Sonic Localization Cues for Classrooms: A Structural Model Proposal

Abhijit Mitra and Cemal Ardil

Abstract— We investigate sonic cues for binaural sound localization within classrooms and present a structural model for the same. Two of the primary cues for localization, interaural time difference (ITD) and interaural level difference (ILD) created between the two ears by sounds from a particular point in space, are used. Although these cues do not lend any information about the elevation of a sound source, the torso, head, and outer ear carry out elevation dependent spectral filtering of sounds before they reach the inner ear. This effect is commonly captured in head related transfer function (HRTF) which aids in resolving the ambiguity from the ITDs and ILDs alone and helps localize sounds in free space. The proposed structural model of HRTF produces well controlled horizontal as well as vertical effects. The implemented HRTF is a signal processing model which tries to mimic the physical effects of the sounds interacting with different parts of the body. The effectiveness of the method is tested by synthesizing spatial audio, in MATLAB, for use in listening tests with human subjects and is found to yield satisfactory results in comparison with existing models.

Keywords— Auditory localization, Binaural sound, Head related impulse response, Head related transfer function, Interaural level difference, Interaural time difference, Localization cues.

I. Introduction

CAPABILITY to locate sound sources through sonic cues is of critical importance to many animals as they have to rely mostly on sounds for their living by hunting. Even though humans do not hunt for food by sound any more, we exhibit a reasonable auditory localization [1]-[4] ability through a signal called *binaural beat* [5]. When two sounds with a subtle phase shift arrive at one ear slightly before arriving at the other one, brain integrates these two signals producing a sensation of a third sound called the *binaural beat*, resulting in detecting the phase or frequency difference between the two sounds. This phase difference provides directional information and enables us to determine the physical location of a sound source. Normally, the difference in phase relationship can be detected when sound frequencies are below approximately 1 kHz and it becomes more tedious for us to determine the physical location of a high pitched sound. The importance of this ability through binaural beats was discovered by a German experimenter, H. W. Dove in 1839 [6] and researchers began investigating several sonic cues or parameters that we use for localization hundreds of years ago. However, this topic has received an increased attention only over the last couple of decades with the development of more advanced computers, sensitive measurement devices and an increased interest in virtual environments where the focus has been to synthesize three dimensional (3-D) sounds for scientific, commercial and entertainment purposes [7]-[8]. Although many models are presently existing to demonstrate such 3-D sound synthesis, most of them are complicated from implementation view point. We propose here another modeling method which is simple yet effective.

With two different ears receiving binaural input from the same sound source (*monaural*), the sound has to travel farther to reach one of the two ears. Therefore, the ear that is farther from the sound source will not have a direct path for the sound wave to follow, as the head is in the way, resulting in the ear farther from the sound source receiving a copy of the sound a little later than the near ear, creating an interaural time difference (ITD) between the ears. Also, the sound wave reaching the far ear will be attenuated compared to the near ear, creating an interaural level difference (ILD) between the ears. We generally perceive a sound source to be originating closer to the ear where it arrives earliest and with the greatest amplitude. It is found that if we vary the sound source location by azimuth (lateral position as related to directly in front of the listener), the ITDs and ILDs would vary as well. It would, however, have been a trivial matter, had the head not been in the way, to calculate the additional distance the sound would have to travel in order to reach the far ear and consequently deducing the appropriate ITD and ILD. The problem is also significantly complicated by the presence of the head in another aspect as there may not even be direct paths from the source to the far ear which may cause to account for the waves reaching the ears after propagating along the surface of certain irregular objects like different organs of human body. In fact, the diffraction of sound waves by such organs like human torso, shoulders, head and outer ears (pinnae) change its spectrum to a certain extent before it reaches the ear drums [4]. These changes may be described by the head related transfer function (HRTF) or head related impulse response (HRIR), which varies in a complicated way with azimuth, elevation, frequency and range, as well as gets altered considerably from person to person [9]-[10] as each individual has unique physical shapes. These HRTFs/HRIRs can be measured for an individual by placing small microphones deep in the ear canal and playing bursts of broadband noise at different spatial locations around the fixed head. In fact, it is seen that certain interesting temporal features, that are hidden in the phase response of the HRTF, can be revealed in HRIR and some researchers thus prefer HRIR synthesis over HRTF. However, the two

Manuscript received January 26, 2006.
A. Mitra is with the Department of Electronics and Communication Engineering, Indian Institute of Technology (IIT) Guwahati, North Guwahati - 781039, India. E-mail: a.mitra@iitg.ernet.in.
C. Ardil is with the Azerbaijan National Academy of Aviation, Baku, Azerbaijan. E-mail: cemalardil@gmail.com.

domains are mathematically equivalent because any model that captures the time-domain response without error also captures the frequency-domain response correctly.

Many generalized computational models [11]-[13] have also been proposed to put back the measured HRTF's in order to simplify spatial audio systems. Some researchers have investigated the effects of synthesizing spatial audio using distorted versions of the measured HRTFs, while some other repeated experiments using progressively *smoothed* versions of the individual's HRTFs and it is found that the HRTFs could be smoothed drastically, retaining only the gross features of the original measurements and still be surprisingly effective at generating spatialized sound. These computed HRTFs for a location, however, would contain both the ITD and ILD cues and the spectral filtering effects, primarily due to the pinnae. Because the elevation of the sound source relative to both ears is same, the spectral filtering effects would also be expected to be very similar. With this spectral cue being roughly the same at both ears, a significant interaural frequency difference is never found to compare between the ears. This explains the empirical observation that, lacking visual cues, we are much better at localizing sound in the azimuthal plane than in the elevation plane. However, there are significant variations in elevation perception from person to person. Additionally, front sounds frequently appear to be very close, as well as front/back reversals are also common.

We propose a computationally simple signal processing model of the HRTF/HRIR in this paper for synthesizing binaural sound from a monaural source in a room. The model reflects a precise one-to-one association with the room model, shoulders, head, and pinnae, where each parameter adds up a specific feature to the impulse response/transfer function. Note that the proposed structural model is primarily developed based on the approach taken in [7]. However, the main virtue of the proposed approach is to include a realistic room model for front end processing of the monaural sound. Additionally, a new azimuth cue is introduced to have better computational results. Also, a reasonable time delay is proposed, better than that of [7], as pinnae feature including elevation cue.

The paper is organized as follows. Section II briefly discusses about different HRTF models, starting from Rayleigh model to existing structural models in different subsections. The proposed structural model is given in Section III by providing basic pole/zero form and parameter values along with azimuth and elevation cues. Section IV deals with computational results and observations and the paper is concluded in Section V.

II. Models of HRTF

A. Rayleigh Model

HRTF can be calculated theoretically with a direct approach by solving the wave equation, subject to the boundary conditions presented by the physical parameters like torso, shoulders, head, pinnae, ear canal and ear drum. The theoretical approach, however, takes a lot of analysis thereby making it computationally appalling. The first attempt to solve such a problem with less computational complexity was made by

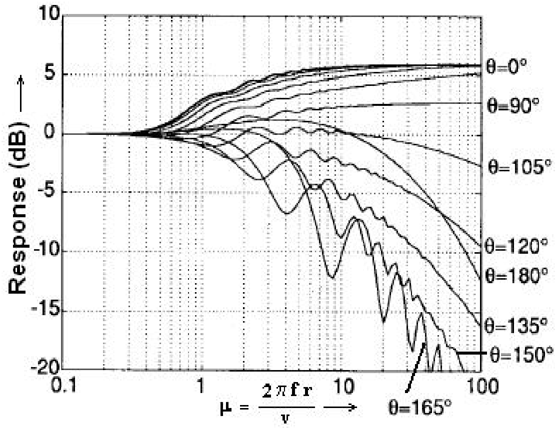

Fig. 1. Rayleigh head model frequency response plot, where f denotes frequency, r is the radius and v represents the speed of sound.

Lord Rayleigh many years ago by modeling the head as a sphere and solving the equations for wave propagation around this rigid sphere [14]. There, the deduced transfer function $H(\omega, \theta)$ is expressed as the ratio of phasor pressure at the surface of the sphere to the phasor free-field pressure, where ω is the radian frequency and θ is the angle of incidence. The results are given in terms of normalized frequency μ, scaled with the radius r, as

$$\mu = \frac{\omega r}{v} \tag{1}$$

with v being the speed of sound ($\approx$ 343 m/s at normal temperature). For a common 8.75 cm radius adult human head [15], $\mu = 1$ corresponds to a frequency of about 624 Hz. Fig. 1 shows the normalized amplitude response of the well known Rayleigh head model accounting for the so-called *head shadow*, the loss of high frequencies when the source is on the far side of the head. For $\mu > 1$, the time difference ($\Delta t(\theta)$) between the wave that arrives at the observation point and the wave that would arrive at the center of the sphere in free space is approximated [3] as

$$\Delta t(\theta) = \begin{cases} -\frac{r}{v} cos\theta & \text{for } 0 \leq \theta < \frac{\pi}{2} \\ -\frac{r}{v}(\frac{\pi}{2} - |\theta|) & \text{for } \frac{\pi}{2} \leq \theta < \pi. \end{cases} \tag{2}$$

On the other hand, the relative delay becomes approximately 1.5 times of the value predicted by eq. (2) when $\mu < 1$ and $\mu \to 0$ [15].

A first-order approximation model of HRTF can be given by simple linear filters that provide the relative time delays specified by eq. (2). This offers useful ITD cues, but no ILD cues. Additionally, the resulting ITD will be independent of frequency, contradicting Kuhns observations [15]. Both of these problems, however, can be treated by adding a minimum phase filter to account for the magnitude response. In [7], it has been reported that effective results can be obtained by cascading a delay element corresponding to eq. (2) with the

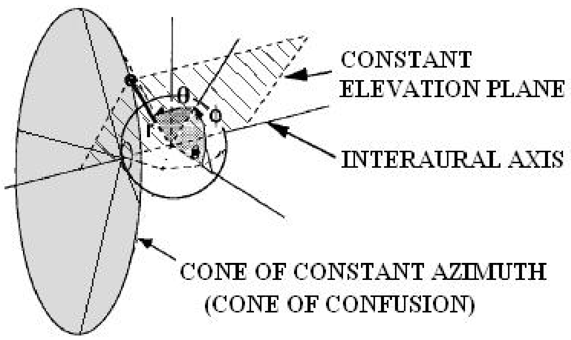

Fig. 2. Side view of *cone of confusion*, where different spatial locations produce identical ITDs and ILDs. Note that only the elevation distinguishes the front from back.

following single-pole, single-zero head-shadow filter

$$H_{HS}(\omega,\theta)=\frac{\omega_0+j\alpha\frac{\omega}{2}}{\omega_0+j\frac{\omega}{2}} \quad (3)$$

where $j=\sqrt{-1}$, ω_0 corresponds to normalized frequency $\mu=1$, i.e., $\omega_0=\frac{343}{r}$ rad/sec and the coefficient α is a function of angle of incidence θ with the range [0, 2] and is mainly concerned with controlling the location of zero. If α is chosen as

$$\alpha(\theta)=(1+\frac{\alpha_{min}}{2})+(1-\frac{\alpha_{min}}{2})cos(\pi\frac{\theta}{\theta_{min}}) \quad (4)$$

with $\alpha_{min}=0.1$ and $\theta_{min}=5\pi/6$, the results produce a considerable good approximation to the ideal solution shown in Fig. 1. At low frequencies, H_{HS} also introduces a group delay

$$T_g=\frac{(1-\alpha)}{2\omega_0}=\frac{r(1-\alpha)}{2v} \quad (5)$$

which appends to the high-frequency delay given by eq. (2). Note that at $\theta=0^o$, the head-shadow filter provides a low-frequency delay of 1.5 times the value given in eq. (2) which is mentioned in [15]. An approximate but simple signal processing implementation of Rayleighs *head shadow* solution can thus be provided by the model based on eq. (2)-(4).

The Rayleigh head model is appealing because it is relatively simple and it has proven to be successful in simple localization experiments. The above solution also highlights some interesting aspects of the ITDs and ILDs. First, ILDs are not linear with frequency. Low frequencies do not appear to be shadowed much by the head because they have a wavelength which is greater than the diameter of the head. Frequencies above 1.3 kHz have a wavelength smaller than the diameter of the head, and can be attenuated a great deal. It follows then that high frequency components are much more important to us when evaluating ILDs because they are most affected by the head shadowing resulting from the location of the sound source. Secondly, for frequencies above 1.3 kHz ITDs become ambiguous. When the wavelength of the sound wave is smaller than the diameter of the head, an ITD is greater than one period of the wave. In this case, comparing the phase difference between the waves arriving at each ear does not provide a unique ITD due to an aliasing effect. It follows from this fact that low frequency components of a sound are much more important to us in evaluating ITDs because the phase difference between the ears can still provide us with a unique ITD.

Nevertheless, it has long been recognized that approximate Rayleigh model is not a complete solution of our sonic cues for localization. It is apparent even considering a constant elevation of the sound source that there are multiple locations which would produce identical ITDs and ILDs. If the elevation of the sound source is allowed to vary, the problem is compounded and we get a whole cone of points in three dimensional space that would produce the same ITDs and ILDs. This phenomenon has been labeled the *cone of constant azimuth*, or, *cone of confusion* because of the ambiguity of the ITDs and ILDs that are generated from these locations. Fig. 2 illustrates the "cone of confusion" where different spatial locations would produce identical ITDs and ILDs. Note that, in these cases, the only way to distinguish the front ear from the back ear is by elevation angle. Thus, models based on elevation estimation is reviewed next.

B. Models with Pinna Elevation Effects

The pinna effects have been investigated mainly because of its importance for elevation estimation. Batteau [16] first proposed a simple two-echo model of the pinna that produced elevation effects. In measuring elevation, the interaural polar system should be given emphasis as in that system surfaces of constant azimuth are cones of essentially constant interaural time difference, as shown in Fig. 2. Here the azimuth θ is confined within the interval $[-\pi/2,\pi/2]$ and the elevation ϕ ranges from $-\pi$ to π. This alternately indicates that with interaural-polar coordinates, it is elevation rather than azimuth that distinguishes front from back. For simplicity, the measurements on human can be subjected to the frontal half space, so that the elevation is restricted to the interval within $[-\pi/2,\pi/2]$ as shown in [7], [17]. It is reported in both the works that the obtained results contain sharply positive sloping diagonal events due to a shoulder reflection and its replication by pinna effects. Other fainter patterns are seen, perhaps due to resonances of the pinna cavities (cavum concha and fossa), or to paths over the top of the head and under the chin. However, it is particularly interesting that the timing patterns, including the azimuth variation, are remarkably similar for the ipsilateral and the contralateral (near and far) ears.

C. Series Expansion Models

On a physical basis, if HRTF's are completely modeled by a relatively small number of physical parameters – the average head radius, head eccentricity, maximum pinna diameter, cavum concha diameter, etc., then, this suggests that the intrinsic dimensionality of the HRTFs might be small, and that their complexity primarily reflects the fact that we are not viewing them correctly.

Several works have applied principal components analysis (or, equivalently, the Karhunen-Loéve expansion) to the log magnitude of the HRTF [9], or to the complex HRTF itself [18] for simpler representations. This produces both a directionally independent set of basis functions and a directionally dependent set of weights for combining the basis functions. The usage of Fourier series expansion [19] is also possible to exploit the periodicity of the HRTF. The results of expanding the log-magnitude can be viewed as a cascade model, and is useful for representing head diffraction, ear-canal resonance, and other operations that occur in sequence. The results of expanding the complex HRTF can be considered as a parallel model to account for multipath phenomena such as shoulder echoes, pinna echoes etc.

In all of these cases, it has been found that a relatively small number of basis functions are sufficient to represent the HRTF, and series expansions have proved to be an effective tool, although not accurate, for studying the characteristics of the data.

D. Structural Models

Another approach to HRTF modeling banks on a simplified analysis of the physics of wave propagation and diffraction. Lord Rayleigh's spherical model can be viewed as a stepping stone in this direction, as can Batteaus two-echo theory of the pinna and the more sophisticated analysis in [20]. A mentionable work along these lines is the model developed by Genuit [21], where 34 measurements that characterize the shape of the shoulders, head, and pinnae have been identified. To formulate the problem analytically, he approximated the head and pinnae by tubes and circular disks and proposed combining these solutions heuristically to create a structural model.

The model in [21] has several appealing characteristics. First, each parameter is responsible for some well identified and significant physical phenomenon. Second, its economical and well suited nature for real-time implementation. Third, its ability to relate filter parameters to human physical measurements. Furthermore, it is neither a cascade nor a multipath model, but rather a structural model that is naturally suited to the problem for which it has been successfully incorporated in a commercial product [22]. However, to the best of our knowledge, it has neither been objectively evaluated, nor its detailed structure has been revealed.

III. The Proposed Structural Model

It is known that the HRTF and the HRIR are mainly functions of four variables – three spatial coordinates and either frequency or time. As mentioned earlier, both the functions are quite intricate and they differ significantly from person to person. It should be noted that the most effective systems for binaural sound synthesis should store large tables of digital filter coefficients derived from HRTF/HRIR measurements for individual subjects. Replacing such tables by functional approximations with simpler structure has also been investigated by researchers. This can therefore be fit as simply a system identification problem, for which there are well known standard procedures. However, the major problems that complicate the system identification task are (a) difficulty to approximate the effects of wave propagation and diffraction by low-order parameterized models that are both simple and accurate, and (b) non-availability of quantitative criterion for measuring the ability of an approximation to capture directional information that is perceptually relevant.

In order to overcome the setbacks as mentioned above, and mainly the last one, we propose here a structural model consisting of basic pole/zero models which reflect an approximately simple practical solution to control spatial/temporal conception. We present this in the following.

A. The Pole/Zero Model

Although eq. (3) is a simple example of a rational function approximation to an HRTF which can produce fairly satisfactory azimuth effects, its effectiveness can be considerably enhanced by adding an all-pass section to reflect the propagation delay and thus the interaural time difference. This leads to a head-shadow model of the form

$$H_{HS}(\omega, \theta - \theta_{ear}) = \frac{2\omega_0 + j\alpha'(\theta - \theta_{ear})\omega}{2\omega_0 + j\omega} e^{-j\omega T_d(\theta - \theta_{ear})} \tag{6}$$

where $T_d(\theta)$ is obtained by adding r/v to eq. (2) for keeping the delays causal, ω_0 is the same as given earlier and θ_{ear} implies the location of the entrance to the ear canal, e.g., $+120^o$ for the right ear and -120^o for the left ear. The parameter $\alpha'(\theta)$, however, is not the earlier one as given in eq. (4) and is specified with a new simplified form as

$$\alpha'(\theta) = 2 - 0.68\theta^2 \tag{7}$$

where θ is expressed in radians. This rudimentary model has many of the characteristics that we desire. In listening tests, the apparent location of a sound source varies smoothly and convincingly between the left and right ear as θ varies within $[-\pi/2, \pi/2]$. It can be also individualized by adjusting its parameters r and θ_{ear}. Although it does not factor into a function of frequency times another of azimuth, it is sufficiently simple for implementation purpose in real-time. Furthermore, the elevation effects are included in our model via time delay parameters to capture the pinna response with an easy approximation formula. This is discussed in details in the subsection, estimation of parameters.

B. Overall Structure of the Model

We start our structural model with a normal room configurations. It has long been recognized that the reflections in a room give us sonic cues about the ambient environment, which add to our perception that a sound presented over headphones is external, rather than coming from inside the head. To try and use these effects, we start with adding a room model component to our proposed approach at the front end. This is a simple component, which produces an attenuated and delayed copy of the original sound and is then fed through the head and pinna model with the same direction parameters as the sound to be localized. This means that we are simulating that the first room echo to arrive at the listener is also coming

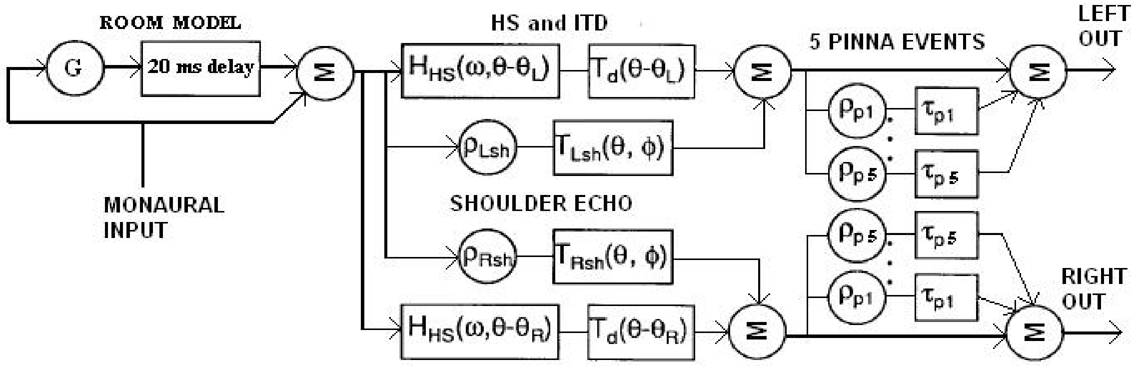

Fig. 3. A block diagram representation of the proposed HRTF modeling. Here separate modules account for room delay, head shadow, shoulder reflections and pinna effects.

from the same direction as the sound source (possibly from a wall directly on the other side of the source from the listener location). This should not only aid in the sense of externalization, but also provide an additional cue for localization of the primary sound. The delay time we have used for this room echo was 20 ms, corresponding to a reflection coming from a wall approximately 3.43 meters behind the sound source. This time delay was kept constant, meaning essentially we model a spherical room with a radius of 3.43 meters. This doesn't necessarily reflect any real physical environment, but work fairly with aiding the externalization and localization in informal listening tests.

We then propose an alternative structural model to account for head-shadow, shoulder and pinnae effects. It is based on combining an infinite impulse response (IIR) head-shadow model along with a finite impulse response (FIR) shoulder-echo model and an FIR pinna-echo model, reflecting the overall structure of eq. (6). A specific form of our model is shown in Fig. 3. Here the ρ's are reflection coefficients and the T's and τ's are different time delays. Such a structure has been chosen mainly because sounds can reach the neighborhood of the pinnae via two major paths: diffraction around the head and reflection from the shoulders. The incident waves in both the cases are changed by the pinna before entering the ear canal. Note that our approximation model can be further refined or simplified depending on assumptions. For example, investigation of the shoulder echo patterns reveals that the shoulder reflection coefficients ρ_{Lsh} and ρ_{Rsh} vary with elevation [7]. The other considerable point is that since the shoulder echoes reach the ear from a different direction than the direct sound, they should propagate through a different pinna model. Further, these directional relations would change if the listener turns his/her head. Informal listening tests, however, indicated that the modeled shoulder echoes did not have a strong effect on the perceived elevation of a sound, and this component can be eluded from our evaluation tests.

In [4], pinna model is divided into three classes: resonator, reflective, and diffractive. Our model is mainly reflective, as well as negative values are also allowed for two of the pinna reflection coefficients, which might correspond to a longitudinal resonance. These coefficients and their corresponding delays alter with both azimuth and elevation. For a symmetrical head, the coefficients for the left ear for an azimuth angle θ should be the same as those for the right ear for an azimuth angle $-\theta$. Although the pinna events definitely depend on azimuth, it has been found that their arrival times are quite similar for the near and the far ear. This is, however, not true for the reflection coefficients as ILD is dependent on elevation. Thus, differences in the pinna parameters are required to correspond for the variation of the ILD with elevation. We discuss it next.

C. Estimation of Parameters

A thorough investigation of the impulse responses from the available database indicated that most of the pinna activity occurs in the first 0.6 ms, which corresponds to 26 samples

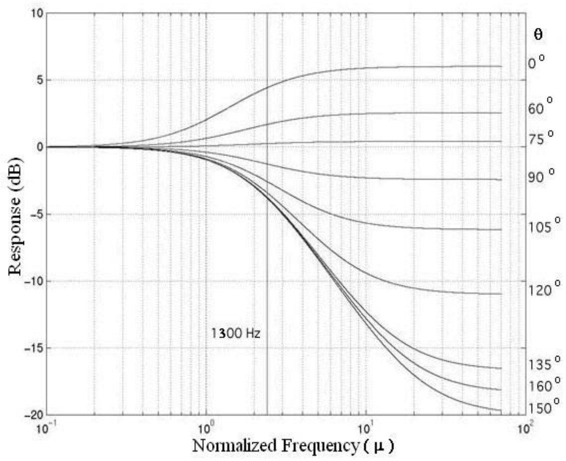

Fig. 4. Proposed head-shadow model frequency response plot with eq. (5), implemented in MATLAB.

at a 44.1 kHz sampling rate. Thus, a single 26-tap FIR filter was selected for the elevation model. Five pinna events were decided to be sufficient to capture the pinna response. There are two quantities associated with each event, a reflection coefficient ρ_{pn} and a time delay τ_{pn} for $n = 1, 2, ..., 5$ as shown in Fig. 3. While the values of the reflection coefficients were not critical and can be assigned constant values, the time delays can vary with azimuth, elevation, and the listener and thus has to assigned very cautiously. Because functions of azimuth and elevation are always periodic in those variables, it would be simple to approximate the time delay using simple sinusoids. We propose the following formula, to some extent similar to what has been found empirically in [7], that seem to offer a satisfactory agrement to the time delay for the nth pinna event:

$$\begin{aligned} \tau_{pn}(\theta, \phi) &= A_n[sin(\pi/4 + \theta/2 - \phi/2) \\ &\quad + sin(\pi/4 - \theta/2 - \phi/2)] + O_n \end{aligned} \tag{8}$$

where A_n is an amplitude and O_n is an offset. In general, these parameters should be tested and given the best fit value from experimental results. In our case, we used $A_2 = 0.5$ and 2.5 for A_3 to A_5, and $O_2, O_3, O_4, O_5 = 2, 4, 7, 11$ respectively. As the time delays computed by eq. (8) hardly coincided exactly with sample times, therefore, in implementing the pinna model as an FIR filter, linear interpolation has been used to split the amplitude between surrounding sample points.

IV. Computational Results and Discussions

The proposed model has been simulated in MATLAB with eq. (6), where the delay has been chosen according to $T_d(\theta)$ and eq. (8) and the simulations have shown considerable good approximations. The resulting head-shadow filter response is shown in Fig. 4. The total structure has also been simulated with a chosen range of spatial locations at discrete intervals in azimuth and elevation making them far enough apart to be distinguishable and was played to a listener over earphones. For each location and each model we created a '.wav' file of pulses of broadband Gaussian white noise, bandlimited to 0.2-10 kHz. Each pulse was 750 ms in duration and was enveloped at the beginning and end with a 20 ms sine squared curve. The first two pulses of each sample were not filtered by the HRTF in order to provide the listener with a frame of reference. These pulses should sound as if they are coming from inside the head. The remaining four pulses were convolved with the transfer function corresponding to the elevation and azimuth that were under test. This proposed model has been tested on one subject with a total of five tests and the standard deviation for the model validation was found to be 15.5^o. A large number of tests on many subjects is also believed to yield satisfactory result with the same model.

Note that our investigation was limited to the frontal half space, and we did not address front/back discrimination problems to maintain the simplicity of the model. While head motion is highly effective in resolving front/back confusion, the shoulder echo may play an important role for static discrimination. Further improvement may be done in this area. An even more important area for improvement is the introduction of range cues. Although there are many cues for range, they go beyond HRTF's intrinsically, and involve the larger problem of how to create a virtual auditory space.

Another drawback of our evaluation process has been by limiting the model to how well it substituted for experimentally measured HRIR's. Unfortunately, this does not answer the question of how well the model creates the deceptive appearance that a sound source is located at a particular point in space. In particular, even though ignoring timbre differences between the model and the measured HRIR would have been ideal, timbre matching may have occurred. It is expected that future works would treat the absolute localization tests with importance.

V. Conclusions

A computationally simple yet effective signal processing model for synthesizing binaural sound from a monaural source in a common classroom has been proposed. Apart from the room model, the structure contains separate components for azimuth (head shadow and ITD) and elevation (pinna and shoulder echoes). The simplicity of the IIR head-shadow filters and FIR echo filters enables inexpensive real-time implementation without the need for special purpose DSP hardware. Additionally, the parameters of the model can be adjusted to match the individual listener and to produce individualized HRTFs. The proposed structural model has primarily been developed based on the approach taken in [7]. However, the main virtue of the proposed approach is to include a realistic classroom model for front end monaural sonic processing. In addition to that, a new azimuth cue is introduced to have better computational results. Also, a reasonable time delay is proposed, better than that of [7], as pinnae feature including elevation cue. The model is found to yield satisfactory results in comparison with existing models and therefore is believed to play a key structure in future spatial auditory interfaces.

REFERENCES

[1] E. M. Wenzel, "Localization in virtual acoustic displays," *Presence*, vol. 1, pp. 80-107, Winter 1992.
[2] G. Kramer, Ed., *Auditory Display: Sonification, Audification, and Auditory Interfaces*. Reading, MA: Addison-Wesley, 1994.
[3] J. P. Blauert, *Spatial Hearing*, rev. ed. Cambridge, MA: MIT Press, 1997.
[4] S. Carlile, "The physical basis and psychophysical basis of sound localization," in *Virtual Auditory Space: Generation and Applications*, S. Carlile, Ed. Austin, TX: R. G. Landes, 1996, pp. 27-28.
[5] VREPAR Project (1996, November). Binaural Technology in Virtual Auditory Environments [Online]. Available: http://www.ehto.org/ht_projects/vrepar/SOUND.HTM.
[6] S. Coppin *et. al.* (2000, April). Sound Localization using Head Related Transfer Functions [Online]. Available: http://www-ece.rice.edu/c̃rozell/courseproj/431report.
[7] C. P. Brown and R. O. Duda, "A Structural Model for Binaural Sound Synthesis," *IEEE Trans. Speech Audio Processing*, vol. 6, no. 5, pp. 476-488, Sept. 1998.
[8] D. R. Begault, *3-D Sound for Virtual Reality and Multimedia*. New York: Academic, 1994.
[9] D. J. Kistler and F. L. Wightman, "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," *J. Acoust. Soc. Amer.*, vol. 91, pp. 1637-1647, Mar. 1992.
[10] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization using nonindividualized head-related transfer functions," *J. Acoust. Soc. Amer.*, vol. 94, pp. 111-123, July 1993.
[11] P. H. Myers, "Three-dimensional auditory display apparatus and method utilizing enhanced bionic emulation of human binaural sound localization," U.S. Patent no. 4 817 149, Mar. 28, 1989.
[12] R. O. Duda, "Modeling head related transfer functions," in *Proc. 27th Ann. Asilomar Conf. Signals, Systems and Computers*, Asilomar, CA, Nov. 1993.
[13] B. Shinn-Cunningham, "Adapting to Remapped Auditory Localization Cues: A Decision Theory Model," *Perception and Psychophysics*, vol. 62, no. 1, pp. 33-47, 2000.
[14] J. W. Strutt (Lord Rayleigh), "On the acoustic shadow of a sphere," *Phil. Trans. R. Soc. London*, vol. 203A, pp. 87-97, 1904; *The Theory of Sound*, 2nd ed. New York: Dover, 1945.
[15] G. F. Kuhn, "Model for the interaural time differences in the azimuthal plane," *J. Acoust. Soc. Amer.*, vol. 62, pp. 157-167, July 1977.
[16] D. W. Batteau, "The role of the pinna in human localization," *Proc. R. Soc. Lond. B*, vol. 168, pp. 158-180, 1967.
[17] Y. Hiranaka and H. Yamasaki, "Envelope representation of pinna impulse responses relating to three-dimensional localization of sound sources," *J. Acoust. Soc. Amer.*, vol. 73, pp. 291-296, 1983.
[18] J. B. Chen, B. D. Van Veen, and K. E. Hecox, "A spatial feature extraction and regularization model for the head-related transfer function," *J. Acoust. Soc. Amer.*, vol. 97, pp. 439-452, Jan. 1995.
[19] R. O. Duda, "Elevation dependence of the interaural transfer function," in *Binaural and Spatial Hearing in Real and Virtual Environments*, R. H. Gilkey and T. R. Anderson, Eds. Mahwah, NJ: Lawrence Erlbaum, 1997, pp. 49-75.
[20] E. A. Lopez-Poveda and R. Meddis, "A physical model of sound diffraction and reflections in the human concha," *J. Acoust. Soc. Amer.*, vol. 100, pp. 3248-3259, 1996.
[21] K. Genuit, "Method and apparatus for simulating outer ear free field transfer function," U.S. Patent 4 672 569, June 9, 1987.
[22] K. Genuit and W. R. Bray, "The Aachen head system," *Audio*, vol. 73, pp. 58-66, Dec. 1989.