Modelling Binaural Receivers in Finite Difference Simulation of Room Acoustics

Jonathan Sheaffer, Craig Webb and Bruno Fazenda

Author's copy - version submitted to Proceedings of the International Conference on Acoustics, Montreal, Canada, June 2-7, 2013

ABSTRACT

Binaural room impulse responses are important for auralisation as well as for objective research in room acoustics. In geometrical room simulation methods, obtaining such responses is easily achieved by convolving each computed reflection tap with a corresponding pre-measured angle-dependent head-related impulse response. Unfortunately, employing such an approach in wave based methods is challenging due to temporal overlap of room reflections in the calculated response. One alternative is to physically embed a listener geometry in the grid. Whilst this method is straightforward, it requires voxelization of a geometrically-complex object. Furthermore, with non-conformal boundary conditions, the voxelized geometry is sample-rate dependent, meaning that numerical consistency may become compromised. In this paper we discuss the merits and drawbacks of embedding different listener geometries in the grid, ranging from a simple rigid sphere to a fully featured laser-scan of a Kemar mannikin. Results are analysed in terms of generated auditory cues, and are discussed in light of room acoustics modelling using the FDTD method.

INTRODUCTION

Binaural rendering of a soundfield is an efficient way to perform auralisation, and to objectively study spatial acoustic properties of a room. In the *Finite Difference Time Domain* (FDTD), which is becoming increasingly popular in room acoustics, modelling binaural receivers is not as straightforward as in geometrical methods, as the wave nature of the method and the need to bandlimit the grid results in temporal overlap between the various reflected components of the soundfield. Thus, in order to convolve each reflection with a corresponding pre-measured HRIR, one would first have to decompose the response into its discrete reflections, which is non-trivial, and requires post-processing. A simple way to approximate a binaural receiver is to record the sound pressure at two spaced receivers. However, in the absence of an efficient shadowing object between the two probes, interaural level differences (ILDs) are not correctly captured, and interaural time differences (ITDs) are only roughly estimated. Murphy and Beeson [1, 2] have shown that in a 2-D Digital Waveguide Mesh, approximation of correct ITD values is possible by embedding a circular object between two spaced receivers. Whilst it is known that for broadband signals the auditory system mostly relies on ITD cues to resolve the locus of the sound source, ILD cues are important for resolution of high frequency components and still play an active role in the overall spatial evaluation of a broadband source [3]. Furthermore, in reverberant listening conditions, the superposition of reflections increase the temporal variability of ITD cues [4], and they consequently become less reliable. Correct representation of ITDs and ILDs allow a good localisation judgement on the horizontal plane, however a more complete binaural response should also include comb filtering effects, caused by reflections from the pinna and shoulders.

Accurate modelling of complete HRIRs, including all binaural and monaural cues, was successfully achieved in [5, 6] using specialised FDTD schemes for propagation of sound in inhomogeneous media. Whilst their results are in good agreement with measured responses, employing such an approach in room acoustics FDTD is challenging. First, the spatial period required to achieve such accuracy is in the order of a single millimetre which is computationally expensive given a uniformly sampled room, even with the aid of GPGPU technology [7, 8, 9]. Furthermore, it is not clear whether sufficient accuracy can be achieved using efficient numerical schemes more commonly used for room acoustics simulation [10, 11], as they are principally different and normally employed in a coarser grid setup.

A more versatile yet indirect method is to capture and spatially encode the output of an FDTD model [12]. Using this approach, n^{th}-order ambisonic impulse responses are obtained, which can be further post-processed to generate auralisations. In this paper we investigate a method of obtaining binaural impulse responses by directly embedding a listener geometry in the grid. Unlike previous studies, in which the goal was to accurately model HRIRs using the FDTD method, here the focus is on room modelling. Thus, in context of numerical schemes and boundary models, priority is given to schemes more suitable for room acoustics simulation. Accordingly, we investigate the performance of embedded listeners under such constraints.

ROOM ACOUSTICS MODELLING USING FDTD

Assuming an inviscid-adiabatic propagation of sound in air, the sound field in a room at $\mathbf{x} = (x, y, z) \in \mathbb{R}^3$, can be described by the wave equation, given by

$$\frac{1}{c^2}\frac{\partial^2 p(\mathbf{x},t)}{\partial t^2} - \nabla^2 p(\mathbf{x},t) = \psi(\mathbf{x},t) \tag{1}$$

where $p(\mathbf{x},t)$ is sound pressure, c is the velocity of sound in air and $\psi(\mathbf{x},t)$ denotes the source driving function. The sound field is sampled such that $(x,y,z,t) \rightarrow (lX,mX,iX,nT)$, where X is the spatial period, T is the temporal period and $\lambda = cT/X$ is the Courant number [10]. Accordingly, the wave equation can be modelled as

$$\left(\delta_t^2 - \lambda^2 \delta_{\mathbf{x}}^2\right) p\big|_{\mathbf{i}}^n = c^2 T^2 \psi\big|_{\mathbf{i}}^n \tag{2}$$

The discrete FD operators are given by

$$\begin{aligned} \delta_t^2 p\big|_{\mathbf{i}}^n &\equiv p\big|_{\mathbf{i}}^{n+1} - 2p\big|_{\mathbf{i}}^n + p\big|_{\mathbf{i}}^{n-1} & \delta_x^2 p\big|_{\mathbf{i}}^n &\equiv p\big|_{l+1,m,i}^n - 2p\big|_{l,m,i}^n + p\big|_{l-1,m,i}^n \\ \delta_y^2 p\big|_{\mathbf{i}}^n &\equiv p\big|_{l,m+1,i}^n - 2p\big|_{l,m,i}^n + p\big|_{l,m-1,i}^n & \delta_z^2 p\big|_{\mathbf{i}}^n &\equiv p\big|_{l,m,i+1}^n - 2p\big|_{l,m,i}^n + p\big|_{l,m,i-1}^n \end{aligned} \tag{3}$$

where $\delta_{\mathbf{x}}^2 = \delta_x^2 + \delta_y^2 + \delta_z^2$ for a standard rectilinear grid, and $\mathbf{i} = [l,m,i]$ is the discrete index position. For a more detailed introduction of the method, including explicit update equations, readers are referred to [10]. When modelling a room, the edges of the domain represent walls and one applies boundary conditions respectively. In this paper, we simulate anechoic boundaries, which is accomplished by making the domain large enough such that any reflections can be windowed out of the resulting response.

Modelling Binaural Receivers

Embedding objects directly in the FDTD grid is a straightforward approach to modelling binaural receivers. Whilst it is possible to achieve accurate geometrical representations of objects and numerically consistent results using conformal boundary methods [13], many authors use the FDTD method for its simplicity and ease of implementation, and as such, refrain from employing complex realisations. Furthermore, many room models do not feature curved or oblique boundaries, and non-conformal methods are the appropriate efficient choice. In this paper, all embedded objects are calculated using a boundary model corresponding to

$$\frac{\partial p(\mathbf{x},t)}{\partial t} = c\beta \mathbf{n} \cdot \nabla \mathrm{p}(\mathbf{x},\mathrm{t}) \tag{4}$$

where $\mathbf{n}$ is the normal to the surface at boundary, and β is a simple loss coefficient, as detailed in [14].

Rigid Sphere Model

Using a rigid sphere to approximate the human head is a simple means of obtaining interaural localisation cues [15]. In context of this investigation, it is useful for two other reasons. First, lacking any spectral cues, responses generated using a rigid sphere model are easy to visually inspect. Secondly, there exist a closed-form solution for diffraction around a rigid sphere, which may be used as an exact reference. For comparison with the numerical results in this paper, the analytical model described in [16] is used.

KEMAR Mannikin

A more accurate representation of a listener, is an actual model of a human head. Here we use laser scans of a KEMAR mannikin, originally used in [17], which have been polygon-reduced and voxellized using a ray-tracing algorithm [18], as shown in Figure 1. The amount of spatial detail

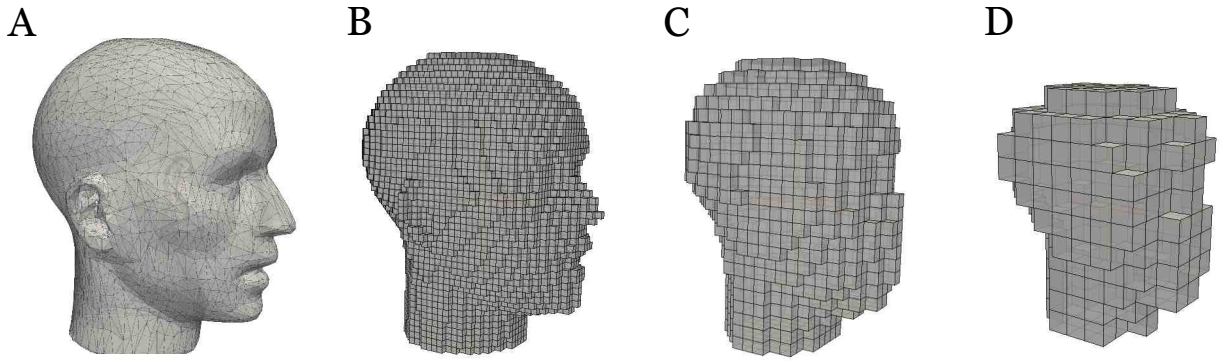


FIGURE 1: Geometry of a KEMAR mannikin used in this study. A. Polygon reduced laser scans, B. Voxellized for 88.2kHz (X = 6.75mm), C. Voxellized for 44.1kHz (X = 13.5mm), D. Voxellized for 22.05kHz (X = 27mm).

which is preserved after discretising the model is dependent on the temporal sample rate, as such, the shape of the head approaches that of a sphere as the sampling frequency is reduced. As reference to the numerical model, we use the measured KEMAR set from MIT [19].

RESULTS

A 3x3x3 room model was calculated using the standard rectilinear FDTD grid, with an object approximating the listener placed at the centre of the domain. A Physically-Constrained (PCS) sound source [20] with a high cutoff frequency at 10kHz, low frequency resonance at 100Hz and Q = 0.7 was placed at 1m radial distance away from the object and the resulting sound pressure was recorded on the surface of the object, at the positions of the ears. The simulation was repeated 360° around the object at 5° intervals. To study the different numerical effects, the experiment was conducted at three different sample rates, namely 22.05kHz, 44.1kHz and 88.2kHz. The cutoff frequency and all other source parameters were held constant across all different sample rates, to allow the responses to be contaminated with numerical dispersion. All simulations were executed using a parallelised *Compute Unified Device Architecture* (CUDA) algorithm on a Tesla C2070 general-purpose GPU, as detailed in [14]. At 88.2kHz, the 27m^3 domain requires grids that contain 90 million points. For efficiency, the computation is split into two parts. Firstly, the entire domain interior is updated using a SIMD (Single Instruction Multiple Data) kernel operating over free space. Then, a small number of separate threads are used to compute the bounding domain that contains the head model. This necessitates storing three grids of domain data, and in total requires 1.1Gb of global memory at single precision.

FDTD Sphere Model

The geometry of a sphere of radius 17cm was analytically generated and embedded into the model. Resulting left-ear HRTFs are shown in Figure 2. As expected, the higher the sample rate, the more the modelled results resemble the exact solution. It can also be seen that similarity is decreased as the frequency is increased, which is a result of the respones being contaminated by numerical dispersion.

FDTD KEMAR Model

The geometry of a KEMAR model, as discussed in Section 3.2, was embedded into the model. Resulting left-ear HRTFs are shown in Figure 5. Unlike the sphere model, in which there was a relatively good match between modelled and exact responses, here even at the highest used sample rate, only the general shape of the HRTF is similar.

Analysis

Whilst direct visual inspection of the HRTFs may indicate how similar the modelled results are to their references, in terms of perception, it is important to evaluate the similarity of the cues they provide to the auditory system. In this section we analyse the modelled results in terms of their inherent spatial cues, namely interaural differences and spectral cues.

Interaural Cues

Interaural cues, specifically interaural time differences (ITDs) and interaural level differences (ILDs), are foremost responsible for providing the auditory system with the information required to localise sources, particularly on the frontal-horizontal plane [21]. Consider two signals, $p_L(t,\theta,\phi)$ and $p_R(t,\theta,\phi)$, denoting the sound pressure due to an impulsive source which is captured at the left and right ears of a listener respectively. These head-related impulse-responses (HRIRs) characterise the physical filtering process due to the head. In the frequency domain, the head-related transfer-functions at the left and right ears are obtained by taking the Fourier transform of the HRIRs, denoted $\mathbf{H}_L(\omega,\theta,\phi)$ and $\mathbf{H}_R(\omega,\theta,\phi)$ respectively. The interaural time differences denote the amount of time delay of each frequency component, which can be calculated from the phase-delay of the two signals,

$$\Delta T(\omega,\theta,\phi) = -\frac{1}{\omega}\angle\,\frac{\mathbf{H}_L(\omega,\theta,\phi)}{\mathbf{H}_R(\omega,\theta,\phi)} \tag{5}$$

Figure 3 shows the computed ITDs, for the rigid sphere and the KEMAR models as well as for their references. It can be seen that even at a relatively low sample rate of 22,050Hz there is good matching with the reference models for both KEMAR and the rigid sphere. An interesting point in case, is the highest frequency at which interaural phase differences can be unambiguously resolved by the auditory system. This frequency can be identified as the tip of the abrupt colour change in Figure 3, which for both reference models is situated at around 750Hz (for the two extreme lateral angles). In both numerical models, this frequency is considerably lower, at around 500 – 600Hz. Since ITDs are calculated from phase-delays, then this implies a numerical inaccuracy in the phase response of the boundary model used to represent the geometry approximating the head. The interaural level differences show the difference, in decibels, between the magnitude of pressure at the two ears. This quantity can be calculated from

$$\Delta L(\omega,\theta,\phi) = 20\log\left|\frac{\mathbf{H}_L(\omega,\theta,\phi)}{\mathbf{H}_R(\omega,\theta,\phi)}\right| \tag{6}$$

Figure 6 and 4 show the computed ILDs for the KEMAR and rigid sphere models, respectively. For the case of the rigid sphere, it can be seen that the 88.2kHz model (pane D) produces results which rather closely match the reference model, with some divergence at high frequencies which is likely due to numerical dispersion. As expected, the accuracy at which cues are reproduced is restricted by the bandwidth of the FDTD scheme, which is especially evident at lower sample rates (panes B and

C). For the case of the KEMAR model, it appears that even at a high sample rate of 88.2kHz, only the general shape of the ILD pattern is reproduced. Whilst the inability to reproduce many of the subtle ILD features would be intuitively attributed to the limited spatial resolution of the grid, it is likely that the accuracy of the non-conformal boundary model plays a role, as many of these features are physically linked to a fine representation of the head geometry.

Spectral Features

Spectral cues, which are normally produced by reflections from different areas of the pinna, are best observed at frequencies above 7kHz [22]. In this work, the 10kHz cutoff frequency of the excitation signal was chosen due to computational constraints, and accordingly many spectral cues which exist at higher frequencies are not calculated. Spectral cues below 10kHz, which are evident as low pressure areas in Figure 5A at angles of incidence above 180°, are not correctly reproduced even in the 88.2kHz model. At models of lower resolutions, some distinct pressure patterns are evident, however being at frequencies above the cutoff frequency for the standard rectilinear scheme [10], they would not represent any meaningful result.

Discussion and Conclusion

In this paper we investigated how well binaural cues can be obtained from a simple FDTD model, by means of directly embedding a listener geometry in the grid. In contrast to previous studies, the purpose of this paper was to asses the accuracy of modelling binaural listeners as part of a room model, rather than to show how the FDTD method could be used to model HRTFs themselves. Results indicate that ITDs match their reference values even at sample rates as low as 22.05kHz, as they are mostly used for localisation of lower frequencies and as such, do not require intense oversampling in order to be accurately reproduced. The general pattern of ILDs is reproduced reasonably well, with the sample rate being the foremost limiting factor. Spectral features caused by Pinna reflections are hardly evident even in the 88.2kHz grid. This is mostly attributed to the spatial resolution limited by memory and computational constraints. It may be possible to obtain better results with higher sample rates, however, even with the aid of GPGPU technology this is still not possible for larger physical domains. It could be argued that comparable cues could be obtained by using a rigid sphere model, whose geometry is trivial to generate analytically. However, it should be borne in mind that some assymetries evident in the human head, do result in more unique interaural cues which are distinguishable from those of a rigid sphere. Thus, there is some advantage to modelling the shape of the head, even when more coarse geometrical representations are considered. Another option is to embed a combination of geometrical shapes approximating a human listener, known as the *snowman model* [15].

Furthermore, whilst straightforward to implement, the method of directly embedding a listener geometry in the grid, does not provide numerically consistent results, as the discretisation resolution varies with the sample rate. We therefore conclude that the method is mostly suitable for cases where one is only interested in obtaining interaural cues. In room acoustics modelling, the designer's choice of numerical parameters need to be independent of the source and receiver models. Accordingly, future work should also look at alternative ways to include binaural listeners in an FDTD model, which could possibly involve approximation of head related impulse responses using digital filters, rather than by means of direct computation.

ACKNOWLEDGMENTS

Laser scans of the KEMAR mannikin are courtesy of Yuvi Kahana, and the Institute of Sound and Vibration Research (ISVR), the University of Southampton.

REFERENCES

[1] D. Murphy and M. Beeson, "Modelling spatial sound occlusion and diffraction effects using the digital waveguide mesh", in *Proc. AES 24th International Conference, Multichannel Audio*, 26–28 (2003).

[2] D. Murphy and M. Beeson, "The KW-boundary hybrid digital waveguide mesh for room acoustics applications", Audio, Speech, and Language Processing, IEEE Transactions on **15**, 552–564 (2007).

[3] E. Macpherson and J. Middlebrooks, "Listener weighting of cues for lateral angle: the duplex theory of sound localization revisited", The Journal of the Acoustical Society of America **111**, 2219 (2002).

[4] A. Ihlefeld and B. Shinn-Cunningham, "Effect of source spectrum on sound localization in an everyday reverberant room", The Journal of the Acoustical Society of America **130**, 324 (2011).

[5] T. Xiao and Q. Liu, "Finite difference computation of head-related transfer function for human hearing", The Journal of the Acoustical Society of America **113**, 2434 (2003).

[6] P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato, "Computer simulation of hrtfs for personalization of 3d audio", in *2nd Intl. Symp. on Universal Communication*, 435–440 (IEEE) (2008).

[7] L. Savioja, "Real-time 3D finite-difference time-domain simulation of low-and mid-frequency room acoustics", in *13th Int. Conf on Digital Audio Effects* (2010).

[8] J. Sheaffer and B. Fazenda, "FDTD/K-DWM simulation of 3d room acoustics on general purpose graphics hardware", in *Proc. of the Institute of Acoustics*, volume 32 (Institute of Acoustics) (2010).

[9] C. Webb and S. Bilbao, "Computing room acoustics with cuda-3d fdtd schemes with boundary losses and viscosity", in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, 317–320 (IEEE) (2011).

[10] K. Kowalczyk and M. van Walstijn, "Room acoustics simulation using 3-D compact explicit FDTD schemes", IEEE Trans. on Audio, Speech, and Lang. Proc. **19**, 34–46 (2011).

[11] S. Bilbao, "Optimized FDTD schemes for 3D acoustic wave propagation", IEEE Trans. on Audio, Speech, and Lang. Proc. 1–1 (2012).

[12] A. Southern, D. Murphy, and L. Savioja, "Spatial encoding of finite difference time domain acoustic models for auralization.", IEEE Trans. on Audio, Speech, and Lang. Proc. **PP**, 1 (2012).

[13] J. Tolan and J. Schneider, "Locally conformal method for acoustic finite-difference time-domain modeling of rigid surfaces", The Journal of the Acoustical Society of America **114**, 2575 (2003).

[14] C. Webb and S. Bilbao, "Computing room acoustics with CUDA - 3D FDTD schemes with boundary losses and viscosity", in *Proc. IEEE Int. Conf. on Acoustics, Speech and Sig. Proc.* (Prague) (2011).

[15] V. Algazi, R. Duda, R. Duraiswami, N. Gumerov, and Z. Tang, "Approximating the head-related transfer function using simple geometric models of the head and torso", The Journal of the Acoustical Society of America **112**, 2053 (2002).

[16] D. Brungart and W. Rabinowitz, "Auditory localization of nearby sources. head-related transfer functions", The Journal of the Acoustical Society of America **106**, 1465 (1999).

[17] Y. Kahana and P. Nelson, "Boundary element simulations of the transfer function of human heads and baffled pinnae using accurate geometric models", J. Sound Vib. **300**, 552–579 (2007).

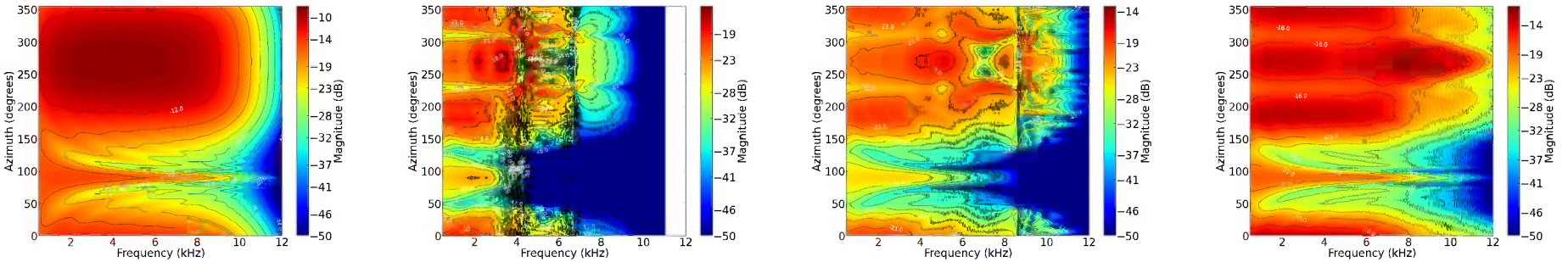
[18] S. Patil and B. Ravi, "Voxel-based representation, display and thickness analysis of intricate shapes", in *9th Intl. Conf. Computer Aided Design*, 6–pp (IEEE) (2005).

[19] B. Gardner, K. Martin, *et al.*, "HRTF measurements of a kemar dummy-head microphone", (1994).

[20] J. Sheaffer, M. van Walstijn, and B. Fazenda, "A physically-constrained source model for FDTD acoustic simulation", in *Proc. of the 15th Intl. Conf. DAFx-12* (2012).

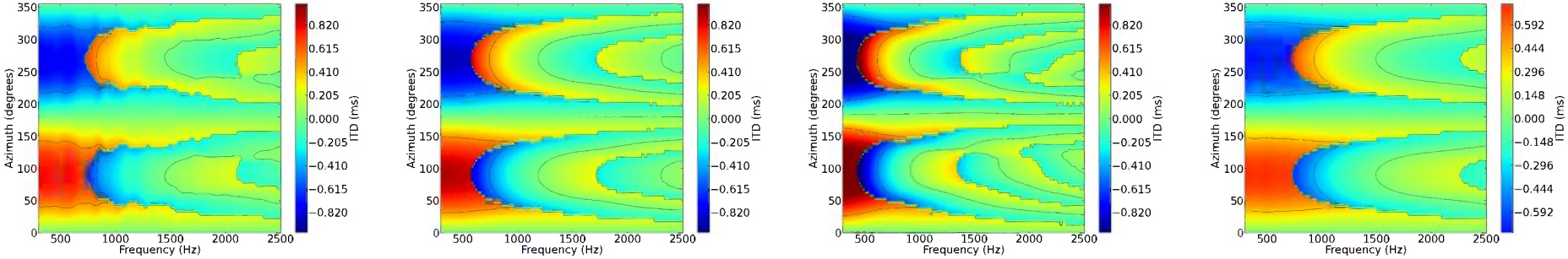
[21] J. Blauert, "Spatial hearing (revised edition)", Cambridge, MA: Massachusetts Institute of Technology (1997).

[22] J. Middlebrooks and D. Green, "Sound localization by human listeners", Annual review of psychology **42**, 135–159 (1991).



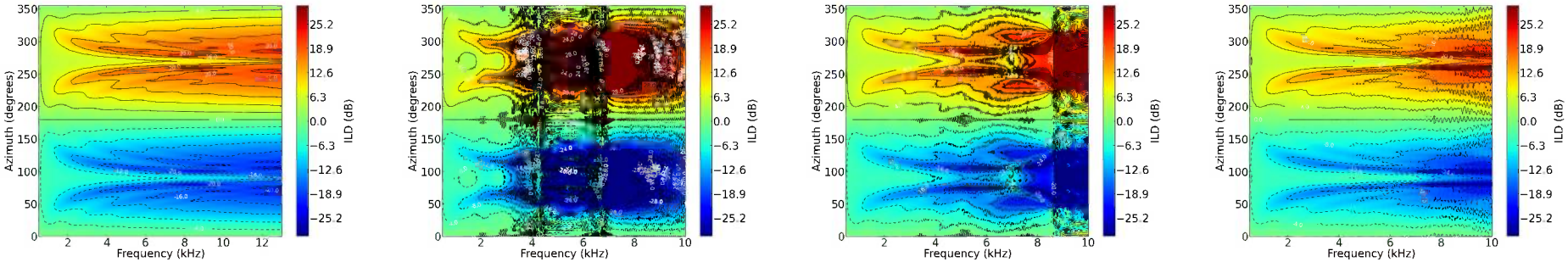
(A) Reference (B) 22.05kHz (C) 44.1kHz (D) 88.2kHz

FIGURE 2: Left ear HRTFs for a rigid sphere.



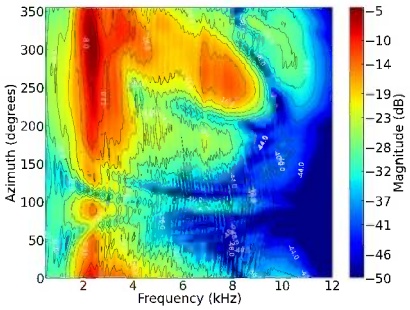
(A) Reference KEMAR (B) 22.05kHz Sphere (C) 22.05kHz KEMAR (D) Reference Sphere

FIGURE 3: Interaural time differnces (ITDs) calculated for the KEMAR and rigid sphere models.



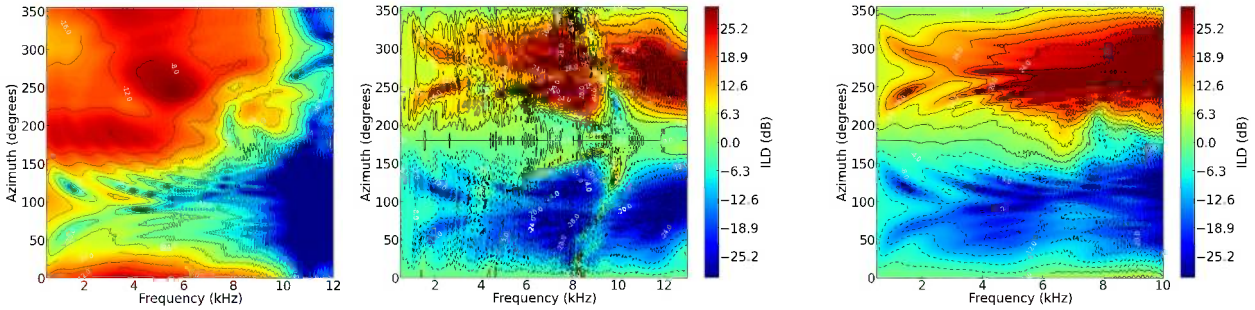
(A) Reference Sphere **(B)** 22.05kHz Sphere **(C)** 44.1kHz Sphere **(D)** 88.2kHz Sphere

FIGURE 4: Calculated ILDs for the rigid sphere model.



(A) Reference **(B)** 88.2kHz

FIGURE 5: Left ear HRTFs for a KEMAR mannikin.



(A) Reference KEMAR **(B)** 88.2kHz KEMAR

FIGURE 6: Calculated ILDs for the KEMAR model.