Title
The Use of Head-and-Torso Models for Improved Spatial Sound Synthesis

Authors
Algazi, Ralph
Duda, Richard O
Thompson, Dennis M.

Publication Date
2002

Peer reviewed

Audio Engineering Society
Convention Paper
Presented at the 113th Convention
2002 October 5–8 Los Angeles, CA, USA

THE USE OF HEAD-AND-TORSO MODELS FOR IMPROVED SPATIAL SOUND SYNTHESIS

V. Ralph Algazi[1], Richard O. Duda[1], and Dennis M. Thompson[1]

[1] *CIPIC Interface Laboratory, University of California, Davis CA, 95616-8553, USA*

Correspondence should be addressed to Richard O. Duda (`rod@duda.org`)

ABSTRACT

This paper concerns the use of a simple head-and-torso model to correct deficiencies in the low-frequency behavior of experimentally measured head-related transfer functions (HRTFs). This so-called "snowman" model consists of a spherical head located above a spherical torso. In addition to providing improved low-frequency response for music reproduction, the model provides the major low-frequency localization cues, including cues for low-elevation as well as high-elevation sources. The model HRTF and the measured HRTF can be easily combined by using the phase response of the model at all frequencies and by "cross-fading" between the dB magnitude responses of the model and the measurements. For efficient implementation, the exact snowman HRTF is approximated by two time delays and two first-order IIR filters. Because the poles are independent of the location of the virtual source, this supports a simple real-time implementation that allows for arbitrarily rapid head and source motion.

1. INTRODUCTION

Many systems for spatial sound synthesis employ experimentally measured head-related transfer functions (HRTFs). Unfortunately, it is both time-consuming and experimentally difficult to measure HRTFs accurately. Low-frequency measurements are particularly problematic, partly because large loudspeakers are required, and partly because even good anechoic chambers reflect long-wavelength sound waves. The use of windowing to eliminate room reflections also modifies the response to low frequencies. Without low-frequency compensation, music synthesized with HRTFs often sounds "thin" and lacking in bass. Ad hoc methods for boosting the bass can improve the sound quality, but such compensation can disturb localization cues. Finally, most HRTF-based systems are not able to convincingly position virtual sound sources at very low elevations.

The low-frequency performance of HRTF-based systems can be significantly improved by the judicious use of theoretical HRTFs derived from simple models. The simplest of these is the classical spherical-head model [1, 2]. It provides the low-frequency interaural time delay (ITD) and the interaural level difference (ILD), which are the two most important cues for azimuth. The spherical head model all by itself does not provide any elevation cues, and sounds synthesized using this model alone are weakly externalized and sometimes experienced as being elevated.

This paper describes a simple model that accounts for the contributions of the torso as well as the head to the HRTF. This so-called "snowman model" is based on approximating both the head and the torso by rigid spheres. The effect of these spheres on sound waves is in turn approximated by simple signal-processing modules. These modules consist of two frequency-independent time delays and two fixed-pole, variable-zero, first-order IIR filters. The time delays account for the propagation of sound across the head and the reflection by the torso for sounds at high elevation. The filters accounts for head shadow for sources on the contralateral side and for torso shadow for sources at low elevation. We show how this filter model can be used to compensate experimentally measured HRTFs to produce virtual auditory sources that have a natural sounding bass response while retaining good source localization cues.

We begin by reviewing some of the observed characteristics of HRTFs that are caused by the torso and that have been explained and simulated by a simple head-and-torso geometric model. We then present the exact HRTF for the snowman model at all elevations. This HRTF indicates the major modes of behavior that occur as the elevation ranges from overhead to underneath. A snowman filter model is then developed by building upon a filter model for a single sphere and by analyzing the combined geometry of the two spheres that compose the snowman. Filter implementation and applications are briefly discussed, with emphasis on the low-frequency compensation of measured HRTFs.

2. THE SNOWMAN MODEL

Although it has been known for some time that the torso has a measurable influence on the HRTF, reports on the psychoacoustic importance of torso effects for sound localization have been inconclusive (see [3] for a review of previous investigations). The best known effect is the so-called "shoulder bounce," the specular reflection from the upper torso that occurs when the source is above the subject. The reflected wave introduces a series of comb-filter notches into the HRTF spectrum. For overhead sources, the first notch appears at frequencies as low as 600 Hz, which makes it potentially important for sources that do not contain much high-frequency energy.

In a previous study, we showed that this torso reflection provides a weak but significant elevation cue for sources that are away from the median plane [3]. Furthermore, the effect can be simulated with a simple model. When the source spectrum was limited to 3 kHz to prevent the pinnae from providing elevation cues, essentially the same localization performance was obtained when the measured HRTF was replaced by the HRTF for a spherical-head model augmented by torso reflections computed from an ellipsoidal torso model.

In general, the longest time delay between the direct pulse and the reflected pulse occurs when the source is overhead. As the source elevation is reduced, the delay time shrinks, eventually becoming zero when the ray from the source to the ear becomes tangent to the torso. The set of tangent rays defines what we call the "torso-shadow cone" (see Figure 1). A sound source outside of the torso-shadow cone produces a torso reflection. The reflection disappears when the source is inside the torso-shadow cone, where the source is shadowed by the torso instead.

Although an ellipsoid fits the human torso better than does a sphere, the geometry of a spherical torso is easier to analyze. This gives rise to the so-called "snowman model," in which the body is approximated by a spherical head directly above a spherical torso. The snowman model is defined by three parameters — the head radius a, the torso radius b and the neck height h (see Figure 2a).[1] In principle, the model can be adapted to fit

[1] One can further elaborate the model in various ways. For

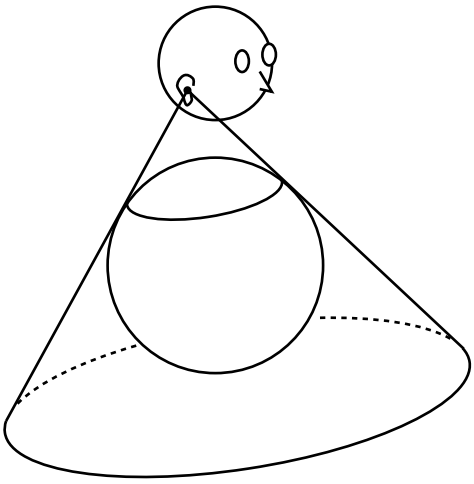

Fig. 1: The torso shadow cone, which is defined as the set of rays from the ear that are tangent to the torso. A source outside the torso shadow cone gives rise to a torso reflection. A source inside the torso shadow cone is shadowed by the torso.

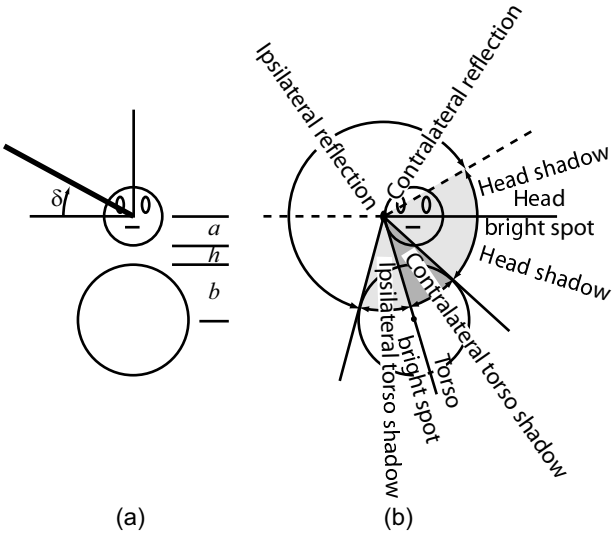

Fig. 2: (a) Frontal view of the snowman model. The head and torso are coaxial spheres. The defining parameters are the head radius a, the torso radius b and the neck height h. The range for the elevation angle δ is $-90^\circ \leq \delta < 270^\circ$. (b) Zones for the right-ear response. The torso-shadow cone is divided into an ipsilateral and a contralateral zone by the torso bright spot. Similarly, the head-shadow zone is divided by the head bright spot.

specific individuals by determining the parameters from anthropometric data [4]. In particular, we have found that the following values are appropriate for the KEMAR mannequin: $a = 8.7$ cm, $b = 16.9$ cm, and $h = 5.3$ cm. These are the values that are used in this paper.

3. THE IMPULSE RESPONSE OF THE SNOWMAN MODEL

Although there is no analytical solution for the HRTF for the snowman model, it is possible to compute the HRTF quite accurately using numerical methods.[2] In general, the HRTF is a function of azimuth, elevation and range. We restrict ourselves to the case where the source is infinitely distant, which is a good approximation for human listeners whenever the source is more than a meter or so away. Although it is conventional to examine the behavior of the HRTF as a function of azimuth in the horizontal plane, or as a function of elevation in the median plane, the behavior of the HRTF in the frontal plane is more revealing. Thus, all of the results that we present are for the frontal plane.

We used the multipole-reexpansion method described in [6] to compute the HRTF in the frontal plane. The complex spectrum from 0 to 5 kHz was windowed with a Hamming window and inverted numerically to obtain the corresponding head-related impulse response (HRIR). The resulting right-ear frontal-plane HRIR for the snowman model using KEMAR parameter values is shown in Figure 3.

Several features of this response are of interest. First, we observe that the pulse arrives soonest when the elevation angle δ is near 0° and the source is aimed directly at the ear. The pulse arrives approximately 0.7 ms later when $\delta = 180^\circ$, a time that corresponds to the maximum interaural time difference or ITD. A more detailed

example, the head can be offset in front of the torso, and the ears can be displaced as desired. For simplicity, we assume that the head is directly above the torso and that the ears are located across a head diameter.

[2]The usual tradeoffs between spatial resolution, numerical precision, and computation time limit the accuracy that can be achieved. To reduce the computational requirements, we limited the maximum frequency to 5 kHz. Working with colleagues at the University of Maryland and Duke University, we have investigated three different numerical approaches — finite-difference methods, boundary-element methods, and multipole reexpansion. The computational issues are discussed in more detail in [5].

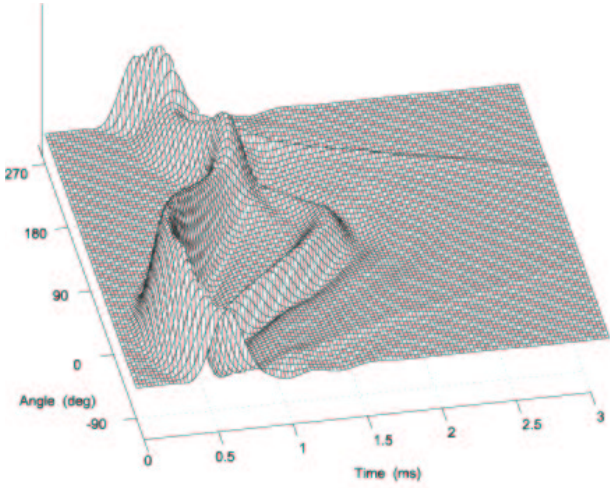

Fig. 3: The right-ear frontal-plane impulse response for the snowman model. The response is bandlimited to 5 kHz. The source is directly below when the angle is either 0° or 270°. The impulse response is strongest and arrives soonest when the source is more or less facing the ear (0°). Notice the presence of a torso reflection for angles roughly between −45° and 180°.

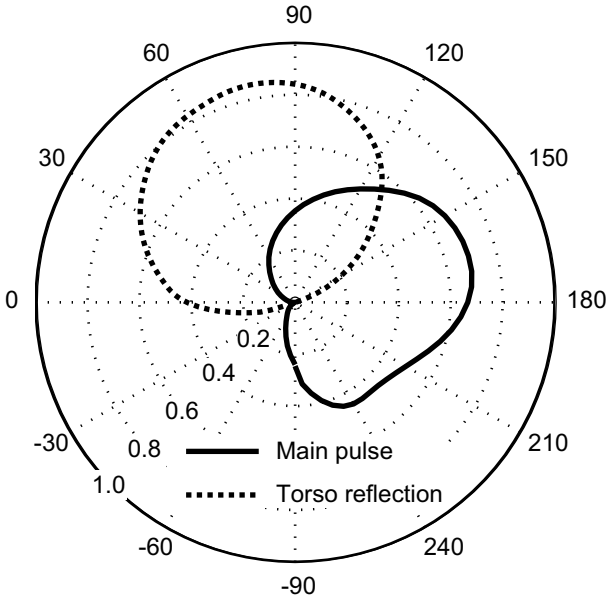

Fig. 4: The arrival time (in ms) for the main pulse (solid) and the torso reflection delay (dotted) for the right-ear frontal-plane impulse response.

picture of the arrival time is provided by the polar plot shown in Figure 4. The solid curve in this plot shows the time for the maximum response, using the earliest arriving pulse to define the time origin.[3] Here we see that the maximum delay time actually occurs closer to $\delta = 170°$. A similar plot for an isolated spherical head would produce a cardiod pattern that is symmetric above and below. Thus, the presence of the torso produces a small but measurable effect on the arrival time.

The height of the main pulse – which is related to the interaural level difference or ILD – also varies with elevation. It is greatest when δ is near 0°, but it also achieves a relative maximum somewhere around $\delta = 180°$. This relative maximum on the contralateral side defines the "bright spot" for the head, and is caused by the waves traveling around the head all coming together in phase at a point on the opposite side [2].[4]

The detailed behavior of the height of the main pulse is easier to see in the polar plot in Figure 5. This plot is slightly surprising, revealing that there are actually three angles at which the peak response achieves a local maximum: (a) $\delta = -40°$, (b) $\delta = 170°$, and (c) $\delta = 255°$. The peak centered at $\delta = -40°$ is very broad, with the value being very close to the value at $\delta = 0°$. Thus, this peak corresponds to direct incidence. The peak at $\delta = 170°$ corresponds to the bright spot for the head. The peak at $\delta = 255°$ is due to the torso. From the geometry in Figure 2, it can be seen that this is the elevation for a ray from the right ear through the center of the torso. Thus, there is a simple explanation for this additional peak — it corresponds to a "torso bright spot" that develops when waves traveling around the torso converge in phase in their path to the ear. This torso bright spot is of theoretical interest, but because it depends so strongly on the perfect symmetry of the spherical torso, it is not expected to appear in human HRIRs.

[3]The time at which the maximum response is achieved is a convenient measure of arrival time. However, it is important to recognize that the arrival time is frequency dependent. A good discussion of various ways to measure arrival time for HRTFs is given in [7].

[4]Although human heads do not have the extreme symmetry of the sphere, we have also observed bright spots in experimentally measured HRIRs for human subjects [8]. Because the ears are usually located below the center of the head, the bright spot is not found at $\delta = 180°$, but is usually observed at a point closer to $\delta = 155°$.

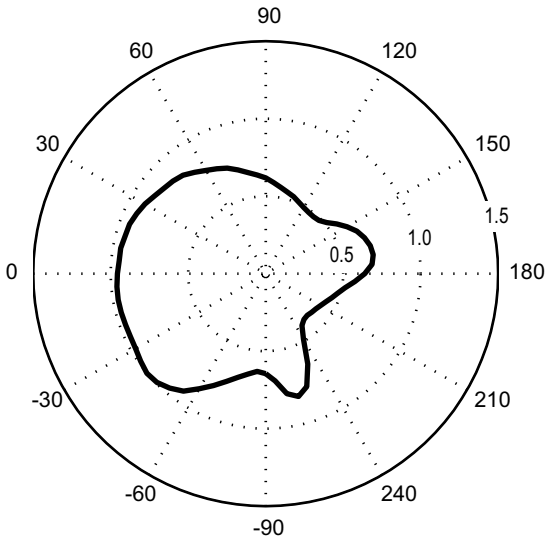

Fig. 5: The peak value of the impulse response for a source in the frontal plane. The response is largest when the elevation angle δ is in the broad range from $-60°$ to $30°$. A bright spot due to the head appears at $\delta = 170°$, and a second bright spot due to the torso appears at $\delta = 255°$.

Returning to Figure 3, one can see a clearly developed second pulse that follows the main pulse for elevations from about $-45°$ to somewhere past $180°$. This is the torso reflection or "shoulder bounce." The time between the arrival of the peak of the main pulse and the arrival of the peak of the torso reflection is shown as the dotted curve in Figure 4. The greatest time delay occurs when the source is overhead and on the ray from the center of the torso to the ear. From the geometry in Figure 2, this maximum time delay can be estimated as $\Delta T \approx 2(\sqrt{(a+h+b)^2+a^2}-b)/c = 0.89$ ms, where the speed of sound c is 343 m/s. This is quite close to the peak-to-peak separation of the pulses at $\delta = 75°$ (see Figure 4). The combination of the main pulse and the reflected pulse produces a comb-filter pattern in the frequency response. The first notch appears when ΔT is half a period, and appears around $f_0 = 1/2\Delta T \approx 560Hz$, which is quite a low frequency for spatial audio effects.

Less prominent but also significant are two thin ridges or tails that develop long time delays that change rapidly with elevation (see Figure 3). These features appear only in the torso-shadow cone. They can be attributed to waves that take the longer path around the torso from the source to the ear [5]. The tail on the contralateral side is particularly important because it interferes with the direct wave that is weakened by head shadow. Depending on the elevation angle, interference between these two waves can be either constructive or destructive, and this results in significant ripples in the frequency response in the torso-shadow zone.

4. THE FREQUENCY RESPONSE OF THE SNOWMAN MODEL

The frequency-dependent behavior of the snowman HRTF is revealed in the frequency response curves shown in Figure 6. Each of these curves approaches 0 dB as the frequency approaches 0 Hz. The three panels essentially correspond to the torso-reflection, head-shadow, and torso-shadow zones shown in Figure 2b.

We start with Figure 6a, which shows the results for δ from $-60°$ to $150°$. In this range of elevations, the response is dominated by the comb-filter patterns produced by torso reflections. As was observed above, the lowest frequency notch occurs at $f_0 = 1/2\Delta T$, where ΔT is the time delay between the main pulse and the torso reflection. This frequency changes rapidly with elevation, being lowest around $\delta = 75°$, i.e., when the source is on the ray from the center of the torso to the right ear. The depth of the first notch also varies with elevation, being approximately 5 dB when $\delta = 75°$. As the elevation angle increases toward $150°$ and enters the head shadow region, this notch rapidly moves up in frequency and becomes very deep.

In Figure 6b we see the development of the bright spot for the head, which appears with the almost flat frequency response at $\delta = 170°$. The torso shadow cone is entered around $\delta = 200°$. The notches that develop for $\delta > 200°$ are the combined result of head shadow and torso shadow, and are quite complicated. However, on average there is a significant loss of high frequencies in the contralateral torso shadow zone where both head shadowing and torso shadowing occurs.

The responses in Figure 6c are mostly in the ipsilateral torso shadow zone, and are more or less flat. The torso bright spot that can be seen at $\delta = 255°$. It is rather surprising that the response is so basically flat in this area, even when the source is directly below at $\delta = -90°$. As was observed above, the sphere is not a good approximation to the human torso, particularly at low elevations, and the detailed low-elevation behavior of the snowman HRTF is not replicated in human HRTFs.

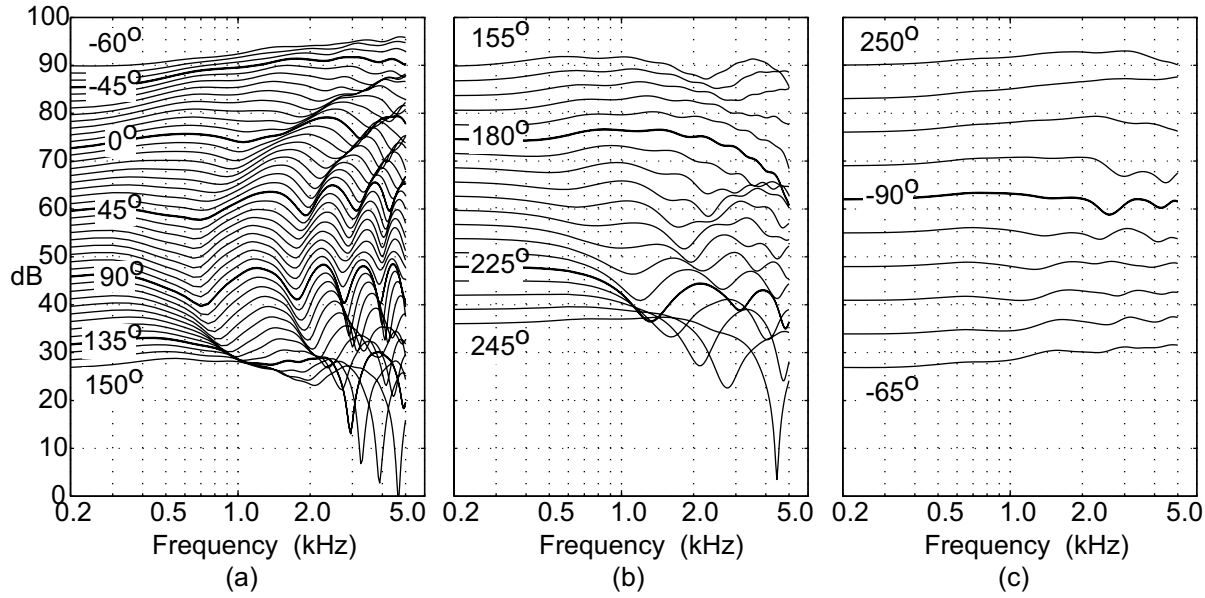

Fig. 6: The frequency response curves for the frontal-plane right-ear HRTF for the snowman. In all cases, the responses approach 0 dB at low frequencies. (a) Torso reflection zone. (b) head shadow zone. (c) Torso shadow zone.

5. FILTER MODELS

Through the use of numerical methods such as multipole-reexpansion, the snowman HRTF can be computed for any sound direction and any size head, neck and torso. For use in spatial sound rendering, the resulting impulse responses can be tabulated, with interpolation used to obtain HRIRs that are not in the table. However, there are advantages to approximating the solution by a filter model that not only allows adaptation to individual listeners but that also supports real-time sound rendering directly.

The approximation of HRTFs by pole/zero models has been thoroughly studied [9-11], and the exact snowman HRTF can be approximated as closely as desired by using a sufficiently high-order model. However, minimizing the error of fit to the theoretical solution does not necessarily lead to a model that subjectively "sounds right." Informal listening tests using the exact snowman HRTF revealed that the bright spots associated with the spheres used in the snowman model produces excessively bright-sounding responses in those areas, and that it is necessary to attenuate the bright spot responses to obtain more natural sounding results. As a result, the filter model we present is an engineering compromise that attempts to capture as simply as possible the first-order effects revealed by the snowman model while reducing the artifacts that stem from the over-idealized geometry of the sphere.

We take a structural approach to developing the filter model. The components of the model approximate the contribution of the separate spheres to the HRTF, either as sound reflectors or as sound shadowers. These partial contributions result in separate filter models for the head, for the head and torso when the torso acts as a reflector, and for the head and torso when the torso acts as a shadower.

6. A SIMPLE FILTER MODEL OF THE SPHERE

We begin the development of a filter model by reviewing a similar filter model for an isolated spherical head that was originally proposed by Brown and Duda [12]. As shown in Figure 7, that model consists of a shadowing filter H cascaded with a time delay ΔT. Both of these components are functions of the radius a of the sphere and the observation angle θ, the angle between a ray from the center of the sphere to the sound source and a ray from the center of the sphere to an observation point `P` on the surface of the sphere. For an HRTF model, `P` specifies an ear location, but it could be any point on

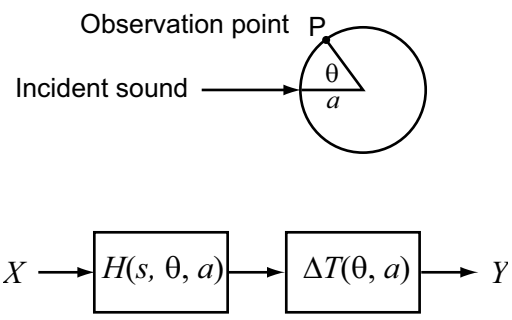

Fig. 7: Block diagram for the filter model for the HRTF of an isolated sphere. When used as a head model, two observation points are used, one at each ear. The shadowing filter H is a fixed-pole/variable-zero filter, where the position of the zero depends on the observation angle θ. The time delay ΔT also depends on θ, but is frequency independent.

the sphere. The time delay is given by Woodworth and Schlosberg's frequency-independent formula[5]

$$\Delta T(\theta, a) = \begin{cases} -\frac{a}{c}\cos\theta & \text{if } 0 \le |\theta| < \frac{\pi}{2} \\ \frac{a}{c}\left[|\theta| - \frac{\pi}{2}\right] & \text{if } \frac{\pi}{2} \le |\theta| < \pi \end{cases} \quad . \tag{1}$$

Because the time-delay component is all-pass, the magnitude spectrum of the HRTF is due entirely to the shadowing filter. As with the time delay, the frequency response is a function of the sphere radius a and the observation angle θ. Figure 8a shows a family of frequency response curves for the theoretical solution for a sphere of radius $a = 8.75$ cm. In all cases, the DC gain is 0 dB. The frequency response is basically flat when the $\theta = 105°$, and it is close to flat at the bright spot when $\theta = 180°$. High frequencies are boosted when $\theta < 105°$, and are cut when $\theta > 105°$. A maximum high-frequency boost of 6 dB occurs when the incidence is direct ($\theta = 0°$). The maximum high-frequency cut occurs near $\theta = 150°$, and is close to 20 dB at 5 kHz.

The Brown/Duda shadowing filter is a simple single-pole, single-zero filter whose response approximates this theoretical solution while reducing the strength of the bright spot. Its transfer function is given by

$$H(s, \theta, a) = \frac{\alpha\tau s + 1}{\tau s + 1} \tag{2}$$

where the parameter α (which we shall discuss) depends on θ and the time constant τ depends on a, with

$$\tau = \frac{2a}{c} \quad . \tag{3}$$

Because $H(0, \theta, a) = 1$, the DC gain is 0 dB for any value of θ or a. The parameter $\alpha(\theta)$ is the asymptotic high-frequency gain. Brown and Duda propose the formula

$$\alpha(\theta) = \left[1 + \frac{\alpha_{\min}}{2}\right] + \left[1 - \frac{\alpha_{\min}}{2}\right] \cos\left[\frac{\theta}{\theta_{\min}}\pi\right] \quad , \tag{4}$$

with the values $\alpha_{\min} = 0.1$ and $\theta_{\min} = 5\pi/6$ rad or 150° providing a good overall match to the theoretical results and an attenuated bright spot.

A key property of the shadowing filter is that the location of the pole depends only on the sphere radius, and is fixed for a particular listener.[6] For the standard 8.75-cm head radius, the corner frequency for this pole is $f_c = 1/2\pi\tau = c/4\pi a \approx 312$ Hz. By contrast, the location of the zero changes with the observation angle. At the angle θ_{flat} where $\alpha = 1$, the zero and pole cancel, and the frequency response is flat. From Eq. 4, it can be shown that this critical angle is given by

$$\theta_{\text{flat}} = \theta_{\min}\left[\frac{1}{2} + \frac{1}{\pi}\sin^{-1}\frac{\alpha_{\min}}{2 - \alpha_{\min}}\right] \quad . \tag{5}$$

If $\alpha_{\min} = 0.1$ and $\theta_{\min} = 150°$, $\theta_{\text{flat}} \approx 77.5°$.[7] If $\theta > \theta_{\text{flat}}$, the shadowing filter cuts the high frequencies, while if $\theta < \theta_{\text{flat}}$ it boosts them. Figure 8b shows a

[5] This formula gives the difference between the time when the sound arrives at the observation point and the time when it would arrive at the center of the sphere if the sphere were absent. Thus, it is negative when $|\theta| < \pi/2$. One can always add a constant greater than a/c to the result to obtain a positive time delay without changing the ITD. Eq. 1, which is the result of an elementary ray-tracing analysis, assumes that the source is infinitely distant. It can be generalized to account for range dependence, but the effects of range on the ITD are small [2]. Because ray-tracing is a short-wavelength concept, this formula gives the high-frequency time delay. It is well known that the low-frequency delay for the sphere is 50% greater than the high-frequency value. As Brown and Duda observe [12], this additional low-frequency phase delay is provided by the shadowing filter.

[6] In any filter, the pole locations determine the resonant or "natural" frequencies. The natural frequencies depend on the system, not on where the response of the system is observed. Thus, it is natural to use a model in which the location of the pole is independent of the observation angle [13]. As will be shown, this also simplifies real-time implementation when the position of the source with respect to the listener changes dynamically.

[7] This value for θ_{flat} is clearly much smaller than the 105° for the theoretical solution. A higher value can easily be obtained by changing $\alpha_{\min}$ and/or $\theta_{\min}$, but this leads to undesirable behavior at other observation angles. To obtain a better fit for the theoretical solution for the sphere, a formula for α that is more complicated than Eq. 4 would be required.

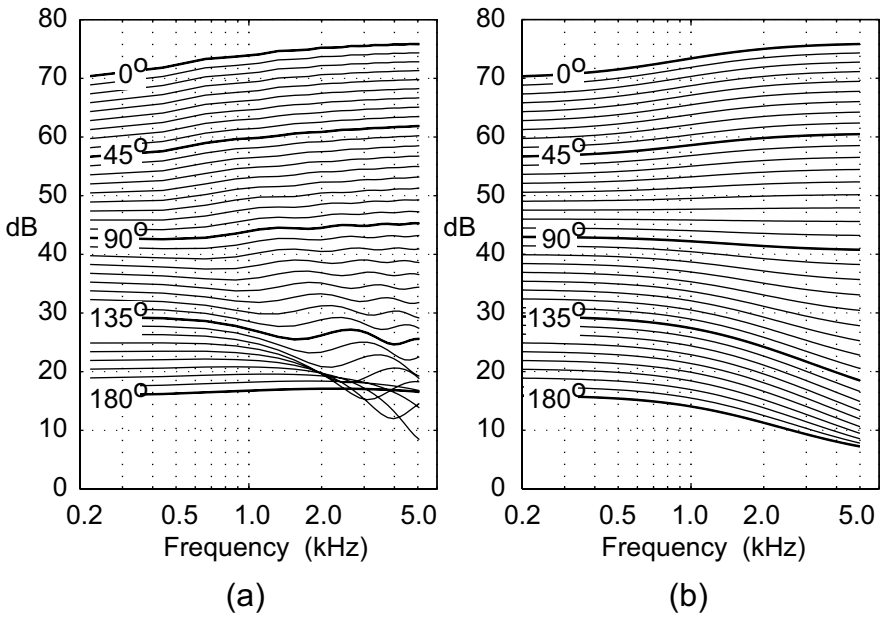

Fig. 8: The frequency response curves for the isolated sphere. The angles shown are the observation angles θ shown in Figure 7. (a) Exact theoretical solution. (b) Approximate shadowing filter.

complete family of frequency response curves. The maximum high-frequency cut is $20\log_{10}\alpha_{\min}$ (20 dB), and occurs when $\theta = \theta_{\min}(150°)$. If $\theta = 0$, the high frequencies are boosted by 6 dB. At the extreme when $\theta = \pi$, the response rises, though not as much as for the theoretical solution, and the filter produces a relatively weak bright spot, as desired.

7. A SNOWMAN FILTER MODEL

7.1 The Torso-Reflection Sub-Model

We now describe an approximate filter model for the snowman HRTF. This model combines the contributions of two individual spheres, approximated by the filters described above. As Figure 9a illustrates, the model switches between a torso-reflection sub-model when the source is outside the torso-shadow cone and a torso-shadow sub-model when the source is inside the torso-shadow cone. To determine when this transition occurs, let $\overline{s}$ be a unit vector pointing in the direction of the infinitely distanct source, and let $\overline{d}$ be a vector of length d from the center of the torso to the ear.[8] From Figure 10, grazing incidence occurs when $\overline{d}\cdot\overline{s} = -\sqrt{d^2-b^2}$. Thus,

$$\overline{s} \text{ is inside the torso-shadow cone if } \overline{d}\cdot\overline{s} < -\sqrt{d^2-b^2} \quad (6)$$

$$\overline{s} \text{ is outside the torso-shadow cone if } \overline{d}\cdot\overline{s} > -\sqrt{d^2-b^2}$$

We begin with the torso reflection sub-model, which is shown in Figure 9b. This sub-model assumes that the sound waves incident on the ear can be divided into two components that are superimposed at the ear: (a) a direct component that arrives from the direction of the source, and (b) a reflected component that arrives after being reflected from the torso. The direct component corresponds to the upper path in Figure 9b, and the reflected component corresponds to the lower path. For either path, the influence of the head is modeled by a head-shadow filter H_H and a propagation delay ΔT_H across the head. The reflected component experiences a further attenuation because of the torso reflection coefficient ρ and an additional propagation delay ΔT_R because of the greater length of the reflected path.

All that is needed to complete the specification of the part of the model for the direct component is to relate

[8] There are different torso-shadow cones for the two ears, and correspondingly different source directions for grazing incidence. When we speak of "the ear," it should be understood that the analysis has to be done separately for each ear.

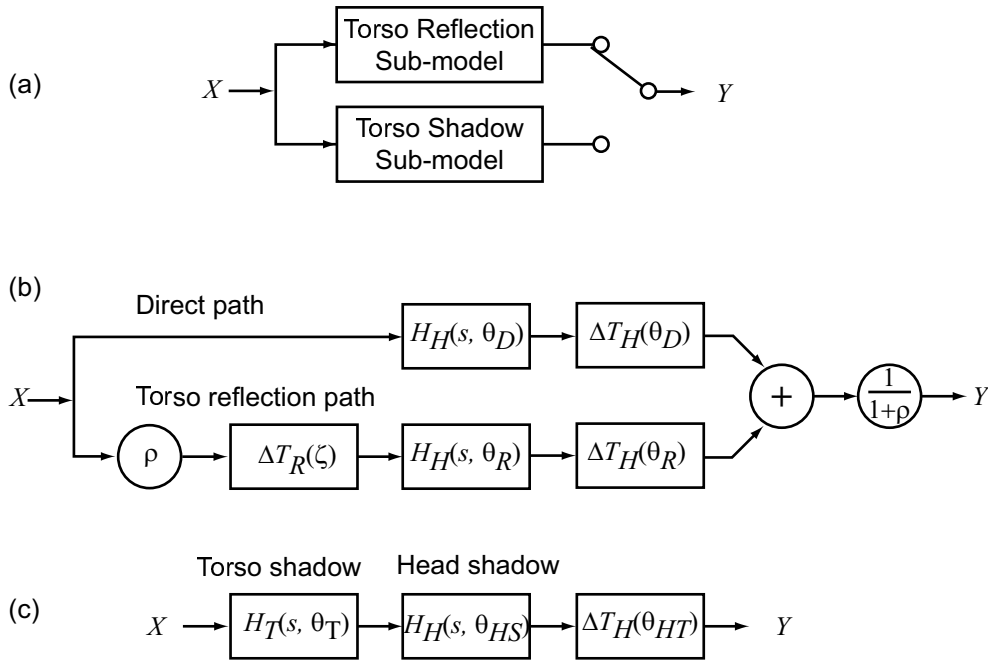

Fig. 9: The filter model for the snowman. (a) Major components. (b) The torso reflection sub-model. (c) The torso shadow sub-model.

the observation angle θ_D to the direction of the source and the location of the ear. Let $\overline{e}$ be a vector of length a from the center of the head to the ear. Because the source is assumed to be infinitely distant,

$$\theta_D = \cos^{-1} \frac{\overline{d} \cdot \overline{c}}{a} \quad . \tag{7}$$

Substitution of this value in Eq. 1 yields ΔT_H, and substitution in Eqs. 2 and 4 yields H_H.

The reflected component arrives at the head from a different direction than the direct component, resulting in a different observation angle θ_R. It also experiences an additional time delay ΔT_R. The analysis in Appendix A uses a ray-tracing argument to derive formulas for these two quantities.

This brings us to the reflection coefficient, ρ. It is clear from Figure 3 that the strength of the torso reflection falls off as the direction of the incoming wave approaches grazing incidence. In addition, reflection is a short-wavelength concept, and becomes meaningless at frequencies for which the wavelength is large compared to the torso. However, for simplicity, we assume that ρ is constant, and we subsequently scale the summed output by the scale factor $1/(1+\rho)$. The main reason for this final scaling is to preserve continuity in moving from outside to inside the torso-shadow cone. At grazing incidence, the angles θ_D and θ_R are equal, and the time delay ΔT_R is zero. Thus, only difference between the upper path and the lower path in Figure 9b is the scale factor ρ in the lower path, and the final scaling results in a transfer function that is equivalent to that of the upper path alone. This scaling also ensures that the sub-model has unity DC gain.[9]

7.2 The Torso-Shadow Sub-Model

We now turn to the torso-shadow sub-model shown in Figure 9c. When the source is in the torso-shadow cone, waves from the source must travel around the torso before reaching the head. The physical situation is rather complex, with wave components that take different paths around the torso to the ear traveling different distances

[9]Because of this constant-ρ approximation, at high frequencies and for overhead incidence, the peak values for the resulting comb-filter pattern are less than they should be by a factor of $1/(1+\rho)$. For $\rho = 0.3$, this introduces a 2.3-dB error. This error can be eliminated by using a reflection coefficient that varies with frequency and orientation, but at the cost of additional complexity.

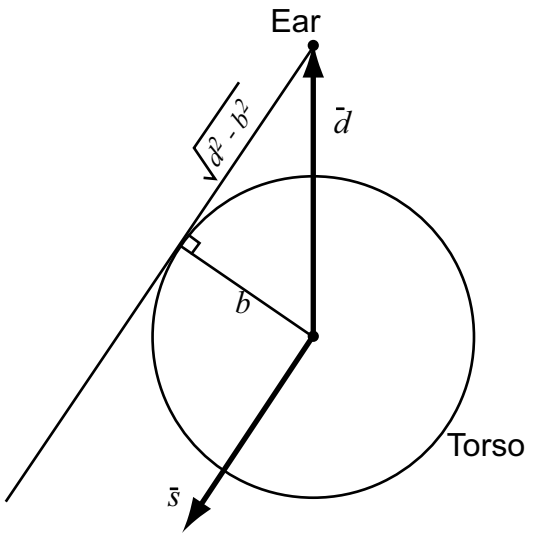

Fig. 10: The geometry for grazing incidence. When a ray from the ear to the source is tangent to the torso, the length of the projection of the vector $\overline{d}$ from the center of the torso to the ear is $\sqrt{d^2 - b^2}$, where $d = ||\overline{d}||$ and b is the radius of the torso sphere.

and arriving at different angles. We approximate this behavior by assuming that all the components are first shadowed by the torso acting as an isolated sphere with an observation point on the torso and an observation angle θ_T, and that these components all arrive at the ear at some effective head angle θ_H. This results in the simple cascade of a torso-shadow filter and a head-shadow filter shown in Figure 9c.

For θ_H, we make different assumptions for head shadowing than for head time delay. For head shadowing, we assume that the waves that travel around the torso leave at the point of tangency and travel on directly to the ear. As Figure 11 illustrates, this leads to a discontinuity in θ_{HS}, which switches from a smaller value when the source is on the ipsilateral side to a larger value when the source is on the contralateral side. Although discontinuities are not desirable, as was observed earlier, the snowman response does in fact change its behavior quite rapidly as the source crosses the torso bright spot, being relatively flat on the ipsilateral side and heavily shadowed on the contralateral side (see Figure 6). Formulas for computing θ_{HS} are derived in Appendix B.

For the time delay, a discontinuity in θ_H leads to a jump in apparent position, and this is not acceptable. We found that good sounding results were obtained by computing θ_{HT} as if the torso were not present. Thus, θ_{HT} is identical to θ_D in Eq. 7.

For the torso-shadow filter H_T we naturally use the torso radius for the sphere radius. The only remaining problem is to determine the observation angle θ_T for determining the torso shadow. Figure 11 suggests an obvious choice — the angle ζ between $\overline{s}$ and $\overline{d}$. However, this choice produces a discontinuity at the boundary of the torso shadow cone, with the response being fairly heavily shadowed inside the cone but almost flat outside. To reduce this discontinuity, we compute θ_T by interpolating between θ_{flat}, the angle for flat response as given in Eq. 5, and π, which is the value of ζ when the source switches from the ipsilateral to the contralateral side (see Figure 11). To be specific, let $\zeta_{\min}$ be the value of ζ at grazing incidence. From Figure 10,

$$\zeta_{\min} = \frac{\pi}{2} + \cos^{-1}\frac{b}{d} \quad . \tag{8}$$

Then interpolation yields

$$\theta_T = \frac{\pi(\zeta - \zeta_{\min}) - \theta_{\text{flat}}(\zeta - \pi)}{\pi - \zeta_{\min}} \quad . \tag{9}$$

8. THE FREQUENCY RESPONSE OF THE SNOWMAN FILTER MODEL

The snowman filter model has six parameters: the head radius a, torso radius b, neck height h, minimum response angle $\theta_{\min}$, minimum α value $\alpha_{\min}$, and reflection coefficient ρ. When the values used earlier ($a = 8.7$ cm, $b = 16.9$ cm, $h = 5.3$ cm, $\theta_{\min} = 150°$, $\alpha_{\min} = 0.1$ and $\rho = 0.3$) are substituted, the frequency response curves shown in Figure 12 are obtained. Comparing these curves to Figure 6, we see generally similar behavior — strong torso-reflection ripples on the ipsilateral side, and significant shadowing on the contralateral side. Notably missing in the model responses are the head and torso bright spots. In particular, where the theoretical solution shows flat or even rising responses between 250° and 270°, the model displays significant high-frequency rolloff. As was mentioned earlier, this difference was intentional. The torso bright spot is an interesting multipath phenomenon that accompanies a spherical torso, but it is not characteristic of human HRTFs. Unfortunately, human HRTF data at these low elevations is not available, and there is currently no basis for saying what the HRTF should be.[10]

[10]The measurement of human HRTFs at very low elevations

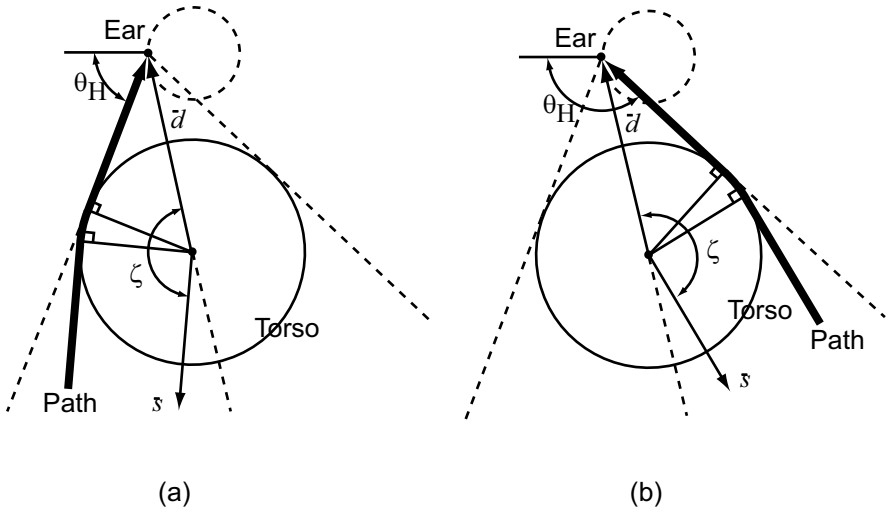

Fig. 11: The dominant paths from the source to the ear when the source is in the torso-shadow cone. The filter model assumes that all of the energy arrives at the observation angle θ_H. Note that θ_H is smaller in (a) when the source is on the ipsilateral side than it is in (b) when it is on the contralateral side.

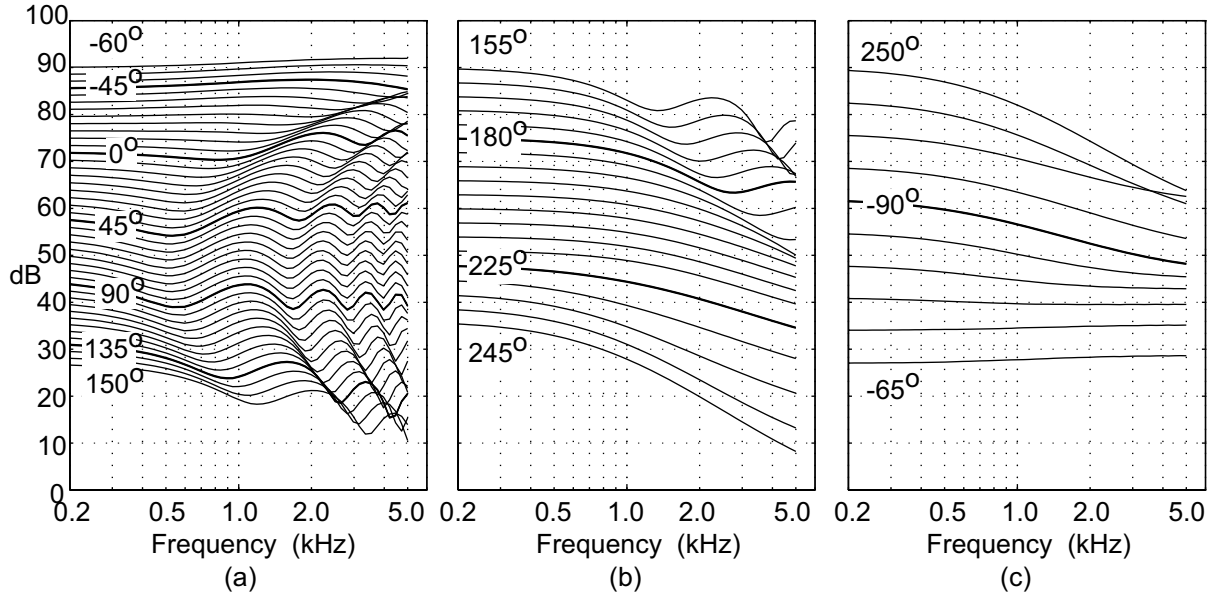

Fig. 12: The frequency response curves for the frontal-plane right-ear HRTF for the snowman filter model. In all cases, the responses approach 0 dB at low frequencies. The three panels can be compared directly to Figure 6, which gives the exact solution.

The resulting snowman filter model incorporates the major low-frequencies and low-elevation features that are revealed by the exact snowman model computation. It can be implemented as a simple combination of low-order IIR filters and time delays that can be realized quite efficiently. We now show how this can be done in a way that allows rapid motion of the listener or of the source.

9. REAL-TIME IMPLEMENTATION CONSIDERATIONS

Most systems for HRTF-based sound rendering convolve the source signals with relatively short FIR filters derived from the HRTFs. Because the filters have very limited duration, rapidly moving sources can be accommodated by rapidly switching between different direction dependent filters. By contrast, the proposed model includes IIR filters, and rapidly changing the parameters in IIR filters can introduce bad transient artifacts.

However, because the locations of the poles never change, it is straightforward to implement the model in a way that allows the model parameters to be changed arbitrarily rapidly without introducing such transient artifacts. To illustrate the general approach, consider a cascade of two first-order filters, such as the torso-shadow submodel shown in Figure 9c. Let $Y = HX$, where

$$H = H_1 H_2 = \frac{b_1 s + b_0}{s + a_1} \frac{c_1 s + c_0}{s + a_2} \quad .$$

Assuming that $a_1 \neq a_2$, we can expand H in partial fractions as

$$H = K_0 + \frac{K_1}{s + a_1} + \frac{K_2}{s + a_2} = K_0 + K_1 G_1 + K_2 G_2$$

where

$$G_1 = \frac{1}{s + a_1} \quad \text{and} \quad G_2 = \frac{1}{s + a_2} \quad .$$

Thus we can write Y as

$$Y = HX = K_0 X + K_1 X_1 + K_2 X_2$$

where

$$X_1 = G_1 X \quad \text{and} \quad X_2 = G_2 X \quad .$$

Although the gain coefficients K_0, K_1 and K_2 vary with the position of the source relative to the head, the poles in the low-pass filters G_1 and G_2 depend only on the dimensions of the listener, and for a given listener are fixed. Thus, two simple recursive filters with one memory tap can be used to compute X_1 and X_2, which are then combined with variable gains to compute Y. This can be done very rapidly, with no artifacts, no need to do any interpolation in HRTF tables, and an ability to adapt the HRTF to the individual listener.

10. USING THE MODEL TO COMPENSATE EXPERIMENTALLY MEASURED HRTFS

The standard way to obtain human HRTFs is through acoustic measurement [14, 15]. To obtain a sound source that is conveniently movable and approximates a point source, experimenters usually use relatively small loudspeakers. Although free-field compensation can partially correct for the loss of low-frequencies, the poor signal-to-noise ratio at low signal levels makes the HRTF measurements unreliable at low frequencies. For the measurements made in our laboratory, we consider 500 Hz to be the minimum frequency for reliable measurements [8].

Most of the interest in HRTF measurements is for investigating the effects of the pinna. Because pinna effects do not even begin to appear until about 3 kHz, for pinna studies this absence of low-frequency information is of no concern. However, when HRTFs are used to render spatial sounds for entertainment purposes, the weakness of the bass response in experimental data presents a serious problem.

One approach to solving this problem is to average-out the low-frequency response through the use of diffuse-field-compensated HRTFs [16]. Another approach is to substitute the HRTF of a theoretical model in place of the measured HRTF at frequencies where the measurements are suspect. Such a method can be used with any model, and we have employed it successfully with both the spherical-head model and the snowman filter model.

The primary question is how to combine the HRTF H_D for experimental data with the HRTF H_M for the model. We have found that an effective approach is to use the phase response of the model for all frequencies, and to "cross-fade" the dB magnitude responses at low frequencies. The rationale behind this approach is that the phase response of the model is more reliable than the measured phase response at low frequencies, and the human auditory system is not sensitive to phase at high-frequencies. Let $A_D(f)$ be $20 \log_{10} |H_D|$ at frequency f, $A_M(f)$ be $20 \log_{10} |H_M|$, and let f_1 and f_2 be two transi-

is problematic. In particular, the exact results are highly posture dependent, and it is not clear that measurements for one particular posture are particularly meaningful. However, it is clear from the physics of wave propagation that the torso will block frequencies whose wavelengths are comparable to or less than torso dimensions, which is what happens with the filter model.

tion frequencies.[11] Then the magnitude $A(f)$ (in dB) for the combined transfer function $H(f)$ is computed from

$$A(f) = \begin{cases} A_M(f) & 0 \le f < f_1 \\ \dfrac{f_2 - f}{f_2 - f_1} A_M(f) + \dfrac{f - f_1}{f_2 - f_1} A_D(f) & f_1 \le f < f_2 \\ A_D(f) & f_2 \le f \end{cases} \tag{10}$$

When the phase of the model is associated with this magnitude and the results are inverse transformed to obtain an impulse response, the waveform can have a very unusual visual appearance, but if the response is limited to a few milliseconds this does seem to be audible. Alternatively, if the interaural time delay is estimated and subsequently reinserted, a shorter and more normal looking impulse response can be obtained by the standard technique of minimum-phase reconstruction [9].

11. DISCUSSION AND CONCLUSIONS

The snowman model is a simple and natural generalization of the spherical-head model that provides insight and useful approximations to the HRTF, principally at low frequencies. By including the torso reflection and torso shadow, the model adds useful elevation cues. A full HRTF model would also provide pinna cues, which are very important when the source contains significant energy above 3 kHz. However, even without pinna cues, the head-and-torso model provides limited but useful control over the apparent elevation of the source, particularly when the source is away from the median plane. For sources above the horizontal plane, torso reflections provide useful elevation cues. For sources at low elevations, the low-frequency effects of torso shadow — effects that occur below 700 Hz and that have often been ignored in previous HRTF studies — are important for imparting a sense that the source is located down below the listener. Such low-elevation cues are also important for proper rendering of floor reflections from nearby elevated sources. The simple snowman head-and-torso model reveals the nature and magnitude of such elevation cues.

However, models based on spheres exhibit bright spots that are at best weak phenomena in human HRTFs. Thus, the proposed filter-model approximation exploits the insight that sphere models provide, but does not attempt to replicate the bright spots seen in the snowman HRTF. The filter model that was developed has several advantages: (a) it is very efficient, (b) it preserves the important azimuth and elevation cues of the snowman model, (c) it allows for arbitrarily high angular resolution, (d) it can respond to rapidly moving sources, (e) it can be adapted to individual listeners, and (f) it provides a good sounding low-frequency response. Because the model lacks pinna features, used alone it provides only a primitive HRTF. However, as a component of a more elaborate structural HRTF model [17], it contributes localization cues and provides a fundamental approach to the proper design of the low-frequency portion of the HRTF. Such a model can also be used to correct for low-frequency deficiencies in experimentally measured HRTF data. By providing both insight and simple implementation techniques, the snowman filter model is a useful addition to the methods for generating spatial sound.

[11] Typical values for these frequencies might be $f_1 = 250$ Hz and $f_2 = 1000$ Hz. In our experience, the choice of these frequencies is not critical.

ACKNOWLEDGMENTS

We are indebted to Nail Gumerov and Ramani Duraiswami for their development and sharing of the multipole reexpansion programs that made this investigation possible. This work was supported by the National Science Foundation under grants IIS-00-97256 and ITR-00-86075. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the view of the National Science Foundation.

APPENDIX A — RAY-TRACING ANALYSIS OF TORSO REFLECTIONS

In this Appendix, a ray-tracing argument is used to determine the additional time delay incurred by sound traveling along the reflected path and the angle at which the torso reflection arrives at the ear. Let $\overline{d}$ be a vector of length d from the center of the torso to the ear, $\overline{s}$ be a unit vector pointing to the infinitely distant source, and $\overline{b}$ be a vector of length b from the center of the torso to the point of reflection. Given $\overline{d}$ and $\overline{s}$, the basic problem is to find $\overline{b}$.

We begin by observing that these three vectors must lie in the same plane, and this plane is defined by $\overline{d}$ and $\overline{s}$ (see Figure 13a).[12] In this plane, we can specify $\overline{s}$ by a single coordinate — such as the angle ζ from $\overline{d}$ to $\overline{s}$ or the complementary angle ϵ given by[13]

$$\epsilon = \frac{\pi}{2} - \zeta = \frac{\pi}{2} - \cos^{-1} \frac{\overline{d} \cdot \overline{s}}{d} \quad . \tag{A1}$$

[12] In the degenerate case where $\overline{s}$ and $\overline{d}$ are colinear, $\overline{b}$ is the projection of $\overline{d}$ onto the sphere.

[13] Because the vector $\overline{d}$ is close to being vertical, the angle ϵ is close to the elevation angle for the source, and we refer to ϵ loosely as the "elevation" angle.

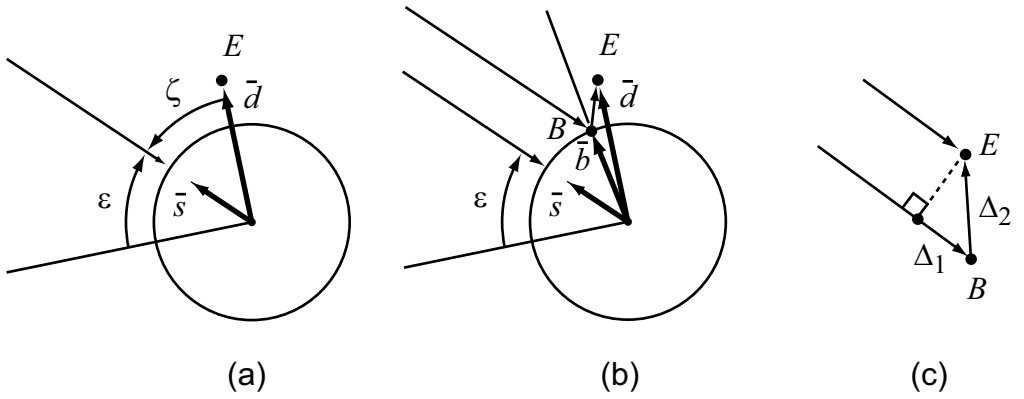

Fig. 13: Normal view of the plane defined by the vector $\overline{d}$ from the center of the torso to the ear and the vector $\overline{s}$ pointing to the source. (a) The "elevation" angle ϵ. (b) The goal is to find the vector $\overline{b}$ to the point of reflection. (c) The reflected ray travels an additional distance that is the sum of Δ_1 and Δ_2.

The vector $\overline{b}$ to the point of reflection lies somewhere on the arc between the projection of $\overline{s}$ onto the torso and the projection of $\overline{d}$ onto the torso (see Figure 13b). Once $\overline{b}$ is determined, it is straightforward to determine the vector from the point of reflection to the ear, and the distance $\Delta = \Delta_1 + \Delta_2$, the difference between the path length of the ray going directly to the ear and the path length of the reflected ray (see Figure 13c).

Let α be the angle between $\overline{b}$ and $\overline{d}$. Because the length of $\overline{b}$ is the known radius b of the torso, $\overline{b}$ is completely determined by the angle α (see Figure 14). The problem is to find α, given ϵ. We introduce the angle β between the reflected ray and $\overline{d}$. From Figure 14,

$$\tan\beta = \frac{b\sin\alpha}{d - b\cos\alpha} = \frac{\sin\alpha}{A - \cos\alpha} \qquad (A2)$$

where

$$A = \frac{d}{b} \quad . \qquad (A3)$$

From Figure 14, we see that the angle of reflection ψ is the sum of α and β:

$$\psi = \alpha + \beta \quad . \qquad (A4)$$

We now exploit the fact that at the point of reflection the angle of incidence must equal the angle of reflection, and thus $\epsilon + 2\psi + (\pi/2 - \beta) = \pi$. It follows that

$$\epsilon = \frac{\pi}{2} + \beta - 2\psi = \frac{\pi}{2} + \beta - 2(\alpha + \beta)$$

$$= \frac{\pi}{2} - 2\alpha - \tan^{-1}\left[\frac{\sin\alpha}{A - \cos\alpha}\right] \quad . \qquad (A5)$$

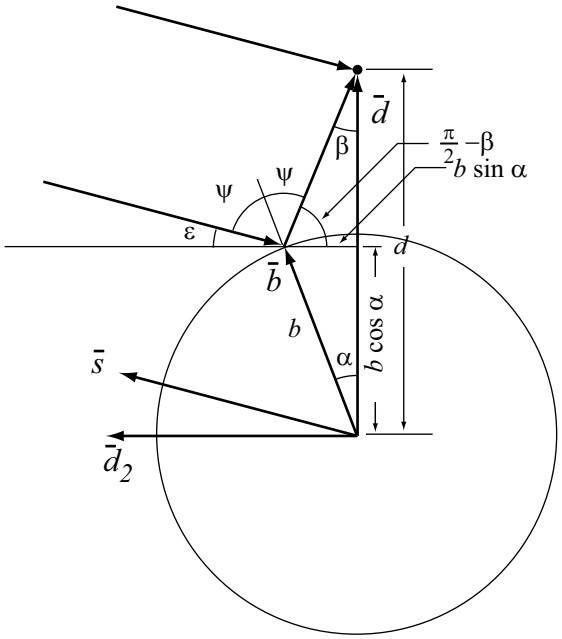

Fig. 14: The vector $\overline{b}$ to the point of reflection, which lies in the plane defined by the vector $\overline{s}$ to the source and the vector $\overline{d}$ to the ear, is determined once the angle α is known. For a given "elevation" angle ϵ, the condition that the angle of incidence ψ equals the angle of reflection uniquely determines α.

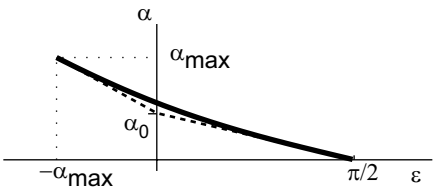

Fig. 15: A piecewise linear approximation to the relation between the "elevation" angle ϵ and the angle α is determined by the two parameters α_0 and $\alpha_{\max}$.

This key equation links α, ϵ, and the ratio $A = d/b$. Although we cannot invert it analytically to solve for α as an explicit function of ϵ and A, the equation is well behaved and is easily solved numerically.

A graph of the solution for a particular value of A is shown in Figure 15, which also shows a piecewise-linear approximate solution. This approximate solution is valuable. It can be used to obtain a rough solution quickly, or to obtain an excellent starting point for iterative methods. To determine this approximate solution, we begin by noting that ϵ approaches $\pi/2$ as α approaches 0. For small α, we can use small-angle approximations to obtain the linear approximation $\epsilon \approx \pi/2 - 2\alpha - \alpha/(A-1)$, or $\alpha \approx ((A-1)/(2A-1))(\pi/2-\epsilon)$. This approximation is reasonably good for $0 \le \epsilon \le \pi/2$ (see Figure 15). For $\epsilon = 0$ it yields

$$\alpha_0 = \frac{A-1}{2A-1}\frac{\pi}{2} \quad , \tag{A6}$$

which is somewhat smaller than but fairly close to the true value. When ϵ is negative, α continues to grow until it reaches a maximum value $\alpha_{\max}$ when the reflected ray is tangent to the sphere. It is clear from Figures 14 and 16 that at that point $\psi = \pi/2$ and $\epsilon = -\alpha_{\max}$. It follows that $\cos\alpha_{\max} = b/d = 1/A$, so that

$$\alpha_{\max} = \cos^{-1}\frac{1}{A} \quad . \tag{A7}$$

This leads to the following linear approximation for α:

$$\alpha \approx \begin{cases} \alpha_0 - \left(1 - \dfrac{\alpha_0}{\alpha_{\max}}\right)\epsilon & \text{if } -\alpha_{\max} \le \epsilon \le 0 \\ \alpha_0\left(1 - \dfrac{\epsilon}{\pi/2}\right) & \text{if } 0 \le \epsilon \le \pi/2. \end{cases} \tag{A8}$$

We now obtain the time delay of the torso reflection from the "elevation" ϵ. Two possible cases are illustrated in Figure 17. In the steep elevation case (Figure 17a), the wave arrives at the ear before it arrives at the torso, and the difference in path lengths $\Delta$is the sum $f + f\cos 2\psi$, where f is the distance from the point of reflection to the ear. In the shallow elevation case (Figure 17b), the wave arrives at the ear after it arrives at the torso, and the difference in path length is the difference $f - f\sin(\beta-\epsilon)$. However, it turns out that both cases are covered by the formula

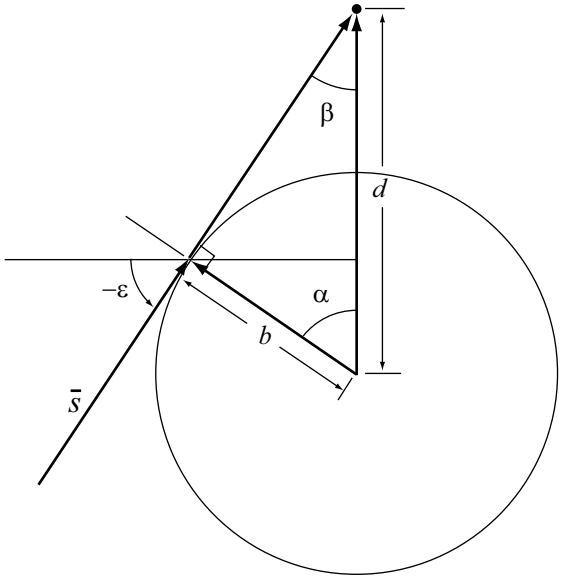

Fig. 16: The limiting case of grazing incidence.

$$\Delta = f\,(1 + \cos 2\psi) \quad . \tag{A9}$$

The reason is that $\epsilon + 2\psi - \beta = \pi/2$ (see Figure 17b), and thus $\sin(\beta-\epsilon) = \sin(2\psi - \pi/2) = -\cos(2\psi)$. From Figure 17 and the law of cosines we have $f^2 = b^2 + d^2 - 2bd\cos\alpha$, which determines f and hence determines Δ. The time delay ΔT is found by dividing Δ by c, the speed of sound.

To summarize, given the torso radius b, the vector $\overline{d}$ to the ear, and the vector $\overline{s}$ to the source, we can find ΔT by the following sequence of calculations:

$$\begin{aligned} A &= d/b \\ \epsilon &= \frac{\pi}{2} - \cos^{-1}\frac{\overline{d}\cdot\overline{s}}{d} \\ \alpha : \text{solve} \quad \epsilon &= \frac{\pi}{2} - 2\alpha - \tan^{-1}\left[\frac{\sin\alpha}{A - \cos\alpha}\right] \\ f &= \sqrt{b^2 + d^2 - 2bd\cos\alpha} \end{aligned}$$

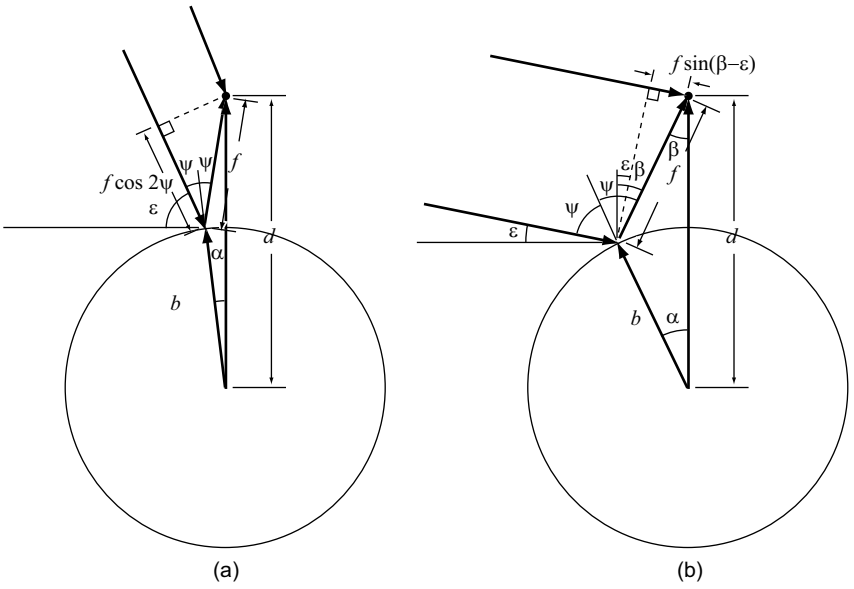

Fig. 17: The path length difference. (a) $\psi < 45°$ (b) $\psi > 45°$

$$\begin{aligned} \beta &= \tan^{-1}\frac{\sin\alpha}{A-\cos\alpha} \\ \psi &= \alpha+\beta \\ \Delta T &= \frac{f}{c}(1+\cos 2\psi) \quad . \end{aligned}$$

To complete the analysis, we need to derive an expression for a unit vector $\overline{r}$ from $\overline{d}$ to $\overline{b}$, so that $\overline{r}$ points in the direction of the incoming reflected wave. We do this by writing $\overline{b}$ as a linear combination of $\overline{d}$ and $\overline{d}_2$, where $\overline{d}_2$ is a vector in the plane of $\overline{d}$ and $\overline{s}$ that is orthogonal to $\overline{d}$ (see Figure 14). It is easy to verify that if $\overline{s}$ and $\overline{d}$ are not colinear, the following vector satisfies these conditions:

$$\overline{d}_2 = d^2\,\overline{s} - (\overline{d}\cdot\overline{s})\,\overline{d} \quad . \qquad (A10)$$

From Figure 14, we see that

$$\overline{b} = b\cos\alpha\,\frac{\overline{d}}{d} + b\sin\alpha\,\frac{\overline{d}_2}{||\overline{d}_2||} \quad . \qquad (A11)$$

Thus, the desired vector $\overline{r}$ is given by

$$\overline{r} = \frac{\overline{b}-\overline{d}}{||\overline{b}-\overline{d}||} \quad . \qquad (A12)$$

The final step is to use $\overline{r}$ in place of $\overline{s}$ in Eq. 7 to compute the observation angle θ_R for the reflected component.

APPENDIX B — THE OBSERVATION ANGLE FOR TANGENT INCIDENCE

In this Appendix, explicit formulas are obtained for the vector $\overline{r}$ from the ear to the point of tangency when the vector $\overline{s}$ to the source is inside the torso shadow zone. As Figure 11 illustrates, two solutions are obtained, one for the case where the source is on the ipsilateral side of the torso and one for the contralateral side. The approach is a continuation of the analysis in Appendix A, and the same definitions and notation are used. However, now the basic goal is to find the vector $\overline{b}$ from the center of the torso to the point of tangent incidence. Once this vector is determined, the vector $\overline{r}$ is obtained at once from Eq. A12.

As in Eq. A11, we write $\overline{b}$ as a linear combination of $\overline{d}$, the vector to the ear, and $\overline{d}_2$, the vector orthogonal to $\overline{d}$ obtained by Eq. A10:

$$\overline{b} = w_1\overline{d} + w_2\overline{d}_2 \quad . \qquad (B1)$$

We determine the coefficients w_1 and w_2 by imposing the conditions that (a) the vector $\overline{b}-\overline{d}$ must be orthogonal to $\overline{b}$ (tangent incidence), and (b) the length of $\overline{b}$ must be the radius b of the torso.[14] From the requirement for

[14]With care, it is also possible to use Eq. A11 for $\overline{b}$ and require that α be at the maximum value $\alpha_{\max}$ given by Eq. A7. Indeed, the value obtained for w_1 in Eq. B2 is exactly $\cos^2\alpha_{\max}$, just as one would obtain by substituting $\alpha=\alpha_{\max}$ in Eq. A11. However, this leads to logical confusion because the derivation of Eq. A11 assumed that $\overline{s}$ was outside of or on the torso shadow cone, and now we have $\overline{s}$ inside the torso-shadow cone.

tangent incidence, we have $(\overline{d}-\overline{b})\cdot\overline{b}=0$ or $\overline{d}\cdot\overline{b}=b^2$. But because $\overline{b}$ and $\overline{d}_2$ are orthogonal, Eq. B1 yields

$$\overline{d}\cdot\overline{b}=\overline{d}\cdot(w_1\overline{d}+w_2\overline{d}_2)=w_1d^2=b^2$$

so that

$$w_1=\left(\frac{b}{d}\right)^2 \quad . \tag{B2}$$

From Eq. B1, the requirement that $||\overline{b}||=b$, and the orthogonality of $\overline{d}$ and $\overline{d}_2$, we have

$$||\overline{b}||^2=w_1^2d^2+w_2^2||\overline{d}_2||^2=b^2=w_1d^2$$

so that

$$w_2=\pm\sqrt{w_1(1-w_1)}\,\frac{d}{||\overline{d}_2||} \quad . \tag{B3}$$

Because Eq. A10 for the orthogonal vector $\overline{d}_2$ is valid regardless of whether $\overline{d}$ is inside or outside of the torso-shadow zone, we have $||\overline{d}_2||^2=||d^2\overline{s}-(\overline{d}\cdot\overline{s})\overline{d}||^2=d^4-2d^2(\overline{d}\cdot\overline{s})^2+(\overline{d}\cdot\overline{s})^2d^2=d^2(d^2-(\overline{d}\cdot\overline{s})^2)$. Substituting this in Eq. B3 leads to

$$w_2=\pm\sqrt{\frac{w_1(1-w_1)}{d^2-(\overline{d}\cdot\overline{s})^2}} \quad . \tag{B4}$$

Thus, we obtain two solutions for $\overline{b}$. Recall that we want $\overline{b}$ to be on the ipsilateral side when $\overline{s}$ is on the ipsilateral side, and on the contralateral side when $\overline{s}$ is on the contralateral side. From Eq. A10, it is clear that $\overline{d}_2$ always points to that side that $\overline{s}$ is on. Thus, we always choose the positive sign, and the direction of $\overline{s}$ will automatically determine the proper solution.[15] The final step is to use Eq. A12 to compute the desired unit vector $\overline{r}$ from $\overline{d}$ to $\overline{b}$, which is then used in place of $\overline{s}$ in Eq. 7 to compute the observation angle θ_{HS} for the head.

[15] As $\overline{s}$ comes into alignment with $\overline{d}$ at the torso bright spot, $d^2-(\overline{d}\cdot\overline{s})^2$ becomes very small, and numerical problems occur at the crossover from one solution to the other. In that case we merely assume that the wave is coming from the direction of $\overline{s}$.

REFERENCES

[1] Blauert, J. *Spatial Hearing* (revised edition) (MIT Press, Cambridge, MA, 1997).

[2] Duda, R. O. and Martens, W. L., "Range dependence of the response of a spherical head model," *J. Acoust. Soc. Am.*, Vol. 104, No. 5, pp. 3048-3058 (1998).

[3] Algazi, V. R., Avendano, C., and Duda, R. O., "Elevation localization and head-related transfer function analysis at low frequencies," *J. Acoust. Soc. Am.*, Vol. 109, No. 3, pp. 1110-1122 (March 2001).

[4] Algazi, V. R., Avendano, C., and Duda, R. O., "Estimation of a spherical-head model from anthropometry," *J. Aud. Eng. Soc.*, Vol. 49, No. 6, pp. 472-478 (2001).

[5] 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," Technical Report CIL-2002-1, CIPIC Interface Laboratory, University of California, Davis, CA 95616 (April, 2002). (Available from http://interface.cipic.ucdavis.edu/CIL_html/CIL_publications.htm)

[6] Gumerov, N., Duraiswami, R., and Tang, Z., "Numerical study of the influence of the torso on the HRTF," *ICASSP'02* (Proc. IEEE Int. Conf. Acoustics Speech and Signal Processing) (2002).

[7] Minnaar, P., Plogsties, J., Olesen, S., Christensen, F., and Møller, H., "The interaural time difference in binaural synthesis," Preprint 5133, *108th AES Convention*, 20 p., Paris, France (Feb. 2000).

[8] Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C., "The CIPIC HRTF database," *WASPAA'01* (Proc. 2001 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics), New Paltz, New York, pp. 99-102 (2001).

[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., Vol. 91, pp. 1637–1647 (1992).

[10] Huopaniemi, J. and Karjalainen, M., "Review of digital filter design and implementation methods for 3-D sound," Preprint 4461, 102nd Convention Audio Engineering Society, Munich, Germany, (1997).

[11] Larcher, V., Jot, J-M, Guyard, J. and Warusfel, O., "Study and comparison of efficient methods of 3D audio spatialization based on linear decomposition of HRTF data," 108th AES Convention, Paris, France (2000).

[12] Brown, C. P. and Duda, R. O., "A structural model for binaural sound synthesis," *IEEE Trans. Speech Aud. Proc.*, Vol. 6, No. 5, pp. 476-488 (1998).

[13] Haneda, Y., Makino, S. and Kaneda, Y., "Common acoustical poles independent of sound directions and modeling of head-related transfer functions," *J. Acoust. Soc. Japan* # "(E)", Vol. 15, No. 4, pp. 277–279 (1994).

[14] Mehrgardt, S. and Mellert, V., "Transformation characteristics of the external human ear," *J. Acoust. Soc. Am.*, Vol. 61, pp. 1567-1576 (1977).

[15] Wightman, F. L., Kistler, D. J. and Perkins, M. E., "A new approach to the study of human sound localization," in Yost, W. A. and Gourevitch, G., Eds., *Directional Hearing*, pp. 26-48, Springer Verlag: NY (1987).

[16] Jot, J-M, and Larcher, V., "Approaches to binaural synthesis," 105th Convention of the Audio Engineering Society, San Francisco, CA (September 1998).

[17] Algazi, V. R., Duda, R. O., Morrison, R. P., and Thompson, D. M., "Structural composition and decomposition of HRTFs," *Proc. WASPAA '01* (2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics), pp. 103-106 (New Paltz, NY, October 2001).