A Physical Model for Binaural Sound Synthesis

Colin Fahy

New York University

3D Audio
Dr. Agnieszka Roginska

May 7, 2016

Abstract

This paper proposes a method for adding binaural localization cues to a monophonic audio signal. Given the azimuth angle, elevation angle, and distance of a sound source as well as the head radius of a listener, the interaural time difference (ITD) and interaural level difference (ILD) can be estimated. The ITD is computed using a spherical head model. The frequency dependent ILD is estimated using a second order filter to approximate the frequency response of a rigid sphere, and a series of peak and notch filters to approximate the pinna at different elevations.

1 Introduction

The recording of and convolution with head-related impulse responses (HRIR) has proven to be an effective means of producing 3D binaural audio signals. However it has it's limitations. It requires many recordings to fully map all directions and a method of interpolating points in between, it is difficult to adjust to individual listeners, and cannot alone account for large changes in distance.

In this paper I propose a method for computing the interaural time difference (ITD) and interaural level difference (ILD) binaural cues based on a spherical head model. These cues can be applied to a monophonic signal by means of a delay and series of filters to create a stereo binaural signal.

Additionally, I have built a graphical user interface for listening to binaural audio files at any azimuth, elevation, and distance in real time. This interface can be used to compare the results of the proposed method with the results of convolution with HRIR recordings, as well as simple stereo equal-powered panning.

2 Background

2.1 Interaural Time Difference

The ITD measures the difference in arrival time of the direct signal to the ipsalateral and contralateral ears. This is often estimated as the difference in travel distance to each ear divided by an approximation of the speed of sound in air. Woodworth and Schlosberg first proposed a model for computing the ITD of a sound source given it's azimuth angle and listener's head radius using a spherical head model, this is given in equation 1 [10].

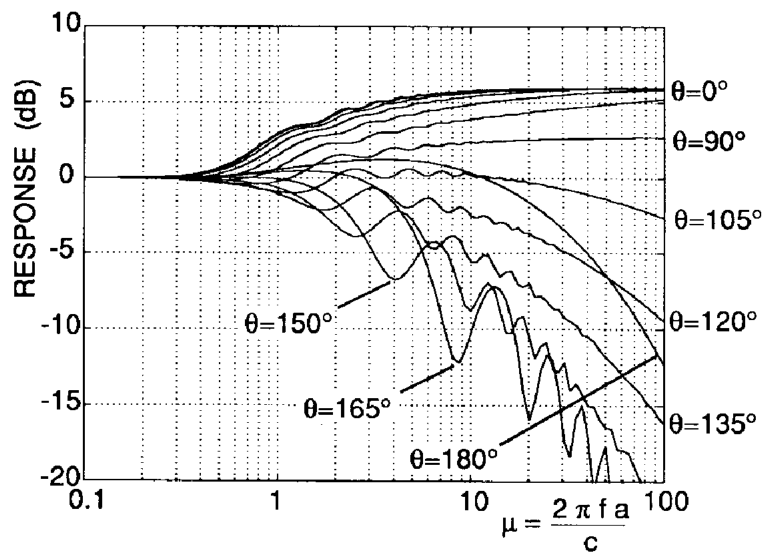
$$ITD = \frac{a}{c}(\theta + sin(\theta)) \tag{1}$$

Where a is the radius of the head and c is the speed of sound in air. This equation assumes the sound approaches as a plane wave and therefore may break down for close sounds that act more like a point source. It also does not take into account elevation. Juras and Miller proposed a method of computing the ITD computes the distance to each ear and calculates the path to the contralateral ear as the sum of the tangent line to the spherical head and the remaining arc length.

2.2 Interaural Level Difference

The ILD is more difficult to predict as it is frequency dependent and can vary significantly based on the individual. The most significant factors include the attenuation associated with the extra distance traveled, especially of high frequencies that are reflected or absorbed by the head, and the spectral filtering caused by the shape of the listener's pinna. In 1998 Brown and Duda proposed a method of estimating head-shadow based on the findings of Lord Rayleigh's solution for the diffraction of a plane wave by a rigid sphere. Brown and Duda created an azimuth-dependent head shadow effect with a second order filter approximating this curve [4].

Figure 1: Diffraction of a plane wave by a rigid sphere. [7]



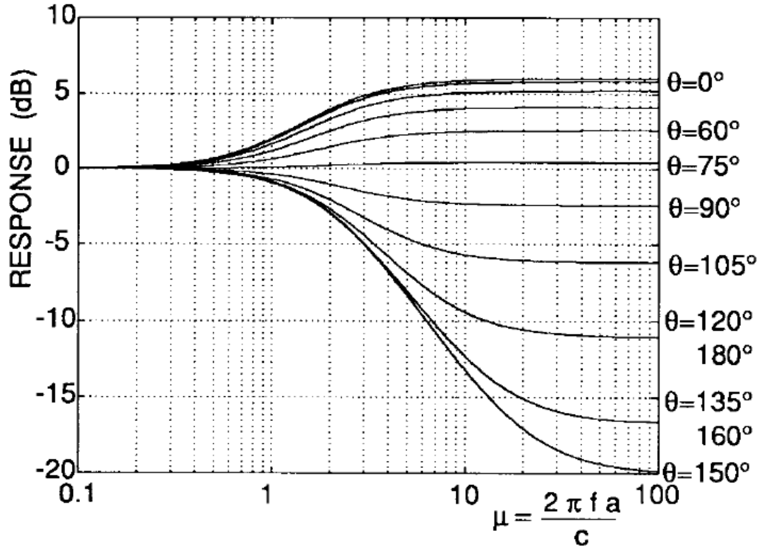
In 2010, Spagnol et al. studied the efficacy of estimating the pinna-related transfer function (PRTF) using peak and notch filters to represented the first major resonances and notches caused by the cavities of the listener's pinna. They measured the frequency response of subjects' HRTFs to determine the frequency and magnitude of the first three resonances and notches at different elevations. Then used these values to create filters that approximately matched the actual frequency response at varying elevation.

3 Methodology

3.1 Interaural Time Difference

For computing the distance to each ear, I used a modified version of the method proposed by Juras and Miller [6]. The same measure of distance can be computed using the azimuth

Figure 2: Frequency response of the 2nd order filter approximation of head shadow. [4]



angle, elevation angle, distance and head radius. The geometry used to compute these distances can be seen in figure 3. The angle of incidence, θ can be derived from the azimuth and elevation angles according to equation 2.

$$\theta = \arccos(\cos(\frac{\pi}{2} - azimuth) \cdot \cos(elevation)) \tag{2}$$

With this angle, the distance to the ipsilateral ear, d_1, can be computed according to equation 3.

$$d_1 = \sqrt{d^2 + r^2 - 2 \cdot d \cdot r \cdot \cos\theta} \tag{3}$$

Where d is the distance from the sound source to the center of the head, and r is the radius of the head.

The distance to the contralateral, d_2, ear can be computed according to equation 4.

$$tangent = \sqrt{d^2 - r^2} \tag{4}$$

$$arc = r \cdot (\pi - \theta - \arccos\frac{r}{d}) \tag{5}$$

$$d_2 = tangent + arc \tag{6}$$

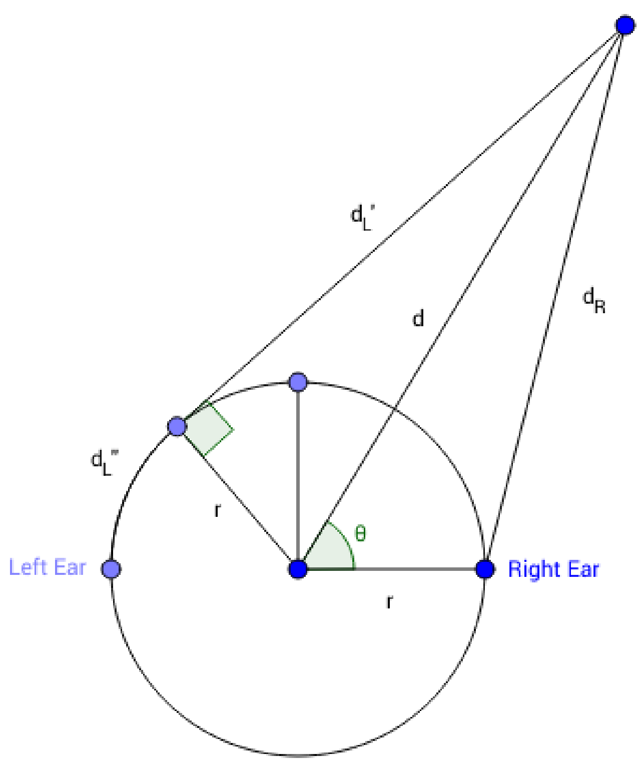
However, for sound sources very close to the head and with small azimuth angle, the source may actually travel along the arc of the head for the ipsilateral ear as well, since the direct path is obscured by the head. This happens when the tangent line is shorter than the computed distance to the ipsilateral ear. Therefore d_2 can be recomputed according to equation 7.

$$\text{if tangent} < d_1: \quad d_1 = tangent + r \cdot (\theta - \arccos \frac{r}{d}) \tag{7}$$

From these distances the ITD can be computed by dividing the difference in distance by the speed of sound according to equation 8, approximating c to be 343 m/s.

$$\begin{aligned} \Delta d &= d_1 - d_2 \\ ITD &= \frac{\Delta d}{c} \end{aligned} \tag{8}$$

Figure 3: The geometry for computing the interaural distance difference. r is the head radius, d the distance from the center of the head to the sound source. θ is the angle of incidence, computed from the azimuth and elevation angle.



3.2 Interaural Level Difference

I have broken down the interaural level difference into three parts: attenuation due to distance traveled, attenuation due to head shadow, and spectral filtering caused by the pinna.

Since the sound must travel further to reach the contralateral ear, it will lose some energy by the time it reaches the ear. Sound level decreases inversely with distance traveled. Therefore the difference in sound pressure level between the two ears can be computed according to equation 9.

$$\Delta L = \left| 20 \cdot \log 10 \frac{d_{ipsilateral}}{d_{contralateral}} \right| \tag{9}$$

However, to have distance effect the perceived overall volume of the sound, each level can be computed separately relative to a reference distance. Here I assumed the input monophonic source to be considered one meter from the listener.

$$d_{ref} = 1$$

$$L_R = L - \left| 20 \cdot \log 10 \frac{d_r ef}{d_R} \right| \tag{10}$$

$$L_L = L - \left| 20 \cdot \log 10 \frac{d_r ef}{d_L} \right| \tag{11}$$

This attenuation does not account for the reflection and absorption of higher frequencies by the head. For this, the second order filter developed by Brown and Duda is used to compute the frequency dependent response that represents this head shadow [4]. The filter is applied to both signals, using the angle of incidence, θ, as input to the ipsilateral filter, and $\pi - \theta$ as the input to the contralateral filter.

The final step is to apply the filtering caused by the pinna. A generalized version of the technique used by Spagnal et al. can roughly approximate the effects of the pinna at varying elevations [8]. Since this model has limited information about the listener, the individual resonance and notch tracking implemented by Spagnol et al. cannot be used. However, a majority of subjects' in their study had peaks around the same frequencies and followed similar trajectories with respect to elevation. Linear interpolation is used to create resonance filters that go from 1000Hz to 4500Hz and 11000Hz to 10000Hz with constant gain as elevation angle changes from $-\pi$ to π. Notches filters were created at ranges 6000Hz to 10000Hz, 10000Hz to 9000Hz, and 10000Hz to 14000Hz, all with decreasing gain as elevation changes from $-\pi$ to π.

4 Graphical User Interface

A user interface was developed to implement this model in a real-time, and allow for quick comparison against other methods of audio panning and with changing variables. The interface was build in Javascript utilizing the Web Audio API, and the Matter.js physics engine. The interface allows for placing sound sources anywhere on a 2D grid to experience the sound processed from the current azimuth relative to the listener. Each source can have it's elevation adjusted individually, though this cannot be visualized in the 2D interface. The proposed model is implemented using a DelayNode, GainNode, and a series of BiquadFilterNodes from the Web Audio API. The listener's head radius as well as the max distance represented by the interface can be adjusted in real-time. This method can be compared to the output of the PannerNode builtin to the Web Audio API, using either HRIR convolution or stereo equal-powered panning. For all algorithms, the source location is updated every 256 samples.

The interface can be found at `http://webaudio.cfahy.com/binaural`.

5 Conclusions and Future Work

No formal assessment has been carried out, but from the author's listening, changes in azimuth and distance sound fairly accurate, though not quite as realistic as the HRTF convolution. There is a slight degrade in audio quality when using the resonance and notch filters. Changes in elevation are noticeable and give a perceived change in the direction that the elevation is changed, but does not give the sense of realism the HRTF convolution does. Going forward, I would like to more accurately estimate the pinna resonance and notch filters. Since the frequency of resonances and notches can be predicted from the size of the concha, it could be worth adding that measurement as a parameter to derive the peak frequencies from.

Sridhar and Choueiri have a method for extending the spherical head model to account for changes in elevation [9]. The used linear regression on the measured values of test subjects to predict a head radius to use in Woodworth's formula according to the listener's head's height, depth, and width. It would be interesting to expand my derivation of the distance to each ear that already includes elevation information to account for these extra head shape dimensions.

References

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

[2] V. R. Algazi, R. O. Duda, R. P. Morrison, and D. M. Thompson, "Structural composition and decomposition of hrtfs," in *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the.* IEEE, 2001, pp. 103–106.

[3] V. R. Algazi, R. O. Duda, and P. Satarzadeh, "Physical and filter pinna models based on anthropometry," in *Audio Engineering Society Convention 122.* Audio Engineering Society, 2007.

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

[5] R. O. Duda, V. R. Algazi, and D. M. Thompson, "The use of head-and-torso models for improved spatial sound synthesis," in *Audio Engineering Society Convention 113.* Audio Engineering Society, 2002.

[6] J. Juras, C. Miller, and A. Roginska, "Modeling itds based on photographic head information," in *Audio Engineering Society Convention 139.* Audio Engineering Society, 2015.

[7] L. Rayleigh and A. Lodge, "On the acoustic shadow of a sphere. with an appendix, giving the values of legendre's functions from p 0 to p 20 at intervals of 5 degrees," *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, vol. 203, pp. 87–110, 1904.

[8] S. Spagnol, M. Geronazzo, and F. Avanzini, "Structural modeling of pinna-related transfer functions," in *In Proc. Int. Conf. on Sound and Music Computing (SMC 2010)*, 2010.

[9] R. Sridhar and E. Choueiri, "Capturing the elevation dependence of interaural time difference with an extension of the spherical-head model," in *Audio Engineering Society Convention 139.* Audio Engineering Society, 2015.

[10] R. S. Woodworth and H. Schlosberg, "Experimental psychology (rev," 1954.

[11] B. Xie, *Head-related transfer function and virtual auditory display.* J. Ross Publishing, 2013.