Binaural Receiver Models for Wave-based Simulations in the Low Frequency Range

Philipp Schäfer, Hark Braren, Janina Fels, Michael Vorländer
Institute of Technical Acoustics, 52074 Aachen, Germany, Email: psc@akustik.rwth-aachen.de

Introduction

A typical example is the simulation of small rooms, such as car cabins [1]. In this context, it is of great interest to get a spatial impression of the room using an auralization of the scene. For this purpose, a binaural receiver can be included in the simulation (see Figure 1).

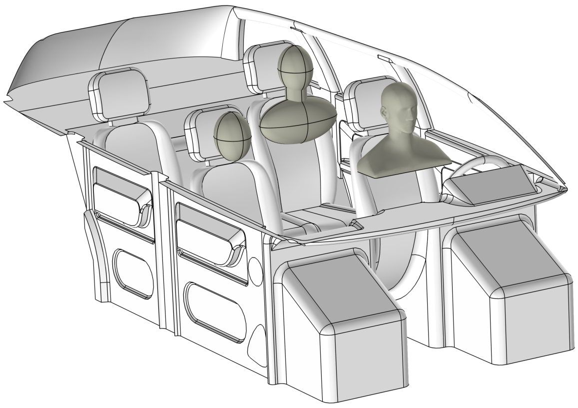

Figure 1: Example of binaural receivers in a car interior.

In a straightforward approach, the geometry of an artificial head can be placed in the room. Nevertheless, such an approach leads to a high density of mesh nodes, especially around fine structures such as ear and pinna, which again results in high computational effort. However, since these simulations are usually limited to the low frequency range ($\ll 4\,\text{kHz}$), the considered wavelengths are generally larger than those details meaning they are acoustically invisible. Thus, a more efficient approach might be the usage of simplified models which are based on simple mathematical bodies, e.g. ellipsoids [2, 3] (see Figure 2). As a result, the computational complexity

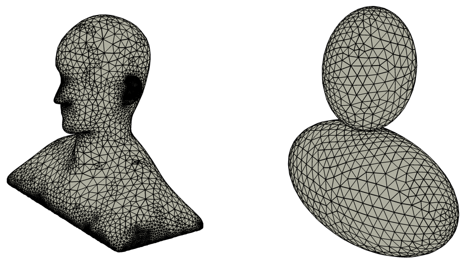

Figure 2: Mesh of an artificial head and a simplified model.

should be reduced while keeping the accuracy of the simulation results on a similar level. An additional advantage of these models is that instead of CAD data, simple parameters can be used to describe the geometry. This also allows an easy individualization of these models.

In the current study, multiple simplified models are derived from an artificial head. For each model as well as the original geometry, an HRTF is simulated using the boundary element method (BEM). Then, the frequency-dependent performance of these models regarding accuracy and required mesh nodes is investigated using the original artificial head as reference.

Simplified models

The binaural receiver models in this study are derived from the *ITA artificial head* [4]. In total, five binaural receiver models are investigated. For all models, the coordinate origin is the center of the interaural axis.

Head models

In the present work, two models that only consider the head geometry are analyzed. The first model, the *Ellipsoid Head*, uses an ellipsoid which is centered at the origin. The ellipsoid dimensions refer to depth, width and height of the reference head. As can be seen in Figure 3,

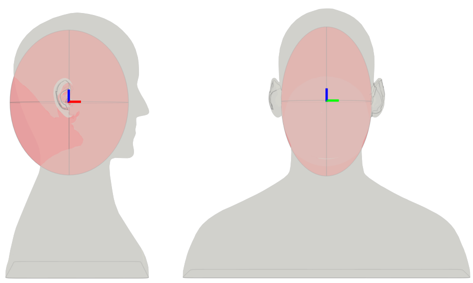

Figure 3: The *Ellipsoid Head* (red) compared to the *ITA artificial head* (gray).

centering the ellipsoid between the ears leads to a great geometrical mismatch. It is known that this mismatch can lead to errors regarding the interaural time difference (ITD) [2]. Thus, a second head, *Ellipsoid Offset Head*, is introduced whose center is shifted towards front and top. Additionally width and depth are increased to better match the original geometry (see Figure 4).

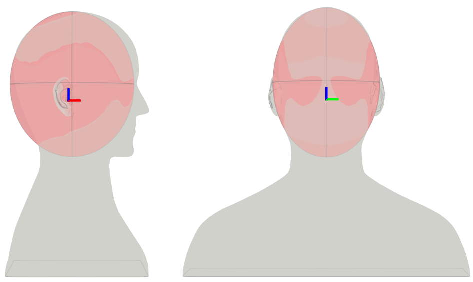

Figure 4: The *Ellipsoid Offset Head* (red) compared to the *ITA artificial head* (gray).

Snowman models

In order to account for the influence of the torso (e.g. shoulder reflections), the *Ellipsoid Head* is extended using a second ellipsoid. This kind of model are referred to as snowman [3]. As shown in Figure 5, the *Ellipsoid Snowman*'s head and torso are tangent to each other. The dimensions of the torso are chosen using a visual fit.

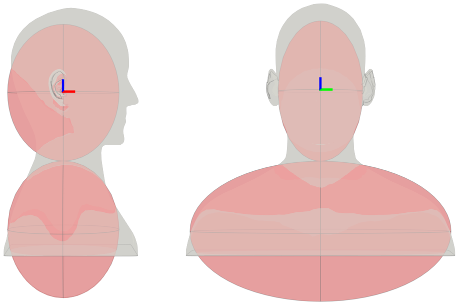

Figure 5: The *Ellipsoid Snowman* (red) compared to the *ITA artificial head* (gray).

By adding the same torso to *Ellipsoid Offset Head*, *Ellipsoid Offset Snowman* is created. It is visible in Figure 6 that due to the head's offset, a gap between head and torso occurs. In order to close this gap, the model is extended by a cylindrical neck leading to *Ellipsoid Offset Snowman with Neck* (see Figure 7). Again, the neck's dimensions are derived using a visual fit. The final geometry is created using a union operation.

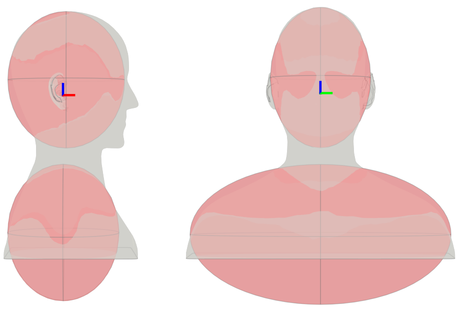

Figure 6: The *Ellipsoid Offset Snowman* (red) compared to the *ITA artificial head* (gray)

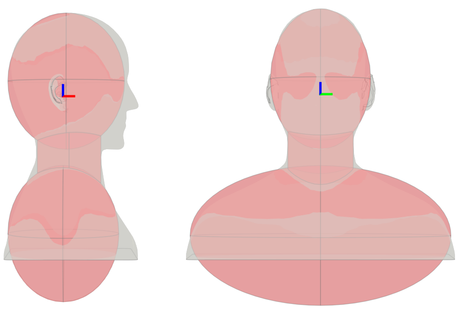

Figure 7: The *Ellipsoid Offset Snowman with Neck* (red) compared to the *ITA artificial head* (gray)

The mathematical bodies used for the for the head, torso and neck of the models are summarized in table 1. In this table, the respective dimensions and the position of their center compared to the center of the interaural axis can be found.

Table 1: Utilized mathematical bodies

	Coordinates [cm]		
Parameter	x (front)	y (left)	z (top)
Head: Centered Ellipsoid			
Semi-axes	9.6	7.2	11.8
Center	0	0	0
Head: Ellipsoid with Offset			
Semi-axes	10.1	8.45	11.8
Center	0.5	0	2.8
Torso: Ellipsoid			
Semi-axes	9.6	22.0	11.8
Center	0	0	-23.6
Neck: Cylinder			
Radius/height	$r_{xy} = 5.5$		$h_z = 10$
Center	1	0	-10

HRTF simulation

All HRTFs are simulated with the BEM of COMSOL Multiphysics 5.4 [5] using 129 bins at a sampling rate of 44.1 kHz. As source, a circular surface with a diameter of 1 cm at the position of the ear canal entrance is excited using a pressure condition (1 Pa) while the results are evaluated on a sphere of 2 m radius using a $1^\circ \times 1^\circ$ equiangular resolution (reciprocal approach). All boundaries are assumed to be rigid.

Frequency-dependent performance

In order to analyze the frequency-dependent performance of the introduced models, three objective HRTF parameters are evaluated in the following. In addition to the interaural time difference (ITD) and interaural level difference (ILD), also the spectral differences (SD) [6] are evaluated using the *ITA artificial head* HRTF as reference. Furthermore, the required number of mesh nodes for the models is analyzed.

Number of mesh nodes

To get an indication of the performance gained through simplification of the binaural receiver geometry, the mesh is generated for each frequency bin. The meshing constraints are chosen in a way that a), the geometry is represented properly and b), it refers to at least 5 nodes per wavelength (quadratic discretization). The result is shown in Figure 8. It can be seen, that up to approx. 3 kHz, the number of mesh nodes is constant and therefore mainly depends on the level of detail of the model. The least number of nodes is required for *Ellipsoid Offset Head*. Compared to the *ITA artificial head*, this refers to a reduction of approximately factor 20. The snowman models, obviously require more mesh nodes but even for the most detailed model *Ellipsoid Offset Snowman with Neck* this factor is still above 4.5.

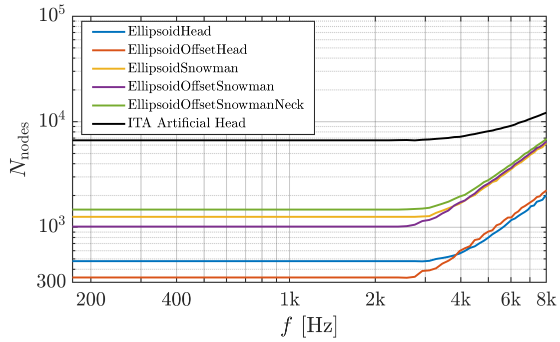

Figure 8: Number of mesh nodes for presented binaural receiver models depending on frequency.

Spectral differences (SD)

The spectral differences (SD) are a frequency-dependent measure to estimate how two HRTF data sets differ from each other. For this purpose, the standard deviation $\sigma_{\theta,\phi}$ of the relation between the HRTF sets is taken over all directions. In order to compensate for the oversampling of the poles using an equiangular sampling, the directions are weighted accordingly [6]. Here, each simplified model is compared to the artificial head.

$$SD(f) = \sigma_{\theta,\phi}\left(20 \cdot \log\left|\frac{H_{\mathrm{model}}(\theta,\phi,f)}{H_{\mathrm{ref}}(\theta,\phi,f)}\right|\right) \qquad \text{[dB]} \quad (1)$$

In Figure 9, it can be observed that above 2 kHz, the SD increase significantly for all models due to the missing pinna geometry. At the lower frequencies, an introduction of a torso geometry leads to a great reduction of the SD. However, above 700 Hz, *Ellipsoid Offset Snowman* performs worse than the other snowman models which is likely caused by the gap between head and torso.

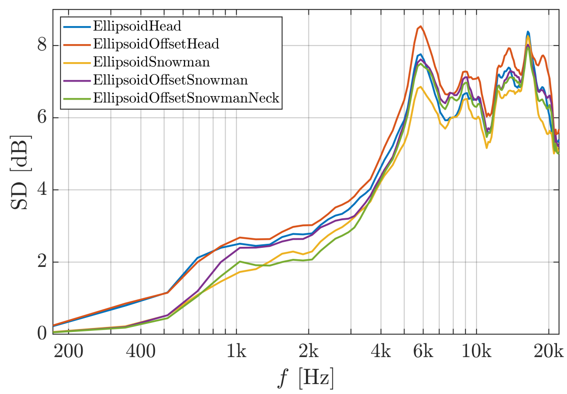

Figure 9: Spectral differences of presented binaural receiver models compared to artificial head depending on frequency.

Interaural time difference error (ΔITD)

The ITD of an HRTF can be estimated using the interaural phase delay difference which is related to sound source localization below 1.5 kHz [7]. By taking the mean over the frequency range $f \in [150, 1500]$ Hz, a frequency-independent parameter is obtained.

$$ITD(\theta,\phi) = \mathrm{mean}_f\left(\frac{\psi_{\mathrm{L}} - \psi_{\mathrm{R}}}{2\pi f}\right) \qquad \text{[s]} \quad (2)$$

Here, ψ_{L} and ψ_{R} refer to the phase of the HRTF of left and right ear respectively. The absolute ITD error between the simplified models and the artificial head is calculated for the horizontal plane ($\theta = 90°$) using:

$$|\Delta ITD(\phi)| = |ITD_{\mathrm{model}}(\phi) - ITD_{\mathrm{ref}}(\phi)| \qquad \text{[s]} \quad (3)$$

The results are shown in Figure 10. As expected, the introduction of an offset for the head leads to great reduction of the ITD error [2]. For the respective models, the error is far below the just noticeable difference (JND) [8]. On the other hand, introducing a torso only leads to a very small change of the ITD (with the exception of *Ellipsoid Offset Snowman*). This suggests, that the ITD seems to be mainly influenced by the head.

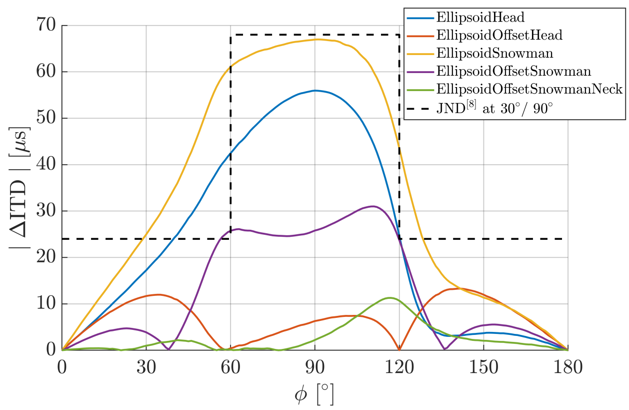

Figure 10: Interaural time difference error (ΔITD) of presented binaural receiver models compared to artificial head in the horizontal plane.

Interaural level difference error (ΔILD)

In contrast to the ITD, the ILD is an important localization cue above 1500 Hz [7]. In general, the ILD error between a model and the artificial head

$$\Delta ILD(f,\theta,\phi) = ILD_{\mathrm{model}} - ILD_{\mathrm{ref}} \qquad \text{[dB]} \quad (4)$$

depends on the direction and frequency. In order to get a direction-independent measure, the mean is taken over all directions in the horizontal plane ($\theta = 90°$) using equation 5.

$$|\Delta ILD(f)| = \mathrm{mean}_\phi\left(|\Delta ILD(f,\phi)|\right) \qquad \text{[dB]} \quad (5)$$

As visible in Figure 11, there is a significant increase of the ILD error above 3 kHz, similar to the SD. Furthermore, it can be seen that an introduction of a torso leads to a great improvement of the ILD at low frequencies. Of all models, *Ellipsoid Offset Snowman with Neck* shows the best performance up to 2.5 kHz. Below 1.3 kHz, the error is even below the JND [9].

Conclusion

In the present work, simplified models for binaural receivers for low frequency wave-based simulations were derived from the *ITA artificial head* using basic mathematical geometries. Their HRTFs were simulated and compared to the simulated HRTF of the original geometry. For these models, it was shown that the required

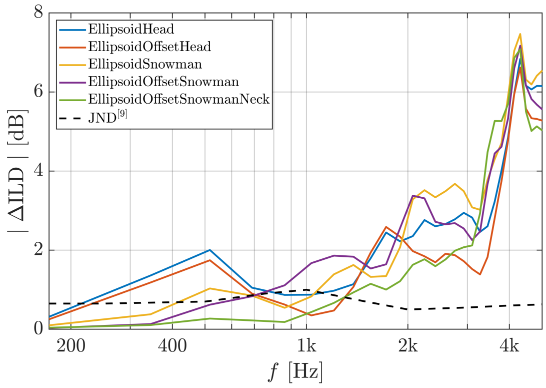

Figure 11: Interaural level difference error (ΔILD) of presented binaural receiver models compared to artificial head averaged over horizontal plane.

number of mesh nodes can be reduced significantly which again reduces the computational effort.

An inspection of the spectral differences (SD) and interaural level difference (ILD) showed that, caused by the missing pinna geometry, the errors increased rapidly above 2 kHz for all models. This indicates that the upper frequency limit for models without pinna could be close to this frequency.

For the interaural time difference (ITD) up to 1500 Hz, it was shown that using a head with an offset greatly improves the results. Since introducing a torso did not have a significant effect on the ITD, it is likely that it is mainly influenced by the head geometry. For the ILD on the other hand, the error could be decreased by adding a torso geometry, especially at low frequencies.

For all investigated parameters, it could be seen that a gap between head and torso leads to a significantly worse performance. Using such a geometry is therefore not recommended.

The best head model in this study is *Ellipsoid Offset Head*. Although its performance regarding SD and ILD is worse than the one of the snowman models, it shows a very good match of the ITD compared to the reference. Thus, this model might be suitable for simulations of very low frequencies where the influence of the torso can be neglected ($\lessapprox$ 1 kHz [7]), especially, since the ITD is a more important localization cue than the ILD in this frequency range. However, if using this model in an actual room, the findings of this study might not hold anymore. In this context, additional tests are required. The major advantage of this head model is its rotational symmetry around the yaw axis (z-axis). Due to this symmetry, all orientations of the binaural receiver within the horizontal plane can be simulated at once.

The best snowman model is *Ellipsoid Offset Snowman with Neck*. While of all presented models, it requires the highest number of mesh nodes, it showed the best performance regarding all error measures. The results suggest that it might be a suitable model up to approx. 2 kHz. Nevertheless, additional investigations are necessary to prove this statement.

Future work should address the simplification of additional artificial head geometries. In this context, it would be interesting whether similar investigations using an IEC conform artificial head lead to equivalent results. Furthermore, the presented models could be extended using simple geometries for the pinna. In this way, the upper frequency limit of the models might be moved towards higher frequencies. Last but not least, the simplified models should be tested within an actual simulation of a room such as a car cabin. Due to the influence of the room, the simplification of the geometry might be less crucial allowing to use the models for higher frequencies. Additionally, it would be of great interest to what extend the simulation can be accelerated exchanging a full detailed artificial head model with a simplified one.

Acknowledgements

The authors would like to thank Robert Henzel and Teresa Pelinski for their help implementing and carrying out the HRTF simulations.

References

[1] Aretz, M.: Combined wave and ray based room acoustic simulations of small rooms. Doctoral dissertation, RWTH Aachen University, Germany, 2012.

[2] Duda, R. O., Avendano, C., and Algazi, V. R.: An adaptable ellipsoidal head model for the interaural time difference. 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99, vol. 2, pp. 965–968, 1999.

[3] Algazi, V. R., Duda, R. O., Duraiswami, R., Gumerov, N. A., and Tang, Z.: Approximating the head-related transfer function using simple geometric models of the head and torso. The Journal of the Acoustical Society of America, vol. 112, no. 5, pp. 2053–2064, 2002.

[4] Schmitz, A.: Ein neues digitales Kunstkopfmeßsystem. Acta acustica united with acustica, vol. 81, no. 4, pp. 416–420, 1995.

[5] COMSOL Multiphysics® 5.4, COMSOL AB, Stockholm, Sweden. `https://www.comsol.com`.

[6] Richter, J.-G., Behler, G., and Fels, J.: Evaluation of a fast HRTF measurement system. Audio Engineering Society Convention, vol. 140, 2016.

[7] Xie, B.: Head-related transfer function and virtual auditory display. Second edition. Plantation, FL: J. Ross Publishing, 2013.

[8] Simon, L. S. R., Andreopoulou, A., and Katz, B. F. G.: Investigation of perceptual interaural time difference evaluation protocols in a binaural context. Acta Acustica united with Acustica, vol. 102, no. 1, pp. 129–140, 2016.

[9] Mills, A. W.: Lateralization of High-Frequency Tones. The Journal of the Acoustical Society of America, vol. 32, no. 1, pp. 132–134, 1960.