diffraction of sound waves around a spherical head [1] which provided a first glance of the so-called head-related transfer function (HRTF). The HRTF is the Laplace transform of the free-field compensated impulse response relative to the path of the sound wave from the source to the eardrum, i.e. the head-related impulse response (HRIR), and contains all of the information relative to sound transformations caused by the human body, in particular by the head, external ears, torso and shoulders. Such characterization allows virtual positioning of sound sources in the surrounding space: consistently with its relative position to the listener's head, the emitted signal can be filtered through the corresponding pair of HRTFs creating left and right ear signals to be delivered by headphones [2]. In this way, three-dimensional sound fields with a high immersion sense can be simulated and integrated into a great variety of contexts.

 This work was supported by the research project Personal Auditory Displays for Virtual Acoustics, University of Padova, under grant no. CPDA135702.

Unfortunately, acoustically obtaining individual HRTFs of a specific listener requires specific facilities, expensive equipment, and lengthy measurement sessions. For these reasons non-individual HRTFs, e.g. measured on dummy heads, are used in most applications. However, individual anthropometric features of the human body have a key role in HRTF-based playback: listening to non-individual spatially rendered sources typically increases the absolute localization error, the front-back reversal rate, and inside-the-head localization [3, 4, 5]. This is the reason why in the last few decades many researchers in the field of binaural audio spent their efforts towards efficient modeling of HRTFs.

The HRTF is a function of four variables: three spatial coordinates (distance, azimuth and elevation) and frequency. Despite the three coordinates being represented by localization cues mainly associated to a specific body part - i.e. azimuth and distance cues to the head; high-frequency elevation cues to the pinnae; low-frequency elevation cue to the torso - previous research failed its attempts at factoring the HRTF into an azimuth-, elevation-, or distance-dependent component. As a consequence, researchers have applied various filter design and/or machine learning techniques in the attempt to fit multiparameter models to experimental data (see e.g. [6, 7]). Unfortunately, real-time HRTF rendering requires fast computations which typically cannot undergo the complexity of the resulting filter coefficients/weights, that are themselves rather complicated functions of both azimuth and elevation, and a sufficiently accurate fit to anthropometry can only be obtained through multiple regression on long anthropometric vectors.

Structural modeling [8] ultimately represents an attractive solution to all of these shortcomings. In structural models the contributions of the listener's head, pinnae, shoulders and torso to the HRTF are isolated and arranged in different subcomponents each accounting for some well-defined physical phenomenon. The linearity of these contributions allows reconstruction of the global HRTF from a proper combination of all the considered effects [9]. Relating each subcomponent's temporal and/or spectral features in the form of low-order digital filter parameters to a subset of anthropometric quantities yields a cheap and customizable HRTF model. Following such an approach, a first-order filter model of source distance in the near field [10] and a low-order model of elevation-dependent pinna reflection patterns in the frontal median plane [11] with the coefficients related to individual pinna contours were recently proposed by the authors. The effectiveness of the two models in individually rendering distance and elevation, whose respective spatial cues are known to be roughly decoupled [12], was subsequently verified through listening tests [13, 14].

In this paper we focus instead on the remaining spatial dimension, azimuth, and on how to relate a spherical head model to individual anthropometry in order to minimize spectral differences between individual and modeled localization cues. Section 2 reports the motivations and background lying behind this study. Section 3 describes an optimization procedure designed with the aim of tuning the radius of the head model onto measured HRTFs of a public database, and Section 4 reports the derivation and objective analysis of a regression formula relating the three main head dimensions to the sphere radius. Section 5 concludes the paper and traces future developments of the presented research.

2. RESEARCH BACKGROUND

Back in 1907, Lord Rayleigh studied the means through which a listener is able to discriminate the horizontal direction of an incoming sound wave. Following his well-known Duplex Theory of Localization [15], azimuth cues can be reduced to two interaural quantities, i.e.

- *Interaural Time Difference* (ITD), defined as the temporal delay between sound waves at the two ears;
- *Interaural Level Difference* (ILD), defined as the ratio between the instantaneous amplitudes of the same two sounds.

ITD is known to be frequency-independent below 500 Hz and above 3 kHz, with a theoretical ratio of low-frequency ITD versus high-frequency ITD of $3/2$, and slightly variable at middle range frequencies [16]. Conversely, frequency-dependent shadowing and diffraction effects introduced by the human head cause ILD to greatly depend on frequency.

Consider a low-frequency sinusoidal signal (up to 1.5 kHz). Since its wavelength is greater than any head dimension, ITD is reduced to a phase lag $\Delta\varphi < 2\pi$ between the signals arriving at the ears [17]. For this reason ITD is seen as a robust cue for horizontal perception in the low-frequency range. Conversely, ILD is not a robust cue because low frequency components trespass the head without causing significant attenuation on the opposite side with respect to the source. Specularly, a high-frequency sinusoidal signal (above 1.5 kHz) yields an ITD that is greater than a period. Being the human ear phase-sensitive only, ITD turns out to be useless in the high-frequency range, apart from detection of sound onsets. Nevertheless, the considerable shielding effect of the human head on high-frequency waves makes ILD the most relevant cue in such spectral range.

Still, the information provided by ITD and ILD can be ambiguous. If one assumes a spherical geometry of the human head, sound sources located at all possible points of a conic surface pointing towards the ear produce the same ITD and ILD values. These surfaces are known as *cones of confusion* and represent a potential hump for accurate perception of sound direction. In practice, ITD and ILD will not be identical at these two azimuth angles because

1. the human head is clearly not spherical;
2. all subjects exhibit slight asymmetries with respect to the median plane;
3. ear canals lie below and behind the horizontal axis [18].

Nonetheless their values will be very similar, and *front-back confusion* is in fact often observed experimentally [19]. Indeed, despite its rough and simplistic geometry, the spherical head model is the most used model of the head in the literature, and provides an excellent approximation to the magnitude of a pinnaless HRTF. Mokhtari *et al*. [20] highlighted that there is roughly no difference between the numerically simulated responses of an unmodified KEMAR head and of a head shape morphed towards a sphere in the median plane.

In the spherical head model each considered spatial location of the sound source is specified through two coordinates: the *incidence angle* α, i.e. the angle between rays connecting the center of the sphere to the source and the observation point, and the distance r between the source and the center of the sphere. Having defined *normalized frequency* as

$$\mu = f\frac{2\pi a}{c}, \tag{1}$$

where c is the speed of sound[1] and a is the sphere radius, and *normalized distance* as

$$\rho = \frac{r}{a}, \tag{2}$$

the theoretical transfer function of the sphere between source and observation point (which we refer to as *spherical transfer function, STF*) can be described as follows, for each $\rho > 1$ [21]:

$$STF(\mu, \alpha, \rho) = -\frac{\rho}{\mu}e^{-i\mu\rho}\sum_{m=0}^{\infty}(2m+1)P_m(\cos\alpha)\frac{h_m(\mu\rho)}{h'_m(\mu)}, \tag{3}$$

where P_m and h_m represent, respectively, the *Legendre polynomial* of degree *m* and the *m*th-order *spherical Hankel function*. Despite the infinite sum in Eq. (3) and the high computational costs of Hankel functions and Legendre polynomials, an approximation algorithm was proposed [22] where both functions are computed iteratively, allowing a relatively fast evaluation. A first-order filter approximation of the STF for $r = \infty$ was proposed by Brown and Duda [8].

Typically, in a spherical head model the two observation points (i.e. the ear canals) are assumed to be diametrically opposed, such that a linear correspondence between incidence angles ($\alpha_{(l)}$ and $\alpha_{(r)}$ for the right and left ears, respectively) and the azimuth angle θ exists in the horizontal plane. However, if a more realistic geometry is considered in which the ear canal points are displaced backwards and downwards by a certain offset, the model provides a better approximation to elevation-dependent patterns both in the frequency and time domains [18]. Also, notice that the STF is a function of the head radius a, the only parameter that can be tuned on the listener. In [23], a procedure was proposed for selecting the optimal sphere radius defined as the one that minimizes ITD differences in a least-square sense with respect to individual anthropometric measures. The optimal radius a_{itd} is a linear combination of head width w_h, height h_h, and depth d_h:

$$a_{\text{itd}} = 0.26w_h + 0.01h_h + 0.09d_h + 3.2 \ \text{ cm}. \tag{4}$$

This result highlights how head height is a relatively weak parameter in ITD definition with respect to head width and depth.

However, in the literature there is evidence that the spherical head model is not accurate in predicting ITD, being the latter variable by as much as 18% of the maximum interaural delay on a cone of confusion [24]. Such an evidence led researchers to consider ellipsoidal head models accounting for such ITD variation, even though not providing any analytical solution for the

[1] Considering dry-air conditions at 20°C temperature, $c = 343.2$ m/s.

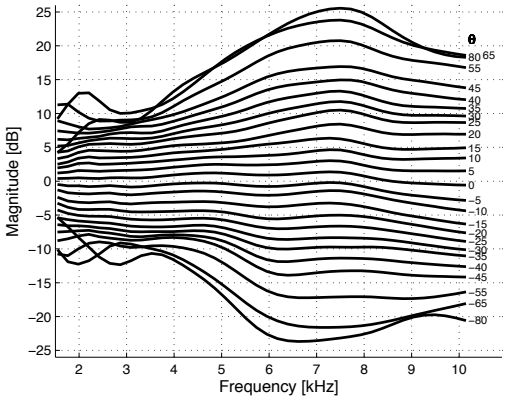

Figure 1: Mean experimental ILDs of CIPIC Subject 021.

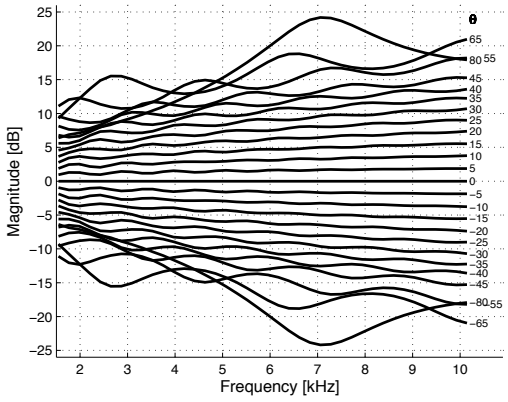

Figure 2: ILDs of a spherical head with radius $a_0 = 10$ cm.

ellipsoid-related transfer function. Consequently, the spherical model should only be used as a filtering element decoupled from ITD, which can in turn be modeled separately as a delay line.

3. ILD-BASED RADIUS OPTIMIZATION

Since modeling the correct individual ILD has been reported to be critical for horizontal localization accuracy, and in particular for resolving front/back reversals [25], we now propose an alternative method to estimate the radius of a spherical head model from individual anthropometric measurements based on ILD, rather than ITD, information. Our reference data set is provided by the CIPIC HRTF database [26], which includes a spatially dense set of HRTFs[2] of 45 subjects measured at $r = 1$ m and a wide range of anthropometric measurements for 37 of them. The following analysis is performed on the HRTFs of these 37 subjects.

Experimental ILDs for each subject and each spatial location (θ, ϕ) are computed in the range $f \in [1.5, 10]$ kHz as the difference between the log magnitudes of the right and left HRTFs smoothed with a constant-Q Gaussian filter with $Q = 5$. The choice of the above frequency range is due to the perceptual irrelevance of the ILD cue below 1.5 kHz, as previously discussed, and to its complex behaviour above 10 kHz due to pinna reflections. Auditory filtering guarantees that the overall ILD envelope is preserved while high-frequency spectral details - irrelevant for ILD perception because of the restricted resolution of the auditory system in the high-frequency region - are smoothed out. Subsequently, in order to discard major elevation-dependent spectral cues, experimental ILDs are averaged on each cone of confusion, i.e. across spatial locations sharing the same azimuth value θ. The resulting average experimental ILDs are formally defined as

$$ILD_{exp}(f, \theta, S_i) = \frac{1}{N_\phi} \sum_{\phi} \frac{|HRTF_r(f, \theta, \phi)|}{|HRTF_l(f, \theta, \phi)|} \qquad (5)$$

[2]Taking as reference an interaural polar coordinate system, azimuth is sampled at $-80°$, $-65°$, $-55°$, from $-45°$ to $45°$ in steps of $5°$, at $55°, 65°$, and $80°$, with positive azimuth values indicating the right hemisphere. Elevation uniformly ranges between $-45°$ and $230.625°$ in steps of $5.625°$, with positive elevation values indicating sources above the horizontal plane.

for each available θ, where $HRTF_r$ and $HRTF_l$ are the right and left smoothed HRTFs of subject S_i, respectively, and N_ϕ is the number of different elevation values (in our case $N_\phi = 50$). These are plotted in Fig. 1 for a representative subject (Subject 021, KEMAR mannequin with small pinnae).

Spherical ILDs for the same azimuth values and the same frequency points as in experimental ILDs are then evaluated through Eq. (3) for different sphere radii ($a \in [0.05, 0.15]$ m) as

$$ILD_{sph}(f, \theta, a) = \frac{|STF(f\frac{2\pi a}{c}, \alpha_r, \frac{1}{a})|}{|STF(f\frac{2\pi a}{c}, \alpha_l, \frac{1}{a})|}, \qquad (6)$$

assuming the ears to lie on the interaural diameter at $\theta = 90°$ (right ear) and $\theta = -90°$ (left ear) so that azimuth θ uniquely defines the α values for the right and left ears as $\alpha_r = |90° - \theta|$ and $\alpha_l = |-90° - \theta|$. Figure 2 reports spherical ILDs for a representative radius, $a_0 = 10$ cm.

If we compare Figs. 1 and 2 we can identify common behaviours of experimental and spherical ILDs, such as increasing direction-dependent differences at high frequencies with respect to low frequencies and a rippled trend in ILDs for lateral angles. In contrast to the perfect left/right symmetry of the spherical ILDs, systematical asymmetries between the left and right hemisphere are observed in most experimental ILDs, see e.g. the nonzero ILD for $\theta = 0°$ in Fig. 1. These are mainly caused by the asymmetry of the human head itself, and especially of the component that has the greatest impact on HRTF measurements, i.e., the pinna. However, asymmetries were also observed in a set of HRTFs of a pinnaless KEMAR mannequin measured with the same apparatus and procedure as the CIPIC database [27]. Thus, measurement noise possibly due to a non-ideal alignment of the measurement system or different positionings of the binaural microphones also has an impact on the found asymmetries. In order to cope with this limitation, that was found to highly influence the subsequent radius optimization step (differences on the order of centimeters were found between optimal radii for left and right directions in several subjects), we define the *asymmetry index* of subject S_i as

$$\gamma(S_i) = \frac{1}{N_f N_\theta} \sum_f \sum_\theta 20 \log_{10} ILD_{exp}(f, \theta, S_i), \quad [\text{dB}] \quad (7)$$

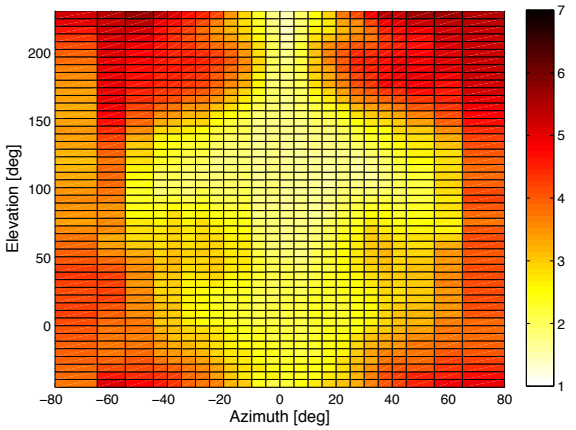

Figure 3: Average spectral distortion [dB] between experimental ILDs and spherical ILDs with individual ILD-optimized radius.

where N_f is the number of frequency bins in the $[1.5, 10]$ kHz frequency range (in our case $N_f = 40$) and N_θ is the number of different azimuth values (in our case $N_\theta = 25$), and use it as a constant normalization factor on the experimental ILDs themselves:

$$\widehat{ILD}_{exp}(f,\theta,S_i) = \frac{ILD_{exp}(f,\theta,S_i)}{10^{\frac{\gamma(S_i)}{20}}}. \tag{8}$$

Now, given a subject S_i and a fixed azimuth θ, we solve the following optimization problem (spectral distortion minimization) for radius a:

$$\min_a \sqrt{\frac{1}{N_f}\sum_f \left(20\log_{10}\frac{\widehat{ILD}_{exp}(f,\theta,S_i)}{ILD_{sph}(f,\theta,a)}\right)^2}, \tag{9}$$

giving the optimal value $\hat{a}(\theta, S_i)$. Since different optimal values result for different θ values, we choose to take the average of the optimal values for the two most lateral azimuth angles, $\theta = -80^\circ$ and $\theta = 80^\circ$, as the *ILD-optimized radius* of subject S_i:

$$a_{\text{opt}}(S_i) = \frac{\hat{a}(-80,S_i) + \hat{a}(80,S_i)}{2}. \tag{10}$$

This choice is due to the facts that (1) the largest individual ILD variations are observed for lateral angles, and (2) as the source approaches the median plane spherical ILDs for different radius values become undistinguishable.

Figure 3 reports the spectral distortion (SD) between experimental ILDs and spherical ILDs with individual ILD-optimized radius for all source positions, averaged on the 37 considered subjects. Along azimuth the SD grows as the source drifts away from the median plane, as expected. Along elevation the SD is generally lower for sources above the head and greater for sources below and behind; this can be attributed to the lack of pinna cues (peaks and notches) in HRTFs for the above locations as opposed to a rich spectral structure when the source is below the horizontal plane [28]. Furthermore, the head's scattering behaviour is most similar to that a sphere when the sound source is above, as the front wave reaching the ears does not encounter facial features or shoulders/torso. Nevertheless, if we consider all source locations, the average SD is less than 4 dB in 77.5% of them. Considering the previously discussed left/right asymmetries which produce individual asymmetry indices as large as $\gamma(S_i) = 3$ dB, this result denotes a close correspondence between experimental and spherical ILDs.

4. RADIUS ESTIMATION FROM ANTHROPOMETRY

When individual HRTFs are not available, so that the ILD of the listener is unknown, a method to estimate the individual head radius from anthropometry is required. Previous literature suggests taking an average head radius $a_{\text{avg}} = 8.75$ cm [29], or half the head width [30], or a weighted sum of the three head dimensions [23]. Since this data is available for the 37 analyzed CIPIC subjects, the ILD-optimized radii can be related to anthropometric parameters through an empirical regression formula derived using the statistics of the population. Multiple linear regression between the three head dimensions and the ILD-optimized radii of all of them was performed, yielding the following regression equation:

$$a_{\text{ild}} = 0.41w_h - 0.15h_h + 0.2d_h + 4.2 \ \text{ cm}. \tag{11}$$

Notice the similarity between this regression equation and the previous Eq. (4) obtained from ITD-optimized radii by Algazi *et al.* [23]. Even though the sign of h_h is negative due to interactions among variables (a linear model with h_h only as regressor would have yielded a coefficient close to zero), in both cases width w_h has the highest coefficient denoting its prominence amongst head dimensions, and a large constant term appears. Indeed, individual radii of the 37 CIPIC subjects estimated through the two different equations show a Pearson correlation coefficient of 0.93. However, Eq.(11) yields radii that are on average 2 cm larger than those predicted through Eq. (4).

Figure 4 provides a deeper insight into different anthropometric radius estimation methods. These include a_{ild} (estimated from Eq. (11)), a_{itd} (estimated from Eq. (4)), the head half-width (i.e. unitary weight to the interaural axis) $a_{\text{wid}} = w_h/2$, and the average of half-head dimensions (i.e. equal weight to the three head axes)

$$a_{\text{eqx}} = \frac{1}{3}\frac{w_h}{2} + \frac{1}{3}\frac{h_h}{2} + \frac{1}{3}\frac{d_h}{2}. \tag{12}$$

The four estimation methods are compared to the average head radius value a_{avg} and to the ILD-optimized values a_{opt} of the 37 CIPIC subjects ordered by increasing head width. From this plot we can notice that

- a_{ild} always has the highest value, comparable on average to a_{opt} because of the previous regression step;
- a_{itd} and a_{eqx} have intermediate values, comparable to the average head radius a_{avg};
- a_{wid} always has the lowest value.

Also notice that a_{opt} scores three particularly high values for subjects 22, 29, and 36. This result indicates that these three subjects present particularly high lateral ILDs, despite their anthropometric measures being not significantly different from those of other subjects in the database. Other factors may have contributed to such a high ILD, e.g. the presence of hair or measurement errors. Since these remain unknown, we chose to act conservatively by keeping these subjects in the regression analysis rather than considering them as outliers.

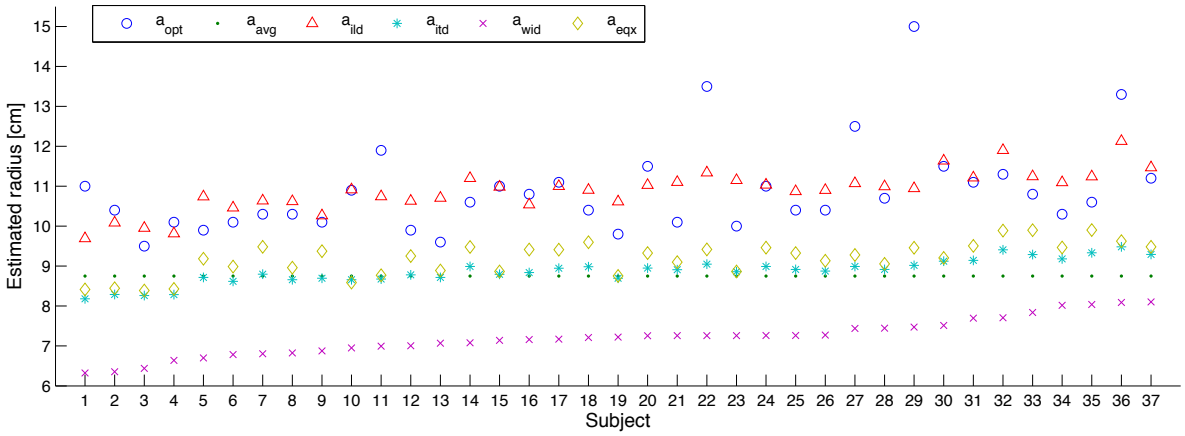

Figure 4: Comparison of different methods for estimating head radius from anthropometric measurements. See text for details.

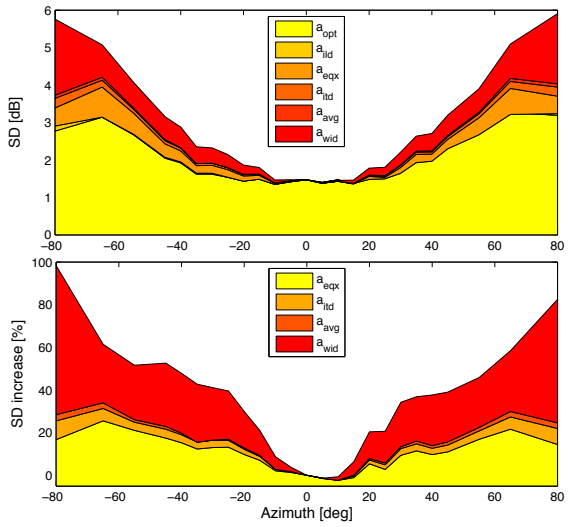

Figure 5: Top panel: average spectral distortion [dB] between experimental ILDs and spherical ILDs with six different radii. Bottom panel: relative SD increase with respect to radius a_{ild}.

In order to quantitatively assess the effectiveness of our estimation method with respect to the others, Fig. 5 provides the average SD between the 37 subjects' experimental and customized ILDs on each cone of confusion for the six defined radii (top panel) and the relative SD increase between radius a_{ild} and each of the other four direct estimation methods (a_{avg}, a_{itd}, a_{wid}, and a_{eqx}, bottom panel). Here we can see that the average SD provided by a_{ild} ranges between 3 dB for the most lateral locations and 1.5 dB for medial locations, and is almost identical to that provided by a_{opt}, attesting the success of our regression. Furthermore, the other four estimation methods score a higher SD for all azimuth angles except for some points near the median plane, where differences among spherical ILDs and thus among radius estimation methods become negligible, and asymmetries of experimental measurements emerge. A within-subjects one-way analysis of variance[3] with radius estimation method as factor was performed on the SD data for each azimuth angle separately, confirming that there is no statistical difference between a_{ild} and a_{opt} for any angle, and that a_{ild} scores significantly lower SD values than any of the other estimation methods in the azimuth ranges $[-80°, -45°]$ and $[45°, 80°]$.

In particular, the average SD of the other estimation methods at these lateral locations is between 15% and 35% higher than that of a_{ild}, with a_{wid} scoring a remarkable relative SD increase lying between 60% and 100% for the most lateral azimuth angles. The bad results associated to radius a_{wid} are in accordance with a previous remark by Katz [30], stating that a sphere with the same interaural distance (i.e., width) as the head is a worse approximation of its acoustical behaviour than a larger sphere with the same volume as the head. Our results on the ILD-optimized sphere, which has in general the largest radius amongst all others, further support his findings.

5. CONCLUSIONS AND PERSPECTIVES

Summing up, we obtained a closed formula relating the three most straightforward anthropometric parameters of the head to the radius of a spherical head model, starting from an analysis of a set of ILDs derived from a public HRTF database. The model does not rely on the use of measured HRTFs, thus allowing a fully synthetic rendering through a low-order filter structure [8], but does not account for the correct ITD. If coupled with an ITD model based on a delay line, the design of an all-pass section counterbalancing the effect that the head filter's phase response has on ITD is necessary.

The proposed estimation method objectively offers better ILD estimations than previous methods proposed in the literature. Although further research is needed in order to assess the effective-

[3] Homoscedasticity of the data set was verified through Levene's test. Mauchly's test was instead used to check data sphericity; in all cases where this test indicated a violation of sphericity, degrees of freedom were adjusted using a Greenhouse-Gasser epsilon correction. The significance level for the data analysis was set to 0.05.

ness of the customized spherical model in rendering the azimuth of a virtual sound source through subjective tests or auditory modeling, the found results already provide substance towards the development and evaluation of structural HRTF models for binaural virtual acoustics.

6. REFERENCES

[1] J. W. Strutt, "On the acoustic shadow of a sphere," *Phil. Trans.*, vol. 203, pp. 87–110, 1904.

[2] C. I. Cheng and G. H. Wakefield, "Introduction to head-related transfer functions (HRTFs): Representations of HRTFs in time, frequency, and space," *J. Audio Eng. Soc.*, vol. 49, no. 4, pp. 231–249, April 2001.

[3] G. Plenge, "On the differences between localization and lateralization," *J. Acoust. Soc. Am.*, vol. 56, no. 3, pp. 944–951, September 1974.

[4] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization using nonindividualized head-related transfer functions," *J. Acoust. Soc. Am.*, vol. 94, no. 1, pp. 111–123, July 1993.

[5] H. Møller, M. F. Sørensen, C. B. Jensen, and D. Hammershøi, "Binaural technique: Do we need individual recordings?" *J. Audio Eng. Soc.*, vol. 44, no. 6, pp. 451–469, June 1996.

[6] E. C. Durant and G. H. Wakefield, "Efficient model fitting using a genetic algorithm: Pole-zero approximations of HRTFs," *IEEE Trans. Speech Audio Process.*, vol. 10, no. 1, pp. 18–27, January 2002.

[7] Q. Huang and L. Li, "Modeling individual hrtf tensor using high-order partial least squares," *EURASIP J. Adv. Signal Process.*, vol. 2014, no. 58, pp. 1–14, May 2014.

[8] C. P. Brown and R. O. Duda, "A structural model for binaural sound synthesis," *IEEE Trans. Speech Audio Process.*, vol. 6, no. 5, pp. 476–488, September 1998.

[9] V. R. Algazi, R. O. Duda, R. P. Morrison, and D. M. Thompson, "Structural composition and decomposition of HRTFs," in *Proc. IEEE Work. Appl. Signal Process., Audio, Acoust.*, New Paltz, New York, USA, October 2001, pp. 103–106.

[10] S. Spagnol, M. Geronazzo, and F. Avanzini, "Hearing distance: A low-cost model for near-field binaural effects," in *Proc. EUSIPCO 2012 Conf.*, Bucharest, Romania, September 2012, pp. 2005–2009.

[11] ——, "On the relation between pinna reflection patterns and head-related transfer function features," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 3, pp. 508–519, March 2013.

[12] S. Spagnol, "On distance dependence of pinna spectral patterns in head-related transfer functions," *J. Acoust. Soc. Am.*, vol. 137, no. 1, pp. EL58–EL64, January 2015.

[13] S. Spagnol and F. Avanzini, "Distance rendering and perception of nearby virtual sound sources with a near-field filter model," 2015, submitted for publication.

[14] S. Spagnol, M. Geronazzo, D. Rocchesso, and F. Avanzini, "Synthetic individual binaural audio delivery by pinna image processing," *Int. J. Pervasive Comput. Comm.*, vol. 10, no. 3, pp. 239–254, July 2014.

[15] J. W. Strutt, "On our perception of sound direction," *Phil. Mag.*, vol. 13, pp. 214–232, 1907.

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

[17] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, 2nd ed. Cambridge, MA, USA: MIT Press, October 1996.

[18] V. R. Algazi, C. Avendano, and R. O. Duda, "Elevation localization and head-related transfer function analysis at low frequencies," *J. Acoust. Soc. Am.*, vol. 109, no. 3, pp. 1110–1122, March 2001.

[19] A. W. Bronkhorst, "Localization of real and virtual sound sources," *J. Acoust. Soc. Am.*, vol. 98, no. 5, pp. 2542–2553, November 1995.

[20] P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato, "Acoustic simulation of KEMAR's HRTFs: Verification with measurements and the effects of modifying head shape and pinna concavity," in *Proc. Int. Work. Princ. Appl. Spatial Hearing (IWPASH 2009)*, Zao, Miyagi, Japan, November 2009.

[21] W. M. Rabinowitz, J. Maxwell, Y. Shao, and M. Wei, "Sound localization cues for a magnified head: Implications from sound diffraction about a rigid sphere," *Presence*, vol. 2, no. 2, pp. 125–129, Spring 1993.

[22] R. O. Duda and W. L. Martens, "Range dependence of the response of a spherical head model," *J. Acoust. Soc. Am.*, vol. 104, no. 5, pp. 3048–3058, November 1998.

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

[24] 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, Signal Process. (ICASSP'99)*, Phoenix, AZ, USA, March 1999, pp. 965–968.

[25] K. Watanabe, K. Ozawa, Y. Iwaya, Y. Suzuki, and K. Aso, "Estimation of interaural level difference based on anthropometry and its effect on sound localization," *J. Acoust. Soc. Am.*, vol. 122, no. 5, pp. 2832–2841, November 2007.

[26] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," in *Proc. IEEE Work. Appl. Signal Process., Audio, Acoust.*, New Paltz, New York, USA, October 2001, pp. 1–4.

[27] V. R. Algazi, Private communications, 2010.

[28] S. Spagnol, M. Hiipakka, and V. Pulkki, "A single-azimuth pinna-related transfer function database," in *Proc. 14th Int. Conf. Digital Audio Effects (DAFx-11)*, Paris, France, September 2011, pp. 209–212.

[29] R. V. L. Hartley and T. C. Fry, "The binaural location of pure tones," *Phys. Rev.*, vol. 18, no. 6, pp. 431–442, December 1921.

[30] B. F. G. Katz, "Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation," *J. Acoust. Soc. Am.*, vol. 110, no. 5, pp. 2440–2448, November 2001.