REAL-TIME BINAURAL AUDIO RENDERING IN THE NEAR FIELD

Simone Spagnol
University of Padova
spagnols@dei.unipd.it

Federico Avanzini
University of Padova
avanzini@dei.unipd.it

ABSTRACT

This paper considers the problem of 3-D sound rendering in the near field through a low-order HRTF model. Here we concentrate on diffraction effects caused by the human head which we model as a rigid sphere. For relatively close source distances there already exists an algorithm that gives a good approximation to analytical spherical HRTF curves; yet, due to excessive computational cost, it turns out to be impractical in a real-time dynamic context. For this reason the adoption of a further approximation based on principal component analysis, which can significantly speed up spherical HRTF computation, is proposed. The model resulting from such an approach is suitable for future integration in a structural HRTF model and parameterization over anthropometrical measurements of a wide range of subjects.

1 INTRODUCTION

The history of binaural 3-D sound rendering dates back to Lord Rayleigh's well known diffraction formula which approximates the behaviour of a sound wave produced by an infinite point source around the listener's head, thus providing a first crude sketch of what we today call a head-related transfer function (HRTF). On the other hand, most of the relevant issues in this field appeared only recently.

HRTF-based spatial audio rendering can be achieved in multiple ways. Approximations based on low-order rational functions (see e.g. [4]) and series expansions of HRTFs [5, 9] were proposed, resulting in simple yet valuable tools for diffraction modeling. Nevertheless, significant computation is required from both techniques when real-time constraints are introduced, due to the complexity of filter coefficients and weights respectively. This is why structural modeling [2] seems nowadays to be an attractive alternative approach. Within this framework, the contribution of the listener's head, ears and torso to the HRTF can be isolated in several subcomponents, each accounting for some well defined physical phenomenon. Due to linearity of all these effects, they can be later combined meaningfully and realistically in an additive fashion to result in a global HRTF. Such a decomposition yields a model which is both economical and well suited to real-time implementations.

In this paper we will conceptually isolate the earless head of the listener and treat it as a rigid sphere, trying to find a suitable way to represent its contribution to the HRTF. Henceforward we will relate to its transfer function by calling it a spherical HRTF. Furthermore, we will concentrate on sources located in the so-called near field – namely within a few meters from the center of the head – for which real-time computation of HRTFs turns to be more troublesome. Section 2 briefly introduces the theory lying behind the problem. Then, Section 3 presents a PCA-based approach for spherical HRTF modeling. Section 4 deals with the problem of efficient filter modeling. Finally, Section 5 concludes with a discussion on the further work to be done in this direction.

SMC 2009, July 23-25, Porto, Portugal

2 THE SPHERICAL HRTF

2.1 Analytical background

Within the assumption of an infinitely distant source from the center of the head, we can describe the response related to a fixed observation point on the sphere's surface by means of the following transfer function, based on Lord Rayleigh's diffraction formula [1]:

$$H(\mu,\theta) = \frac{1}{\mu^2}\sum_{m=0}^{\infty}\frac{(-i)^{m-1}(2m+1)P_m(cos\theta)}{h'_m(\mu)}, \quad (1)$$

where θ is the incidence angle, the angle between the ray from the center of the sphere to the source and the ray to the observation point, and μ is the normalized frequency, defined as [2]

$$\mu = f\frac{2\pi a}{c}, \quad (2)$$

which is directly proportional to the sphere radius a. Figure 1 shows the magnitude of the transfer function on a dB scale against normalized frequency for 19 different values of incidence angle. When we remove the assumption of an infinitely distant source and consider source positions in

[1] Here P_m and h_m represent, respectively, the *Legendre polynomial* of degree *m* and the *m*th-order *spherical Hankel function*. h'_m is the derivative of h_m with respect to its argument.

[2] Parameter *c* is the ambient speed of sound.

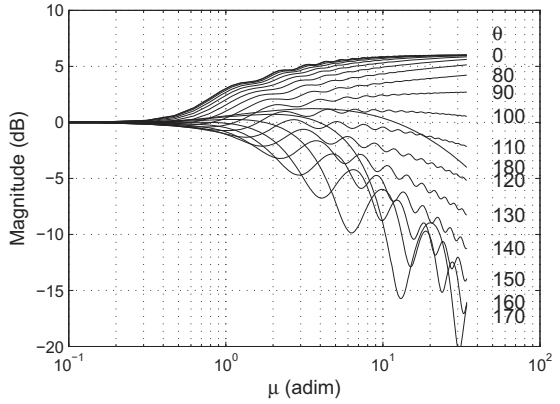

Figure 1. Magnitude response for an infinitely distant source.

the near field, the distance dependence can no longer be ignored. Having defined the normalized distance to the source ρ as the ratio between the absolute distance from the center of the sphere and the sphere radius

$$\rho = \frac{r}{a}, \tag{3}$$

the spherical HRTF can be evaluated by means of the following function [11]:

$$H(\rho, \mu, \theta) = -\frac{\rho}{\mu} e^{-i\mu\rho} \sum_{m=0}^{\infty} (2m+1) P_m(cos\theta) \frac{h_m(\mu\rho)}{h'_m(\mu)}, \tag{4}$$

for each $\rho > 1$. From the analysis of this function we can state a fundamental characteristic of spherical HRTFs: as the source approaches the sphere (ρ tends to 1) the response on the ipsilateral side increases, while the response on the contralateral side decreases [3]. A description of the evaluation algorithm, based on recursion relations, can be found in [8].

2.2 Real-time computation

Let us consider a scenario where the listener is free to move his head with respect to the virtual source to be rendered, and vice versa. It is clear that real-time computation of HRTFs is needed in order to track these movements with enough reactivity, possibly avoiding any discontinuity in the resulting sound. Furthermore, we have to take into account the possibility of having to simulate a complex acoustic environment that includes several independent sound sources, and/or reflections coming from the environment.

Relatively simple HRTF filter structures for sources in the far field have been proposed to date (e.g., Duda's first-order filter in [2]). These turn out to be impracticable in the near field, having no parameterization on source distance. Point-to-point real-time evaluation of Eq. (4) using the algorithm in [8] is computationally still too expensive. Moreover, even if a suitable parameterized filter model is found each source has to be processed with a separate filter. Thus we need to introduce a proper HRTF approximation to speed up the computation. In the next section we discuss such an approximation, which makes use of Principal Component Analysis (PCA) to represent a collection of sample analytical HRTFs.

3 A PCA-BASED APPROACH

3.1 Principal Component Analysis

Principal Component Analysis is used in a number of problems to reduce the dimensionality of an input data set. Its main goal is to provide an efficient representation of a set of correlated measures - in this instance, a set of vectors.

Without delving into deep technicalities (which can be found in [7]), suffice it to say that given a set of n real-valued vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$, each of dimension d, and defining its *covariance matrix* $\mathbf{S}$ as

$$\mathbf{S} = \frac{1}{n} \sum_{k=1}^{n} \mathbf{x}_k \mathbf{x}_k^t, \tag{5}$$

it can be seen that the best p-dimensional representation (with $p \leq d$) of the data set is obtained by taking as basis vectors the p eigenvectors of $\mathbf{S}$ that correspond to the p largest eigenvalues. [3] Each vector $\mathbf{x}_k$ is then projected onto the space defined by the basis vectors as follows:

$$\mathbf{a}_k = \mathbf{C}^t \mathbf{x}_k, \tag{6}$$

where $\mathbf{C}$ is a matrix, the columns of which are the basis vectors. We call principal components the set of weights $\{a_{ki}\}$, $k = 1, \ldots, n$, associated to basis vector i. Now given the set of p-dimensional vectors $\mathbf{a}_k$, $k = 1, \ldots, n$, we can reconstruct an estimate of each original data vector by the inverse equation:

$$\mathbf{x}_k = \mathbf{C}\mathbf{a}_k. \tag{7}$$

Clearly, by increasing the dimension p of the representation the approximation improves. Thus, when dealing with PCA, the main design goal is to extrapolate the value p for which the trade-off between accuracy and data dimensionality is maximized.

PCA has already been used in previous works concerning HRTF modeling [5, 9], with the vectors $\mathbf{x}_k$ representing

[3] An alternative formulation of PCA requires the mean of all vectors in the data set to be subtracted from each one of them before constructing the covariance matrix. However, as the data set we will take into consideration is already well-centered, inclusion of the mean turns out to be quite unnecessary.

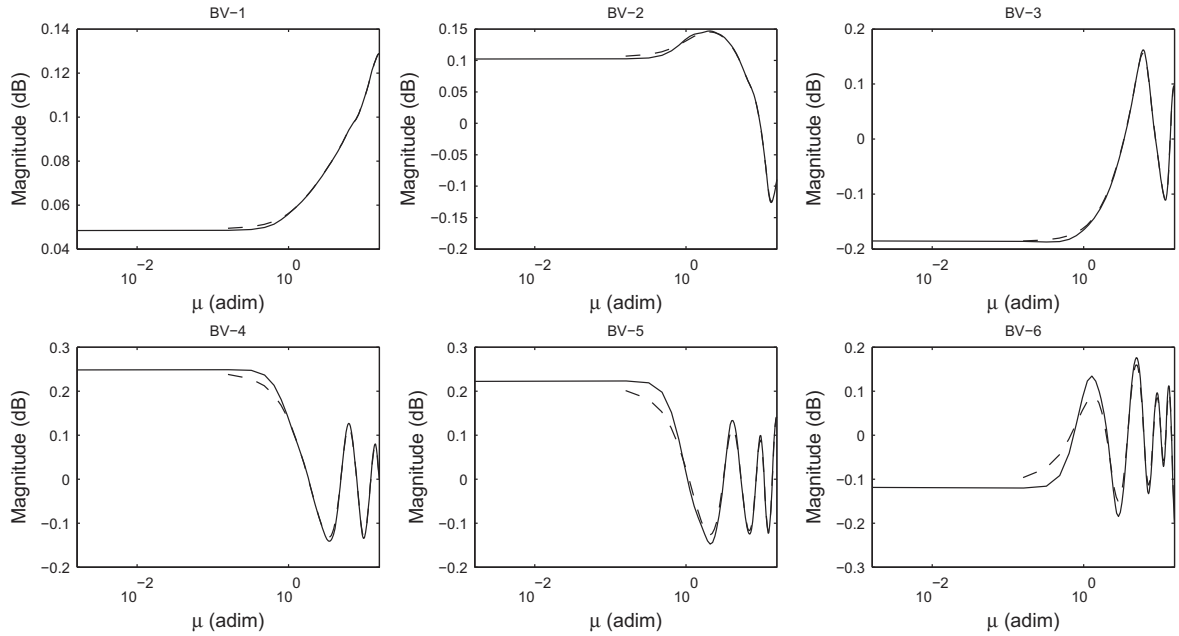

Figure 2. The first six basis vectors (solid lines) and the corresponding least-squares fit 8-th order IIR filters (dashed lines).

magnitude responses of a set of measured HRTFs. However, instead of applying the technique to experimental data, we will exploit it to approximate a collection of spherical HRTF magnitudes sampled from Eq. (4) on a discrete set of frequencies. We will show that, thanks to the decoupling of spatial variables from frequency created by PCA, this approach provides significant computational and storage advantages in the modeling of spherical HRTFs.

3.2 Design choices

We choose to collect a set of spherical HRTFs for sound sorces located at different distances and incidence angles with respect to the ear canal. Being Eq. (4) dependent on only two spatial parameters, in our polar coordinate system we do not consider elevation and restrict these locations to points lying on the horizontal plane. We conventionally assume θ to be the incidence angle at the right ear canal. Therefore $\theta = 0°$, $\theta = 90°$, and $\theta = 180°$ corresponds to a sound source facing the right ear, in front of the head, and facing the left ear, respectively. The set of spherical HRTFs is sampled by fixing the head radius to the standard value $a = 8.75$ cm and varying the following parameters:

- 19 linearly spaced θ values, from $0°$ to $180°$, with $10°$ angle increments;
- 7 exponentially spaced distance values, $\rho = 1.25, 1.5, 2, 4, 8, 16, 32$ (with the last one approximating far field);
- 100 linearly spaced frequency points from 100Hz to 10 kHz, with 100 Hz increments.

We obtain a set of $19 \times 7 = 133$ spherical HRTFs, of which we consider only the dB magnitude responses. Indeed, the HRTF for an ideal sphere appears to be minimum phase for all ranges and incidence angles [8]. In addition, when considering interaural differences for binaural hearing, approximated ITD models (e.g. the Woodworth's formula) can be used to simulate phase lag between right and left ear canal as a simple delay line. Interaural Time Difference (ITD) effects can therefore be cascaded to the HRTF synthesis process.

3.3 Application of PCA

At this point we apply PCA to the set of $n = 133$ real-valued vectors $\mathbf{x}_1$,...,$\mathbf{x}_n$, each of dimension $d = 100$. The first 6 basis vectors of the analysis are sketched in Figure 2. As we can see, after the first one which accounts for the general slope of the majority of HRTFs (with a positive weight for ipsilateral sources and a negative weight for contralateral ones – see Figure 3), each successive basis vector introduces more and more ripples in the frequency response, starting from the most prominent at $\theta = 170°$.

By investigating the trend of principal components 2 to 6 with the varying of distance and incidence angle we obtain a deeper insight of the analysis. As expected from the observations reported in Section 2.1, weights' moduli are am-

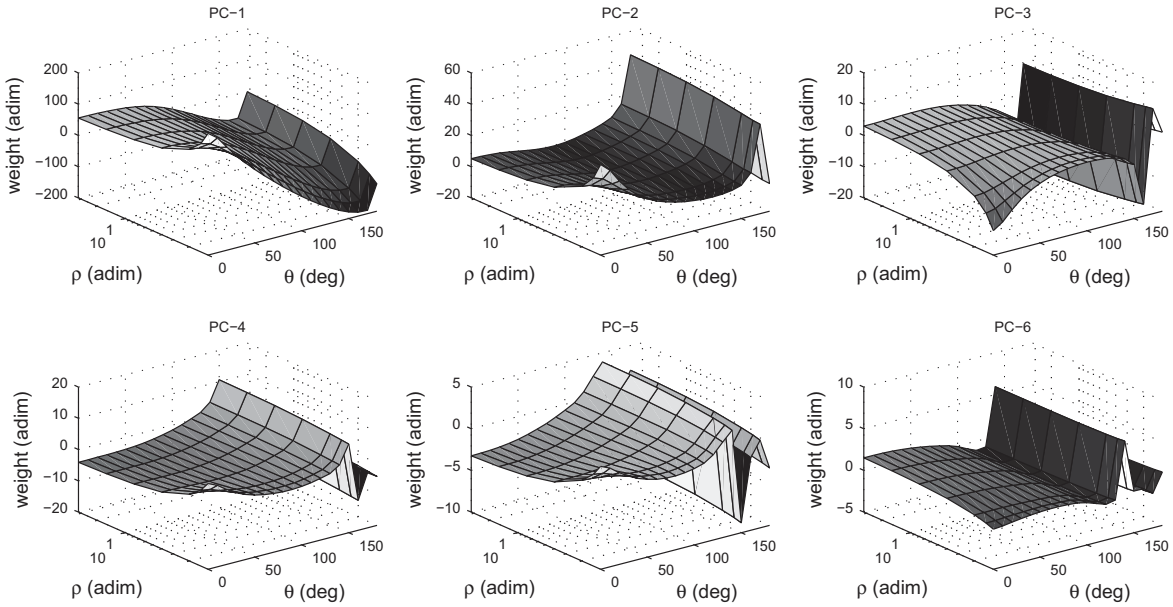

Figure 3. The first six principal components.

plified by decreasing distance; furthermore, Figure 3 shows that each component emphasises its corresponding basis vector only for a limited range of incidence angles, regardless of the distance. This means that the first basis vector retains most of the common variation, while those from the second onwards provide particularized description of the rippled high-frequency behaviour of spherical HRTFs, which varies according to the incidence angle.

3.4 Theoretical optimality

The number of principal components (parameter p) to be included in our model is crucial: as a matter of fact, it denotes the number of filters required to approximate the spherical HRTF by means of the new representation. We need then a proper principle to theoretically quantify the maximum tolerable error, so to extract the minimum p that meets its constraints.

Mills [10] presents a psychoacoustical result which can be used to this purpose. In particular the Interaural Level Difference (ILD) jnd curve as a function of frequency in Figure 4 represents a safe upper bound on the approximation error, owing to insensibility of human hearing apparatus to small changes in ILD (which remarkably denotes the main feature for discriminating source location together with ITD). After having checked that the absolute error between all ILDs derived from a complementary pair of original HRTFs (same distance parameter and sum of incidence

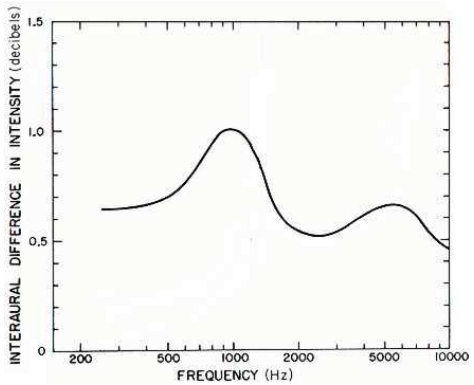

Figure 4. ILD jnd as a function of frequency (figure reproduced from [10]).

angles equal to 180 degrees, assuming diametrically opposite ear canals) and those reconstructed after PCA approximation turns out to lie under the jnd function, we can state there is no significant information loss in our approximation. Note that the jnd function has not been defined for very low frequencies; nevertheless, the dominant localization feature in this frequency range being ITD, ILD information appears to be relevant just for detecting very close distances.

As we can see from Figure 5 the minimum value p for which the total error introduced by the PCA approximation

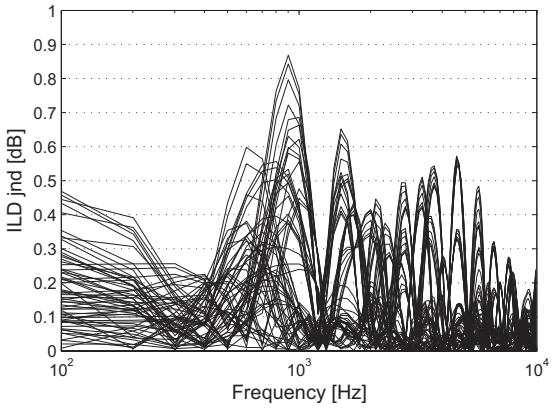

Figure 5. ILD error functions with $p = 7$.

remains below the jnd curve is $p = 7$. In Section 4.2 we will repeat this kind of analysis by including errors due to filter approximation of basis vectors.

4 FILTER REALIZATION

4.1 Filter modeling and interpolation

Each basis vector provides the magnitude response of a filter that has to be realized numerically. Using the least-squares fit procedure provided by the Yule-Walker approach, we design for each basis vector an IIR filter of a desired order, such that it approximates the corresponding magnitude response. In order for the function to work properly we assign a fictional value for zero frequency (we choose this to be the same value as $f = 100$ Hz, as the low-frequency magnitude response is essentially flat) and assume a 20 kHz sampling rate (so that the Nyquist frequency coincides with our 10 kHz limit). Filter coefficients may later be rescaled in case of different sampling rates and different head radii.

It can be seen from Figure 2 that eighth-order filters provide accurate matching of the target magnitude responses. It has to be noted that procedure does not take into account phase requirements. However the resulting filter structures have poles and zeros all inside the unit circle, and are therefore minimum-phase filters.

Having HRTF frequency dependence (now incorporated inside filters characterization) been decoupled from spatial variables dependence, interpolation of spherical HRTFs over spatial points which are not included in the analysis process involves only interpolation of principal components in the form of scalars. To this end, the components plotted in Fig. 3 can be interpolated over distance and incidence angle using simple techniques, e.g 2-dimensional spline interpolation. In particular, in this way any distance value can be rendered (with the upper distance bound in the analysis $\rho = 32$ corresponding to the far field).

Frequency decoupling from spatial variables gives another fundamental advantage. Specifically, the simulation of N independent sound sources located at different positions around the listener head does not require N different filter sets. Instead the set of filters derived above is used for all the sources, with only the components a_i varying for each source. This can be seen in the following equation:

$$\begin{aligned} Y(\mu) &= \sum_{k=1}^{N}\sum_{i=1}^{p} H_{ki}(\rho_k, \mu, \theta_k) X_k(\mu) \\ &= \sum_{k=1}^{N}\sum_{i=1}^{p} H_i(\mu) a_i(\theta_k, \rho_k) X_k(\mu) \\ &= \sum_{i=1}^{p} H_i(\mu) \sum_{k=1}^{N} a_i(\theta_k, \rho_k) X_k(\mu), \end{aligned} \tag{8}$$

where the N input signals, each with frequency response X_k, are linearly combined through spatial coefficients a_i and filtered by the H_i's, resulting in the output signal with frequency response Y. This result, together with the inclusion of distance dependence and near-field effects in the spherical HRTF, represents the main advantage of the proposed approach with respect to the model described in [2].

4.2 Optimality considerations

The filter realization described in the previous section introduces further error between the real-time model and analytical spherical HRTF curves. Hence, in addition to parameter p, choosing the adequate filter orders $o_1, \ldots, o_p$ turns out to be pivotal. To this end, we reapply the ILD jnd criterion in order to determine minimum parameters p and $o_1, \ldots, o_p$ that satisfy the forementioned psychoacoustical constraint.

The analysis must be targeted at finding a satisfactory trade-off between accuracy and efficiency. By keeping the minimum value $p = 7$ determined in Section 3.4, it is verified that eighth-order filters ($o_1 = \ldots = o_p = 8$) provide an error which is still below the jnd curve, while seventh-order filters cause 1 dB low-frequency errors. If p is increased by one or more units, using filters of lower order (e.g., 7) still results in errors which are above the psychoacoustical threshold. Intuitively, this circumstance can be explained as follows. Considering that the very first principal components capture the largest part of variance in the data set and have the corresponding basis functions being multiplied by a relatively high coefficient, adding new principal components does not affect the accuracy of the representation as much as properly designing the filters representing each basis vector. Further inspection shows that, since the magnitude responses H_i become increasingly rippled as i grows, the psychoacustical threshold is satisfied even by choosing filter orders that increase accordingly, i.e. $o_i = i + 1$ ($i = 1 \ldots 7$).

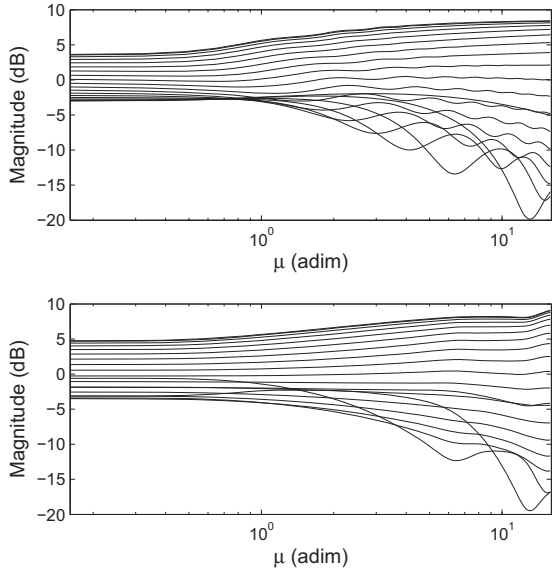

Figure 6. Analytical (top panel) and approximated (bottom panel) spherical HRTF magnitude curves for $p = 3$, $o_1 = o_2 = o_3 = 3$, and $\rho = 4$.

4.3 A low-cost realization

The above discussion is based on purely theoretical assumptions which are very strict. Moreover, the realization proposed in the previous section may have exceedingly high computational costs for real-time applications. In light of this, a more efficient approximation of the spherical HRTF based on a lower number of components and lower-order filters can still be usable even if it does not satisfy the psychoacustical criterion discussed above.

By choosing $p = 3$ and $o_1 = o_2 = o_3 = 3$, the gross magnitude characteristics of the spherical HRTF are still matched, even though the ILD error can be as large as 3 dB at low frequencies. This statement can be verified by looking at Figure 6, which represents reconstructed spherical HRTF magnitude responses for $\rho = 4$ and varying incidence angle. Comparison of the top and bottom panels of the figure confirms that three basis vectors represented with third-order filters already provide a satisfactory approximation.

5 CONCLUSIONS AND FUTURE WORK

In this paper we have presented a PCA-based approach for approximating spherical HRTFs in the near field. We proved that a description in terms of seven eighth-order filters and a set of coefficients turns out to be psychoacustically robust. Much work is still needed in this direction. First, we shall reproduce the analysis in Subsection 3.4 for spatial points that were not included in the synthesis step. Second, the low-cost realization described in Subsection 4.3, possibly along with alternative descriptions, needs to be experimentally evaluated. Third, we need a strong criterion for the personalization of HRTFs based on anthropometrical measurements, analogously to the approach presented in [1]. Finally, we should take into consideration alternative and more realistic head models, like the elliptical one [6].

6 REFERENCES

[1] Algazi, V. R., Avendano, C. and Duda, R. O. “Estimation of a spherical-head model from anthropometry”, *JAES Volume 49, Issue 6, 472-479*, 2001.

[2] Brown, C. P. and Duda, R. O. “A structural model for binaural sound synthesis”, *IEEE Transactions on Speech and Audio Processing, Vol 6, No. 5, 476-488*, 1998.

[3] Brungart, D. S. “Near-field virtual audio displays”, *Presence Vol. 11, No. 1, 93-106*, 2002.

[4] Durant, E. C. and Wakefield, G. H. “Efficient model fitting using a genetic algorithm: pole-zero approximations of HRTFs”. *IEEE Trans. Speech Audio Process., Vol 10, No. 1, 18-27*, 2002.

[5] Chen, J., Van Veen, B. D. and Hecox, K. E. “A spatial feature extraction and regularization model for the head-related transfer function”, *J. Acoust. Soc. Am. 97 (1), 439-452*, 1995.

[6] Duda, R. O., Avendano, C. and Algazi, V. R. “An adaptable ellipsoidal head model for the interaural time difference”, *Proc. IEEE Int. Conf. Acoustics Speech and Signal Processing, Phoenix, AZ, 965-968*, 1999.

[7] Duda, R. O., Hart, P. E. and Stork, D. G. *Pattern Classification: Second Edition*. John Wiley & Sons, New York, 2001.

[8] Duda, R. O. and Martens, W. L. “Range dependence of the response of a spherical head model”, *J. Acoust. Soc. Am. 104 (5), 3048-3058*, 1998.

[9] Kistler, D. J. and Wightman, F. L. “A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction”, *J. Acoust. Soc. Am. 91 (3), 1637-1647*, 1992.

[10] Mills, A. W. “Lateralization of High-Frequency Tones”, *J. Acoust. Soc. Am. 32 (1), 132-135*, 1960.

[11] Rabinowitz, W. M., Maxwell, J., Shao, Y. and Wei, M. “Sound Localization Cues for a Magnified Head: Implications from Sound Diffraction about a Rigid Sphere”, *Presence Vol. 2, 125-129*, 1993.