Title
Elevation Localization and Head-Related Transfer Function Analysis at Low Frequencies

Journal
J. Acoust. Soc. Am., 109

Authors
Algazi, Ralph
Avendano, Carlos
Duda, Richard O

Publication Date
2001

Peer reviewed

Elevation localization and head-related transfer function analysis at low frequencies

V. Ralph Algazi
CIPIC, Center for Image Processing and Integrated Computing, University of California, Davis, California 95616

Carlos Avendano
Creative Advanced Technology Center, 1600 Green Hills Road, Scotts Valley, California 95067

Richard O. Duda
Department of Electrical Engineering, San Jose State University, San Jose, California 95192

(Received 22 May 2000; revised 9 August 2000; accepted 20 December 2000)

Monaural spectral features due to pinna diffraction are the primary cues for elevation. Because these features appear above 3 kHz where the wavelength becomes comparable to pinna size, it is generally believed that accurate elevation estimation requires wideband sources. However, psychoacoustic tests show that subjects can estimate elevation for low-frequency sources. In the experiments reported, random noise bursts low-pass filtered to 3 kHz were processed with individualized head-related transfer functions (HRTFs), and six subjects were asked to report the elevation angle around four cones of confusion. The accuracy in estimating elevation was degraded when compared to a baseline test with wideband stimuli. The reduction in performance was a function of azimuth and was highest in the median plane. However, when the source was located away from the median plane, subjects were able to estimate elevation, often with surprisingly good accuracy. Analysis of the HRTFs reveals the existence of elevation-dependent features at low frequencies. The physical origin of the low-frequency features is attributed primarily to head diffraction and torso reflections. It is shown that simple geometrical approximations and models of the head and torso explain these low-frequency features and the corresponding elevations cues. [DOI: 10.1121/1.1349185]

PACS numbers: 43.66.Qp, 43.66.Pn [DWG]

I. INTRODUCTION

It is well established that the interaural time difference (ITD) and the interaural level difference (ILD) provide the primary cues for the horizontal localization of a sound source, whereas the monaural spectral modifications introduced by the pinna provide the primary cues for vertical localization (Middlebrooks and Green, 1991; Carlile, 1996; Blauert, 1997; Wightman and Kistler, 1997). Pinna effects start to appear at frequencies around 3 kHz, where the wavelength becomes comparable to the pinna size, with the so-called ''pinna notch'' appearing within the octave from 6 to 12 kHz (Shaw, 1997). This supports the general belief that the source must have substantial high-frequency energy over a fairly wide band for accurate judgment of elevation (Roffler and Butler, 1967; Gardner and Gardner, 1973; Butler, 1986; Asano, Suzuki, and Sone, 1990).

The role of the torso in localization is less well understood. The fact that the torso disturbs incident sound waves at low frequencies has been recognized for a long time (Hanson, 1944; Kuhn and Guernsey, 1983). However, the effects of the torso are relatively weak, and experiments to establish the perceptual importance of low-frequency cues have produced mixed results. For example, Theile and Spikofski (1982) concluded from their experiments that the torso does not provide significant cues for front/back discrimination. However, while agreeing that high-frequency spectral cues are needed for front/back discrimination, Asano *et al.* (1990) observed that front/back discrimination is significantly improved when the subjects are provided with the correct low-frequency spectrum.

The effect of the torso on vertical localization in the median plane was first systematically investigated by Gardner (1973), who observed that—although the subjective sense of source location was greatly diminished when high frequencies were removed—it was possible for some subjects to localize sounds from loudspeakers located in the anterior median plane, despite the fact that the source had no spectral energy above 4 kHz. Gardner also measured the head-related transfer function (HRTF) of a mannequin, both with and without pinna occlusion and with and without a torso. By comparing the change in the response at $+18°$ elevation to that at $-18°$ elevation, he concluded that the pinna had no influence below 3.5 kHz, but that the torso introduced important ''clues of a secondary nature'' between 0.7 and 3.5 kHz. However, he cautioned that the mere presence of elevation-dependent low-frequency spectral features does not mean that they can be exploited by the auditory system. Searle *et al.* (1976) identified six localization cues in their statistical model of human sound localization, and used Gardner's data to estimate the variance due to the torso reflection or ''shoulder bounce.'' They concluded that the shoulder bounce provided by far the weakest elevation cue.

Kuhn (1987) used a KEMAR mannequin with and without pinnae and torso in a study of the behavior of the HRTF

for all elevations in the median plane. He showed that median plane directivity is governed by specular reflection from the torso at frequencies below 2 kHz and by complex pinna phenomena for frequencies above 4 kHz. However, the question of whether or not the low-frequency features provided effective elevation cues was not addressed.[1]

Going outside the median plane, Genuit and Platte (1981) showed that the torso introduced both direction- and distance-dependent effects on the HRTF that are limited to the spectral range below 3 kHz, and Genuit (1984) subsequently included separate torso and shoulder submodels in his structural HRTF model. Brown and Duda (1998) observed torso reflections in head-related impulse response (HRIR) data, and also included a ''shoulder echo'' in their structural HRTF model. However, that component was omitted during their formal tests of the model because informal listening experiments had indicated that the simulated torso reflections did not have a significant effect on perceived elevation in the median plane.

This paper reports on psychoacoustic experiments with individualized HRTFs that show that there are significant elevation cues for sources having little high-frequency energy, but the source must be away from the median plane. Some of the experiments used measured HRTFs, and others used a simplified low-frequency HRTF model. The methods used for the psychoacoustic experiments are described in Sec. II. The experimental results obtained with measured HRTFs are reported, analyzed, and discussed in Sec. III. Section IV presents an analysis of the low-frequency characteristics of HRTF that demonstrates that the pinnae do not contribute to the HRTF at frequencies below 3 kHz. Simple geometric models of the head and torso of each subject are then developed and analyzed to establish that the head and torso are the determinant contributors to the HRTFs at low frequencies. Finally in Sec. V, the results of psychoacoustic experiments with synthetic approximations and simple models of the head and torso are reported that confirm the contributions of head and torso to the perceived elevation.

II. METHODS

A. HRTF measurements

The HRTFs employed in this study were measured using the blocked-ear-canal technique (Møller, 1992; Algazi, Avendano, and Thompson, 1999). The probe tubes of two Etymōtic Research ER-7C microphones were attached to plastic ear plugs, which were then inserted into the subject's ear canals. The subjects were seated and, to minimize head movements, were asked to control their head position by viewing their reflection in a mirror; however, they were not otherwise physically constrained. The impulse responses were obtained using Golay codes (Crystal River Engineering Snapshot™ system), played through Bose Acoustimass™ Cube speakers. The speakers were mounted on a 1-m-radius hoop that was rotated about the subject's interaural axis. The sampling rate for the measurements was 44.1 kHz. To remove most room reflections, the resulting impulse responses were windowed and truncated to a duration of 4.5 ms, and were equalized to compensate for the loudspeaker and microphone transfer functions.

The geometry of the HRTF measurement apparatus leads naturally to use of the interaural–polar spherical coordinate system shown in Fig. 1. The origin of this spherical coordinate system is at the interaural midpoint, which is usually somewhat below and behind the center of the head. The azimuth angle θ is measured between the median plane and a ray from the origin to the source. An azimuth angle of $+90°$ corresponds to the right side of the subject, and $-90°$ to the left, with $\theta=0°$ defining the median plane. The elevation angle ϕ is the polar rotation angle, with $\phi=0°$ defining the anterior horizontal half-plane. The elevation sequence $-90°$, $0°$, $90°$, $180°$, and $270°$ corresponds, respectively, to locations below, in front of, above, in back of, and below the subject.[2]

The HRTFs were measured at 1250 locations in space, with elevation increments of $\Delta\phi=5.625°$ for a range $-45°\leqslant\phi\leqslant231°$ and at 25 different azimuth angles with a $5°$ spacing in the front, increasing towards the interaural poles (Algazi *et al.*, 1999).

To a first degree of approximation, in this coordinate system the ITD depends on azimuth alone (Searle *et al.*, 1976; Wightman and Kistler, 1997). A surface of constant interaural–polar azimuth is often called a ''cone of confusion.'' Thus, in principle, knowledge of the ITD would allow one to estimate the azimuth, and hence to constrain the location of the source to a particular cone of confusion. For a constant range, the source moves around a ''circle of confusion'' which corresponds to the trajectory described by one of the loudspeakers as the hoop rotates.

B. Subjects

Six subjects were tested, four males and two females ranging in age from 20 to 42 years. None of the subjects was related to the research and all had normal hearing. All subjects were students or staff members at UC Davis, and had no previous experience with listening tests.

C. Experiments

The experiments involved listening to simulated or virtual auditory sources through headphones. The headphone stimuli were produced by convolving a test signal with the left and right impulse responses for each position tested, and the subjects were asked to report the perceived elevation.

Localization accuracy was measured on the left side of the subject in 16 different situations, one for each of the possible combinations of the following three factors:

Azimuth angle θ: $0°$, $-25°$, $-45°$, $-65°$;
Source location: front, back;
Source bandwidth: 22 kHz, 3 kHz.

The aim of the experiments was to compare the accuracy of the elevations reported by the subjects for full-bandwidth sound sources with that for low-pass-filtered, limited-bandwidth sources. In an ''absolute-judgment'' approach, the subject listened to a presentation of a test signal and used

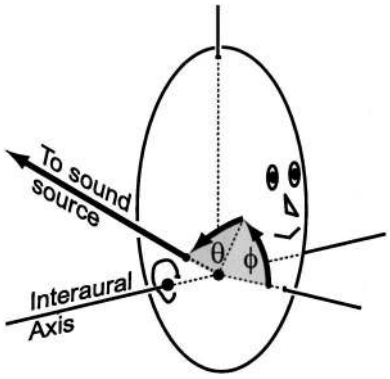

FIG. 1. The interaural–polar coordinate system. A surface of constant interaural azimuth θ is a cone of confusion, while a surface of constant interaural elevation ϕ is a half-plane through the interaural axis.

a graphical interface to select any point on a circle that best corresponded to the perceived elevation. To familiarize subjects with the procedure, test sessions were preceded by a brief description of the coordinate system and a presentation of a subset of the stimuli. Subjects were asked to think of the circle as a projection of the circle of confusion onto a plane. To visualize this mapping, circles of confusion were constructed on the surface of a three-dimensional image of a sphere, and subjects could immediately relate the circles to the trajectories of the loudspeakers at the time when their HRTFs were measured. To provide familiarization with the procedure, each subject was allowed a brief time period in which she or he could follow a marker on the circle and hear the corresponding stimulus. Front and back locations were tested separately and the subject always knew which condition prevailed.[3]

Each of the 16 situations was tested separately. For example, a particular test might be for a low-pass-filtered source at $-45°$ azimuth located in the front. For each test, one of 12 elevation angles was randomly selected, subject to the constraint that each angle would eventually be repeated 10 times. This gave a total of $n=120$ responses per test situation. When the source was in the front, the elevation angles ranged from $-45°$ to 78.75° in 11.25° steps. Subjects were allowed to respond with an elevation anywhere between $-90°$ and 90°. The mirror image locations were used when the source was in back: 225° to 101.25° in $-11.25°$ steps, and subjects could respond anywhere between 90° and 270°. Each test situation required approximately 15 min to complete, with all 16 situations tested in about 4 h. To reduce fatigue, experiments were split into sessions of 2 h each, performed on different days.

D. Stimuli

The 22-kHz test signal was a sequence of two Gaussian noise bursts, sampled at 44.1 kHz and independently generated on each presentation. Each noise burst had a duration of 500 ms, with a 250-ms silent period between bursts. In addition, to increase the effective number of localization ''looks'' (Buell and Hafter, 1988), each noise burst was 100% amplitude modulated with a 40-Hz sinusoid, phased to begin and end with zero slope. Thus, each noise burst was essentially 20 bursts of 25-ms duration each. The 3-kHz test signal was obtained by filtering the wideband signal with a 40th-order Butterworth low-pass filter having a 3-kHz cutoff frequency. The convolution of the test signals with the HRTFs was done numerically in MATLAB. In addition, the resulting signals were filtered by a headphone compensation filter designed following Møller's procedure (Møller, 1992). The resulting sound files were played back through AKG 240-DF headphones using a PC equipped with a Turtle-Beach Tahiti sound board. Although the energy in the test signal was constant, the variation of the HRTF with elevation produced a corresponding small variation in loudness, with an average SPL of 73 dB. Finally, the electrical signals driving the headphones were analyzed with a spectrum analyzer to verify that nonlinearities or noise in the processing and the hardware were not introducing spurious high-frequency signals.

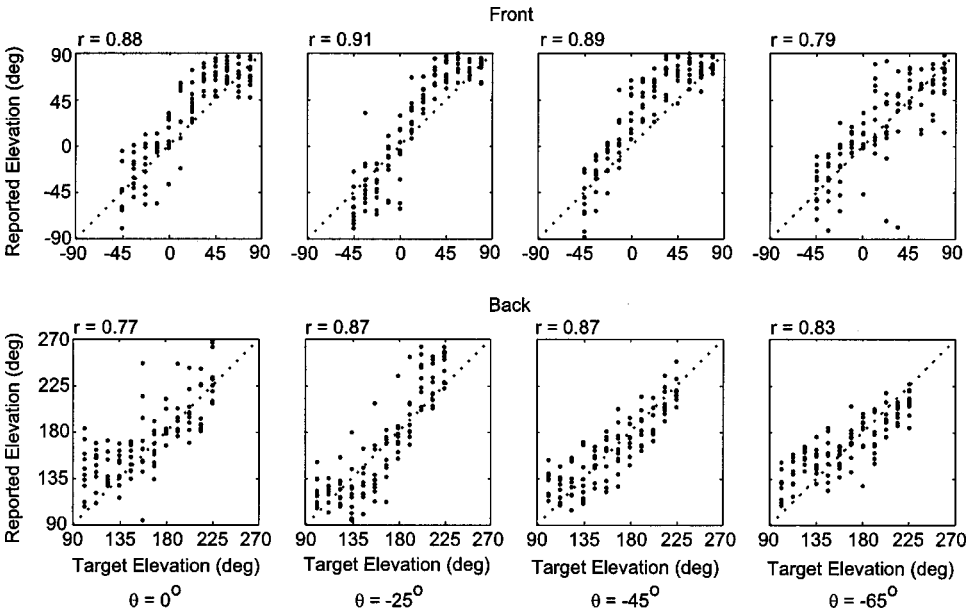

FIG. 2. Scatterplots for judged source elevation versus actual elevation for Subject S6 for a 22-kHz-bandwidth source at four different azimuths. In the top row the sound source was in the front hemisphere, while in the bottom row it was in back. Each plot shows data for 10 judgments at each of 12 different elevations, together with the sample correlation coefficient. The performance is comparable for all azimuths and hemispheres.

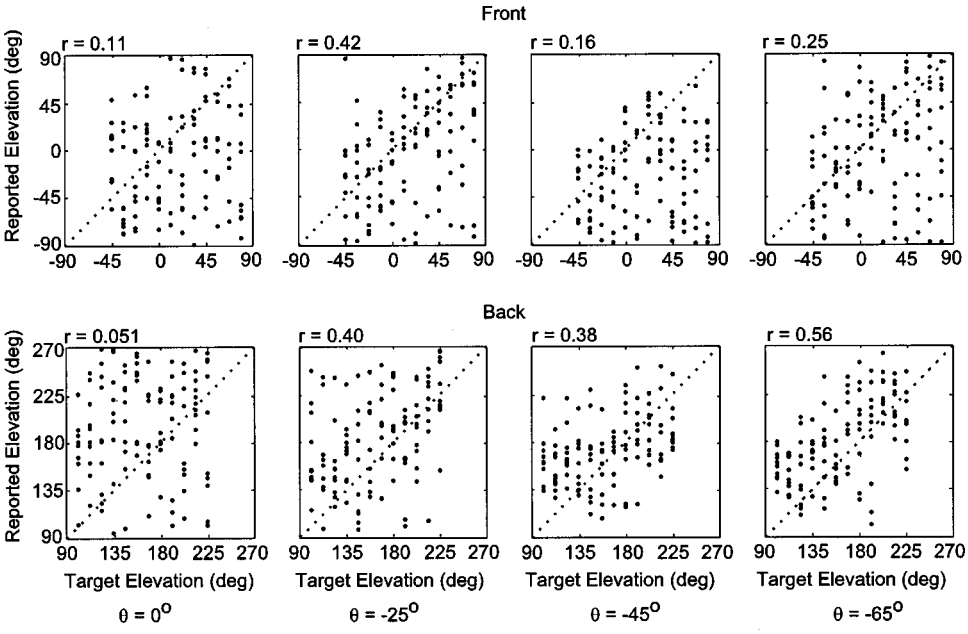

FIG. 3. Scatterplots as in Fig. 2, but with the signal low-pass filtered to remove frequency components above 3 kHz. Performance in the median plane ($\theta = 0°$) is severely degraded. As the magnitude of the azimuth increases, the performance improves, particularly for sources in the back hemisphere.

III. EXPERIMENTS WITH MEASURED HRTFS

Scatterplots of experimental results for a typical subject (S6) using full-bandwidth and 3-kHz low-pass stimuli are shown in Figs. 2 and 3. The eight situations shown in Fig. 2 are for the 22-kHz-bandwidth source at the four different azimuths. All eight cases are quite comparable, showing that the accuracy of judging elevation was not particularly sensitive to whether the source was in the median plane or on any of the cones of confusion, or whether the source was in front or in back. By contrast, Fig. 3 shows that when the maximum signal frequency was reduced to 3 kHz, performance was very poor in the median plane, but improved at other azimuths. Figure 4 shows similar 3-kHz bandwidth results for another subject (S1). Once again, the subject performed very poorly in the median plane, and was more accurate in the back than in front away from the median plane. While wideband results confirm that high frequencies are the major contributors to elevation perception, it is surprising that, away from the median plane, one can still judge elevation with a low-bandwidth source.

The effect of reducing the bandwidth can be measured by the change in the sample correlation coefficient. For Subject S6 we observe that the degradation in the median plane was about 90% in both hemispheres. The performance was better for azimuths away from the median plane and was better in back than in front. Figure 5 shows that this general trend was exhibited by the majority of the subjects tested. This figure compares side-by-side the sample correlation coefficients for full-bandwidth stimuli and for 3-kHz low-pass stimuli for all subjects and all azimuths. The average correlation coefficient r for all subjects is summarized in Table I for both wideband and low-pass tests.

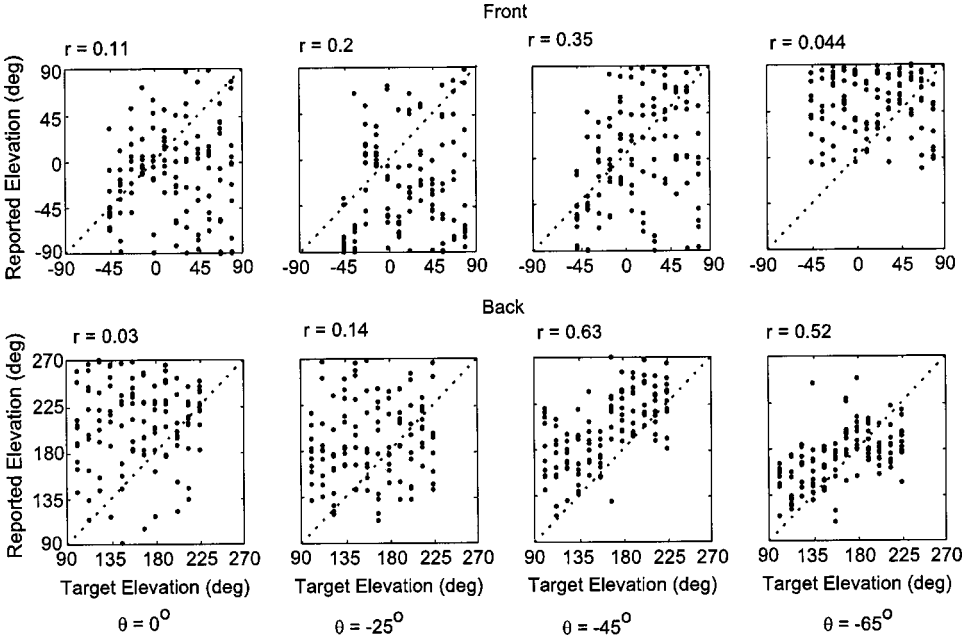

FIG. 4. Scatterplots as in Fig. 3, but for Subject S1. The performance is generally similar. In both cases, performance in the median plane is severely degraded, but a good correlation appears for sources away from the median plane and in back.

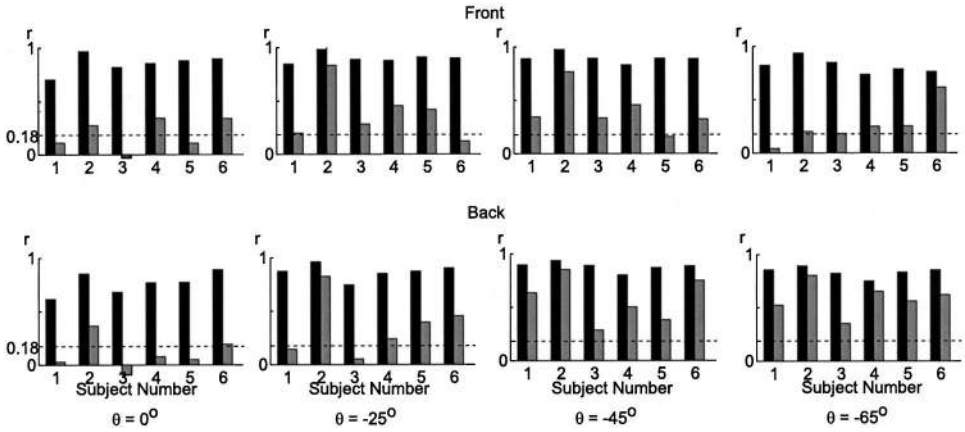

FIG. 5. Comparison of low-pass and full-bandwidth correlation coefficients for all subjects. Black: full bandwidth; gray: 3-kHz low pass. Values of $|r|$ above 0.18 are statistically significant at the 95% level.

A standard significance test for the sample correlation coefficient r is the Fisher z statistic, $z=0.5\ln(1+r)/(1-r)$; if the true correlation coefficient is ρ and if the sample size n is greater than 10, this statistic is approximately normally distributed with mean $0.5\ln(1+\rho)/(1-\rho)$ and variance $1/(n-3)$ (Cramer, 1946). For our data, where $n=120$, any correlation whose magnitude is less than 0.18 is not statistically significant at the 95% confidence level.

Analysis of the performance of individual subjects shows that the correlation was always statistically significant for the full-bandwidth source for all subjects. When the signal was low-pass filtered and the source was in the median plane, the correlation was not significant for most subjects and the degradation in performance was highest. When the source was away from the median plane, the performance improved, as shown in Fig. 5 and Table I, and was best in the back.

Inspection of the scatterplots in Figs. 3 and 4 reveals a significant amount of bias in the subjects' estimates. To be more specific, most of the time the subjects estimated the virtual source location to be lower than it actually was. As a measure of accuracy, the correlation coefficient is invariant to bias, but the rms error includes it.[4] Table II shows both the bias and the rms error (in degrees), averaged over all six subjects for each experimental condition. The rms error for random guessing between $-90°$ and $+90°$ is 51.96°, and the rms values for low-pass stimuli in front or in the median plane indicate performance at the chance level. However, lower rms errors are achieved when the source is away from the median plane and in back. Because bias contributed significantly to the rms error, we believe that the correlation coefficient is a better indicator that low-frequency information is providing an elevation cue.

Finally, we observe that the results were subject dependent. At the extremes, one subject performed poorly in both the wideband and low-pass tests, while another subject had a surprisingly good performance in all the low-pass tests, and at $-45°$ and $-65°$ in the back had an increase in rms error from 20° to 23° (less than 20%) when the bandwidth was reduced from 22 to 3 kHz.

IV. LOW-FREQUENCY HRTF ANALYSIS

The perceptual experiments in the previous section confirmed the existence of low-frequency elevation cues. The physical sources of these cues are reflected in features present in the HRTFs. Given the frequency range in which these features appear, it is natural to assume that they are caused by larger body structures such as the torso and head, whose dimensions are comparable to the wavelengths in question. Although Gardner (1973) and Kuhn (1987) showed that the effects of the pinnae on the spectrum become noticeable above 3.5 kHz, it was important to establish that they were negligible below 3 kHz.

The hypothesis that the low-frequency elevation cues were not due to the pinnae was tested in three ways:

(1) By analyzing and identifying features of measured HRTFs obtained by including or removing different body parts (pinnae or torso);
(2) By synthesizing HRTFs based on simple torso and head models and comparing such synthetic HRTFs to measurements; and
(3) By psychoacoustic tests of perceived elevation for customized approximations to the HRTFs that are based solely on the geometry of the torso and of the head.

Several sets of HRTFs obtained by including or removing the pinnae and torso of a KEMAR mannequin were analyzed. The goal was to separate the effects of the different anatomical structures and to isolate their partial contributions to the low-frequency portion of the HRTFs. Strictly speaking, these contributions cannot be isolated this way, because the combination of structures does not imply the superposition of their acoustic fields. However, the effects of the torso, head, and pinnae are sufficiently separated in time, frequency, and spatial location that they can be observed by

TABLE I. Average correlation coefficient r for four different azimuths. F=front and B=back.

Condition	F, 0°	B, 0°	F, −25°	B, −25°	F, −45°	B, −45°	F, −65°	B, −65°
Wideband HRTF	0.86	0.75	0.90	0.87	0.89	0.88	0.82	0.83
Low-pass HRTF	0.19	0.10	0.39	0.35	0.40	0.57	0.24	0.58

TABLE II. The average rms error and bias. W=wideband, L=3-kHz low pass.

Condition	F, 0°	B, 0°	F, −25°	B, −25°	F, −45°	B, −45°	F, −65°	B, −65°
W rms	25.8	27.9	25.7	22.0	27.7	21.9	28.3	22.3
W bias	5.8	5.5	7.2	5.4	9.8	5.5	6.6	3.7
L rms	55.9	57.5	51.8	47.3	50.0	40.0	53.3	37.1
L bias	18.4	21.8	20.3	18.0	18.3	15.5	19.4	13.2

selecting the domain in which their individual influences dominate.

A. HRTF data

Three sets of HRTFs of a KEMAR mannequin were obtained by including or removing different anatomical structures. The data sets were collected according to the combinations shown in Table III.

The HRTFs of two human subjects were also measured. For each subject, two HRTFs were measured, a standard HRTF and a ''pinna-less'' HRTF, obtained by suppressing the effects of the subjects' pinnae. This was achieved by the use of a rubber swimming cap that covered the outer ears. Adhesive tape was placed on the pinna regions to further smooth the surface. Microphone probe tubes were placed on the outside surface of the tape at positions corresponding to the ear canals. All measurements were made at the same spatial locations and with the techniques described in Sec. II.

B. Contribution of the pinnae

The contribution of the pinnae to the HRTFs at low frequencies can readily be evaluated on a KEMAR mannequin with removable pinnae. Figure 6 illustrates the elevation dependence of the KEMAR HRTF with and without pinnae. The measurements were made for the ipsilateral ear on a cone of confusion at $\theta = -45°$. The squared magnitudes of the HRTFs were smoothed with simple auditory filters ($Q = 8$) and the results were displayed as images. In these image displays, the HRTF data at a particular elevation are displayed along a vertical line, where the gray scale indicates power in decibels. Because 90° elevation is in the center, front/back differences are revealed as lack of bilateral symmetry in the images.

Clearly, the pinnae have a major effect on the spectrum above 3 kHz, but relatively little effect below 3 kHz. Below 3 kHz, the average difference between the spectra with and without pinnae is 0.86 dB. Thus, the pinnae do not appear to contribute significant monaural cues below 3 kHz. However, in both cases, one can see elevation-dependent, arch-shaped notches in the spectrum that extend as low as 700 Hz. These are potential sources of elevation information that are clearly not due to the pinnae.

The contribution of the pinnae to binaural ILD cues at low frequencies was also evaluated. The ILD was computed as the difference between the right and the left dB values of the smoothed HRTF spectra. For frequencies below 3 kHz, a comparison of the ILDs of data set 1 (both pinnae and torso present) and data set 2 (pinnae removed) in the cone of confusion at $\theta = -45°$ is shown in Figs. 7(a) and (b). The magnitude of the ILD is shown in a gray scale as a function of elevation and frequency. We also evaluated the ILD for two human subjects. In Figs. 7(c) and (d) we show the ILDs for the one of these subjects. For the KEMAR mannequin and for both of the two human subjects, the contribution of the pinnae to the low-frequency ILD was insignificant.

The essential identity of the pinnae/no-pinnae ILDs pairs for frequencies below 3 kHz was observed for all azimuths and for all subjects. This is in agreement with the observation of Kuhn (see Fig. 14 in Kuhn, 1977), who attributed the ILD variations he observed in this frequency range to the torso.

C. Contribution of the torso

Now that it has been established that the effect of pinnae is negligible below 3 kHz, what remains to be clarified is the nature of the separate head and torso contributions to the low-frequency cues. To this end we make use of the measurements in data set 2 (pinnae removed). The removal of the pinnae reduces the complexity of the HRIRs, particularly on the contralateral side, and simplifies identification of the head and torso contributions.

Figure 8 shows both the HRIR and the HRTF of KEMAR for an azimuth angle of −45° with torso but no pinnae (data set 2). Both ipsilateral and contralateral responses are displayed as functions of elevation and of time or frequency. The ipsilateral HRTF image is clearly brighter than the contralateral image, which is a consequence of the ILD at −45° azimuth. Notice that the ipsilateral HRTF data (the lower-left panel) are actually the same as in the right panel in Fig. 6; the difference in visual appearance is due to a combination of (a) a linear instead of a logarithmic frequency scale, and (b) a gray scale that encompasses both the high-amplitude ipsilateral data and the low-amplitude contralateral data.

The HRIR images shown in Fig. 8(a) expose features of the HRTF that are hard to see in the frequency domain, and they deserve a more detailed description. In either image, an impulse response at a particular elevation is displayed along vertical line. To reduce the effect of the ILD on ''washing out'' the contralateral image, the impulse responses were scaled so that the maximum magnitude was unity for both the ipsilateral and the contralateral ear. As the color bar on the right indicates, bright values are positive and dark values are negative. The gray band at the very top of either image

TABLE III. KEMAR HRTF data sets.

Set	Pinnae	Torso
1	Yes	Yes
2	No	Yes
3	No	No

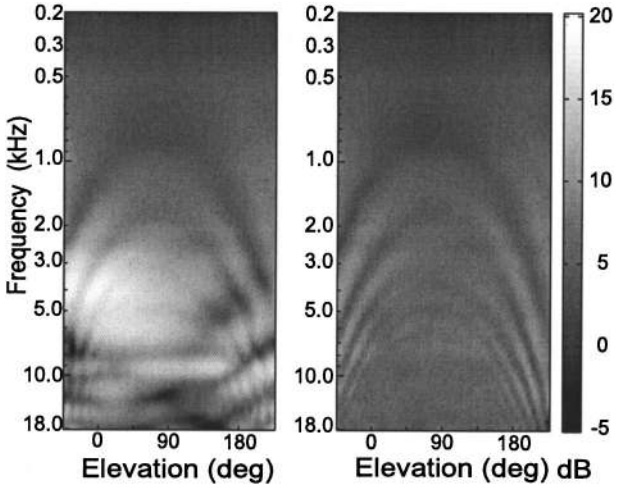

FIG. 6. Comparison of HRTF spectra. The left panel shows the spectrum with the pinnae attached, and the right panel shows the effect of removing the pinnae. The data are for the left ear at $\theta = -45°$, so that these are ipsilateral data. The measurements were smoothed by a constant-Q auditory filter ($Q = 8$). The gray scale indicates the magnitude of the smoothed spectra in decibels. The elevation-dependent arch-shaped patterns that are present in both cases are due to head and torso effects. Notice that they extend down to fairly low frequencies (below 3 kHz).

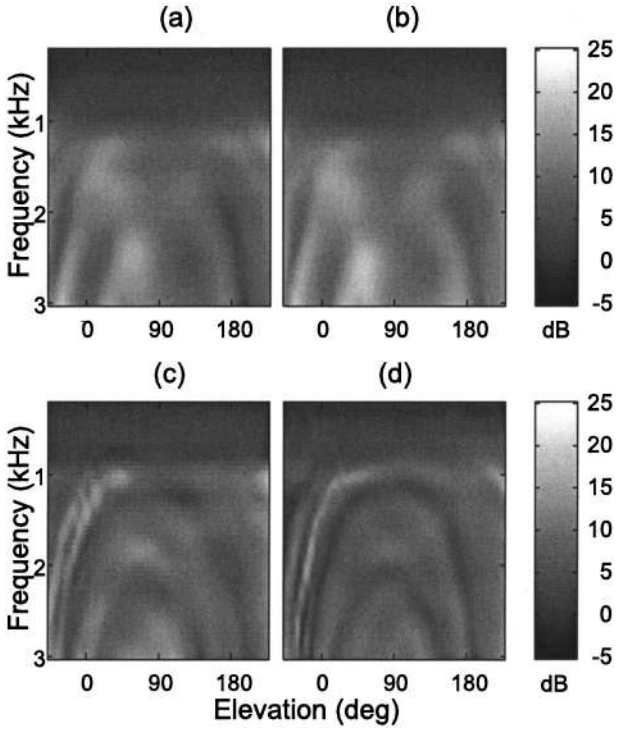

FIG. 7. Comparison of ILDs with and without pinnae. (a) KEMAR with pinnae; (b) KEMAR without pinnae; (c) Subject SA1; (d) Subject SA1 with pinnae ''removed.'' Data shown for $\theta = -45°$ and frequencies below 3 kHz.

corresponds to the zero value before the impulse response starts. The strong white band or ridge near the top corresponds to the initial peak of the response. This peak was actually ''clipped'' to allow the weaker parts of the impulse to be visible. This initial ridge is horizontal in the ipsilateral image because the time of arrival was the same for all elevations. The initial ridge occurs about 0.4 ms later in the contralateral image than in the ipsilateral image, corresponding to the ITD at $-45°$ azimuth. Note that the ITD is actually not constant, but varies by about ± 0.1 ms; this phenomenon is discussed further in Sec. IV E.

The initial pulse is followed by a series of subsequent pulses. We focus on the response of the ipsilateral ear (upper-left panel of Fig. 8) because it is simpler than the response of the contralateral ear. Probably the most prominent feature is the pair of V-shaped ridges, one that is stronger in the front and one that is stronger in the back. From the way that these delays increase and then decrease with elevation, we infer that the reflections come from below the ears. The delays are maximum for sound source locations above the subject (at about $\phi = 90°$). The maximum delay of about 1 ms corresponds to a distance of 33 cm, which is roughly twice the distance from the ear canal to the shoulder. Thus, the pattern of delays suggests that the reflections are indeed due to a specular reflection from the torso. This was further verified using data set 3, where removal of the torso resulted in a loss of these reflections (compare the upper-right panel of Fig. 8 and the upper-middle panel of Fig. 9).

In the frequency domain the torso reflections act as a comb filter, introducing roughly bilaterally symmetric, arch-shaped periodic notches in the spectrum that are particularly clear for the ipsilateral ear [see Fig. 8(b)]. The frequencies at which the notches occur are inversely related to the delays, and thus produce a pattern that varies with elevation. The lowest notch frequency corresponds to the longest delay. Delays longer than a sixth of a millisecond will produce one or more notches below 3 kHz and will contribute to the low-frequency ILD of Fig. 7. Although the complexity of response of the contralateral ear makes it somewhat difficult to see, analysis of data set 3 in the frequency domain confirmed that removing the torso indeed eliminated the large arch-

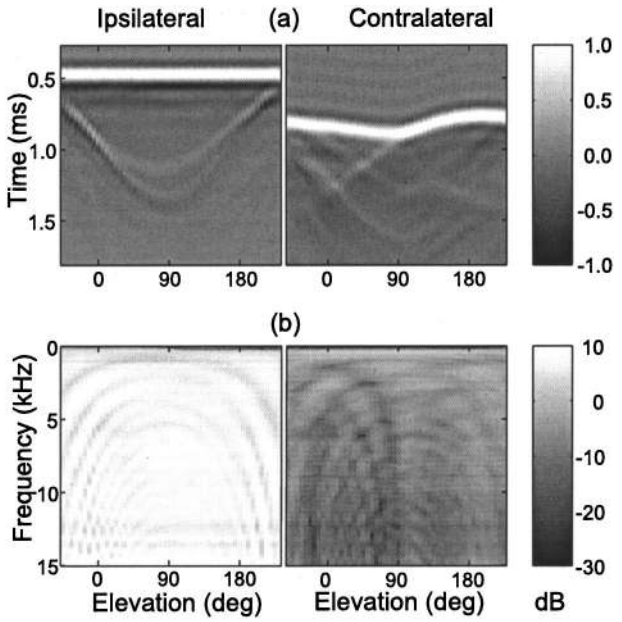

FIG. 8. (a) HRIRs and (b) magnitude HRTFs for KEMAR with no pinnae. The responses are shown for the cone of confusion at $\theta = -45°$ and frequencies up to 15 kHz. In the time-domain plots the amplitude of the HRIRs has been scaled to enhance the gray-scale image.

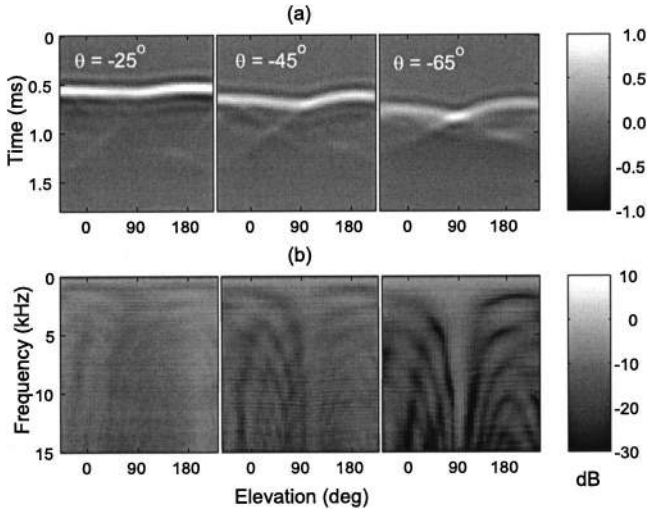

FIG. 9. (a) Right HRIR data for the KEMAR head with no pinnae and no torso. (b) Magnitude of the HRTF. Three azimuths on the contralateral side are shown.

shaped notches [compare the lower-right panel of Fig. 8(b) and the lower-middle panel of Fig. 9(b)].

The contralateral impulse response in the right panel of Fig. 8(a) exhibits similar but weaker torso reflections, with their corresponding notches in frequency domain. The contralateral response displays other features, not explained by torso reflections, that become visible because of the relative weakness of the direct sound and torso reflections. These features are considered further in Sec. IV E.

Next, we develop a simple geometrical model for the torso that accounts for the delayed reflections.

D. Geometric model of the torso

Although the human torso does not have a regular shape, it can be approximated by a simple ellipsoid, illustrated in Fig. A1 in the Appendix. The choice of an ellipsoid is based on analytical simplicity and its small number of parameters, which can be related to and estimated from anthropometry (height, width, depth). An algorithm for computing the delay $D(\theta,\phi)$ of the torso reflection relative to the initial pulse as a function of azimuth, elevation, and the geometrical parameters is outlined in the Appendix.

This algorithm was used to compute the delays using anthropometric measurements for three subjects (KEMAR and two humans). Considering the simplicity of the model, the resulting delays were remarkably close to the measured data. Figure 10 compares the delays produced by the model against the delays measured from the corresponding HRIR data (data set 2—with torso but without pinnae). The three subjects exhibited different torso reflection patterns that depended on body dimensions, and the anthropometry-based geometric model was able to account for these differences. Figure 10 shows that the behavior of the model follows the measured data closely.

E. Contribution of the head

Given its size, the head is the other anatomical structure that may contribute elevation-dependent features at low frequencies. To isolate the effect of the head, we use measurements with both the pinnae and torso removed. The resulting ipsilateral response is rather featureless, because the energy of the direct sound is large relative to the energy of the secondary waves that are diffracted around the head (Avendano, Duda, and Algazi, 1999). Thus, here we focus on the contralateral response.

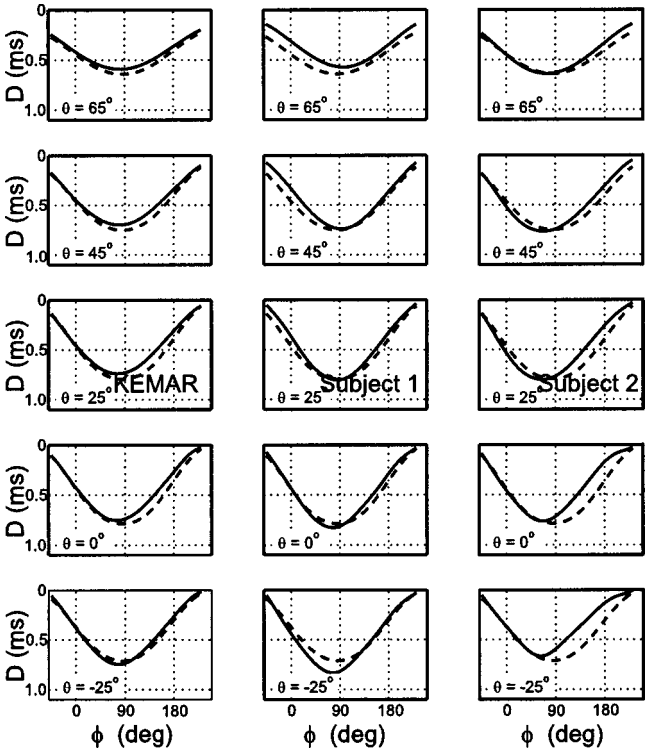

FIG. 10. Comparison of the delay D from the model (dashed) and the measured subject data (solid): (a) KEMAR; (b) Subject 1; and (c) Subject 2. The delays are shown as functions of elevation for five azimuths in each case.

Figure 9 displays contralateral HRTF data in data set 3 (both pinnae and torso removed) for three different azimuths ($-25°$, $-45°$, and $-65°$). The impulse response exhibits a prominent X-shaped pattern, particularly away from the median plane [see Fig. 9(a)]. A simplified explanation is that the incident sound wave travels to the contralateral ear by two paths, one around the front of the head and the other around the back (Duda and Martens, 1998); the upper or primary part of the X-shaped pattern arises from the shorter path, and the lower or secondary part from the longer path.

As we noted earlier, the onset of the primary wave varies slightly as a function of elevation, indicating some elevation asymmetry. This asymmetry has been discussed in Duda, Avendano, and Algazi (1999), where it was observed that the ITD on a cone of confusion is actually not constant, but can vary by as much as 0.12 ms as a function of elevation. For a spherical head, the HRTF can be computed exactly from the head radius and the angle of incidence, the angle between the source and the position of the ear canal (Duda and Martens, 1998). If the ear canals are diametrically opposed, the primary and secondary waves would each have the same delay for all elevations on a cone of confusion, and no X-shaped pattern would be seen. However, an X-shaped

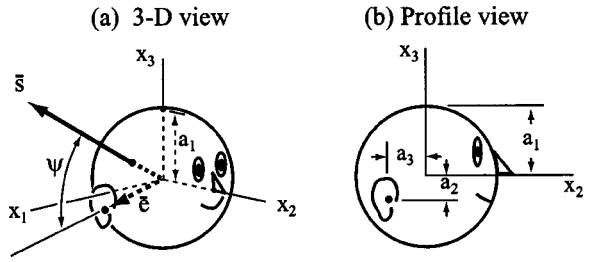

FIG. 11. Geometry for the head model. Here, $\bar{s}$ is a vector from the center of the head to the sound source, $\bar{e}$ is a vector from the center of the head through the entrance of the ear canal, and ψ is the angle between them. The anthropometric parameters are the head radius a_1, the downward offset of the ear a_2, and the backward offset of the ear a_3.

pattern appears if the ears are displaced. Several researchers have noted that human ears typically lie behind and below the horizontal axis (Genuit, 1984; Blauert, 1997). Because the interaural axis defines the axis of rotation, this displacement causes the angle of incidence to change as the source moves around the cone of confusion, with larger changes occurring towards the contralateral hemisphere. Although other factors, such as the nonspherical shape of the head, also affect the time delay (Duda *et al.*, 1999), the ear location is particularly important.

F. Geometric model of the head

A simple spherical-head-with-offset-ears model is now used to account for the features observed in Fig. 9. With this model, both the ILD and the ITD vary on a cone of confusion. The HRTF for the sphere is obtained from Rayleigh's infinite series solution to the equations for the diffraction of sound by a sphere (Duda and Martens, 1998). To compute the transfer function from the source to the ear, three quantities are needed: the distance r to the source, the angle of incidence ψ, and the head radius a_1 [see Fig. 11(a)]. The distance to the source was 1 m for our experimental data. The angle of incidence ψ is the angle between the vector $\bar{s}$ to the source and the vector $\bar{e}$ to the ear: $\psi = \cos^{-1}[(\bar{s}^T\bar{e})/\|\bar{s}\|\,\|\bar{e}\|]$, where $\bar{s}^T$ is the transpose of $\bar{s}$ and $r = \|\bar{s}\|$ is the length of $\bar{s}$ (see Fig. 11). The only anthropometric data needed are the head radius a_1 and the vector $\bar{e}$, which is determined by the offsets of the ear down a_2, and back a_3.

A comparison between the spherical head model with size and offset parameters extracted from KEMAR ($a_1 = 8.5$ cm, $a_2 = 3$ cm, and $a_3 = 0.5$ cm) and the data in data set 3 (both pinnae and torso removed) reveals that the spherical-head-with-offset-ears model provides a good approximation to the elevation-dependent patterns in both the frequency and the time domain (cf. Figs. 9 and 12). Notice that the X-shaped pattern due to the elevation-dependent onset and secondary waves is introduced by the ear offset. As expected, some discrepancies remain, because neither a human head nor KEMAR's head is really spherical, and effects of the neck have not been modeled. However, the basic elevation-dependent features introduced by the head appear to be captured.

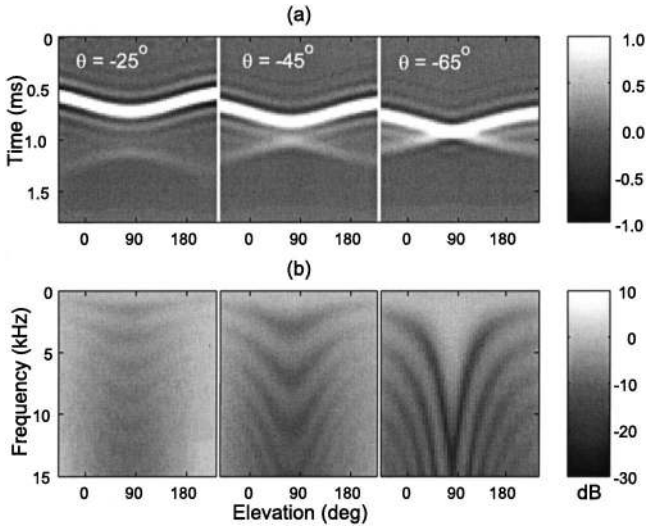

FIG. 12. (a) The HRIR and (b) the magnitude of the HRTF for the head model at three different azimuths on the contralateral side. A comparison with Fig. 9 shows a good general correspondence with the measured data.

V. EXPERIMENTS WITH A HEAD-AND-TORSO APPROXIMATION

In Sec. IV, we demonstrated that a simple geometrical head-and-torso (HAT) model accounts for the behavior of the low-frequency experimental HRTFs. In this section, we report on psychophysical experiments employing a classical spherical-head model and an empirical torso-delay model. Although this HAT approximation does not capture all the details of the experimental HRTFs at low frequencies, it incorporates the principal subject-dependent effects of the head, shoulders, and torso. Thus, the purpose of these new experiments is to assess the elevation cues that are conveyed by simple geometrical features, individualized for each subject.

The spherical-head model was computed from the infinite-series solution to the problem of the scattering of acoustic waves from a point source by a rigid sphere (Duda and Martens, 1998). The resulting HRTF $H_s(i\omega, r, \psi, a_1)$ depends on the angular frequency ω, the distance r from the center of the head to the source, the incidence angle ψ between the ear and the source, and the radius a_1 of the sphere. The HAT model approximates the complete HRTF by assuming that the wave incident on the head is the sum of a direct wave and a weaker torso reflection that arrives after a delay $D(\theta, \phi)$ that depends on azimuth θ and elevation ϕ. For simplicity, it was assumed that the direct wave and the torso reflection arrive from the same direction, so that the HAT HRTF can be written as

$$H_{\mathrm{HAT}}(i\omega) = \alpha[1 + \rho e^{i\omega D(\theta,\phi)}]H_s(i\omega, r, \psi, a_1),$$

where ρ is the torso reflection coefficient, and $\alpha = 1/(1+\rho)$ is a scale factor that guarantees that $H_{\mathrm{HAT}}(0) = 1$.[5]

The resulting H_{HAT} was individualized for each of the six subjects by making separate estimates for the various parameters. For all subjects, we used $r = 1$ m, because that was the range for the measured data, and for simplicity we assumed that $\rho = 1/3$, independent of direction or frequency.[6] The head radius a_1 and the ear locations (which are needed

TABLE IV. Average correlation coefficient r for four different azimuths. F=front and B=back.

Condition	F, 0°	B, 0°	F, −25°	B, −25°	F, −45°	B, −45°	F, −65°	B, −65°
Wideband HRTF	0.86	0.75	0.90	0.87	0.89	0.88	0.82	0.83
Low-pass HRTF	0.19	0.10	0.39	0.35	0.40	0.57	0.24	0.58
Low-pass HAT model	0.11	0.16	0.25	0.47	0.42	0.66	0.05	0.47

to calculate the incidence angle ψ) were individualized for each subject by optimizing a least-squares fit to experimentally measured ITD data estimated from individual HRIR images like those shown in Fig. 8. We could have used the ellipsoidal torso model to compute the delay $D(\theta,\phi)$ of the torso reflection, but, as Fig. 10 illustrates, that would have introduced some additional error into the HAT approximation. Instead, we chose to determine the torso delays from measurements taken from individual HRIR images.

The experiments conducted with the HAT approximation used the signals and methods described in Sec. II. As before, a 3-kHz stimulus was produced by filtering the wideband, amplitude-modulated noise signal with a 40th-order Butterworth filter having a 3-kHz cutoff frequency. That low-pass signal was then convolved with the location-dependent HAT HRTF approximation. Localization accuracy was measured in eight different situations, for azimuth angles θ of 0°, −25°, −45°, −65°, using a source location either in front or in back. The results for the HAT approximation could therefore be compared directly to the results obtained for each subject's measured HRTF with the same 3-kHz low-pass stimulus.

As a whole, the results of these experiments with the HAT approximation complement and confirm the results obtained with measured HRTFs. The results are summarized in Table IV, which adds to Table I the correlation coefficient for all eight conditions for the HAT approximation, averaged over the six subjects used in the study. We note that the HAT approximation and the measured HRTF gave quite similar results. Performance in the median plane was very poor, and the larger correlations occurred away from the median plane and in the back.

However, examination of the details of individual results reveals some interesting differences. For all subjects, the HAT approximation provided a more consistent elevation cue than the measured HRTFs. However, for some subjects the correspondence between intended and perceived elevations was poorer when the HAT approximation was used. These observations are exemplified by the experimental data of Subject S6 (Figs. 3 and 13) and Subject S1 (Figs. 4 and 14). These figures show that the HAT approximation led to substantially less scatter of reported elevations for each target elevation than when the measured HRTF was used. However, with the HAT model, target elevations between 90° and 140° were not well discriminated, with the mean being around 160° regardless of target elevation, while target elevations greater than 140° were more consistently and correctly reported. Thus, the linear correspondence between target and reported elevations that the correlation coefficient measures is only a partial characterization of the differences between the results for the measured HRTF and for the HAT approximation.

VI. DISCUSSION AND CONCLUSIONS

The experimental results reported have clearly established the existence of low-frequency cues for elevation that are significant away from the median plane. The analysis of the HRTFs has shown that the HRTF features below 3 kHz are primarily due to the torso reflection and head diffraction,

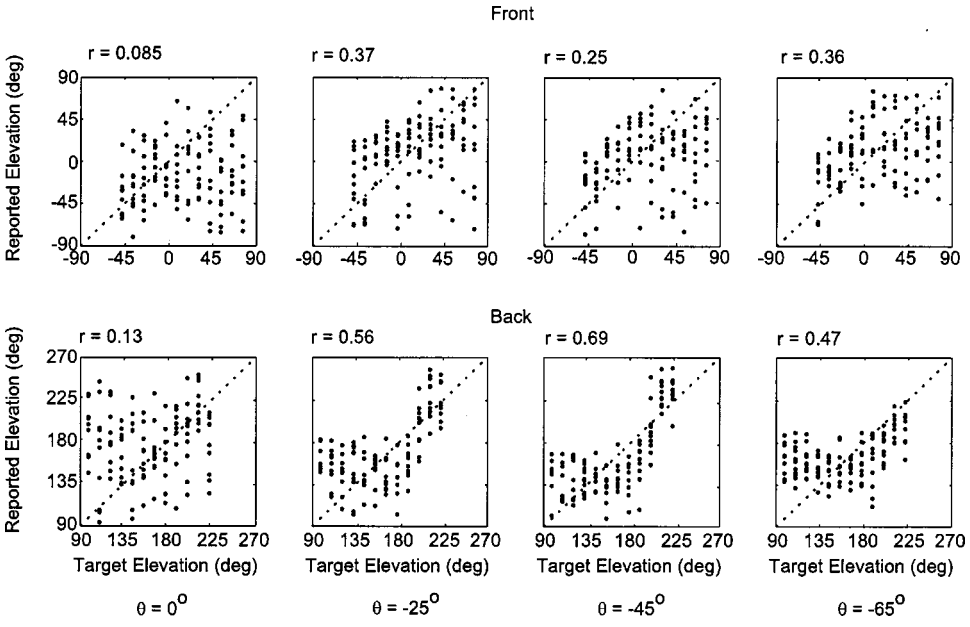

FIG. 13. Scatterplots for the HAT model, 3-kHz bandwidth, Subject S6. A comparison with Fig. 3 where the measured HRTF was used shows very similar results. The ability to localize in back actually appears to be better than the performance with the measured HRTF. However, at azimuths of −25° and −45°, the HAT model seems to lead to more of a bimodal (low/high) response.

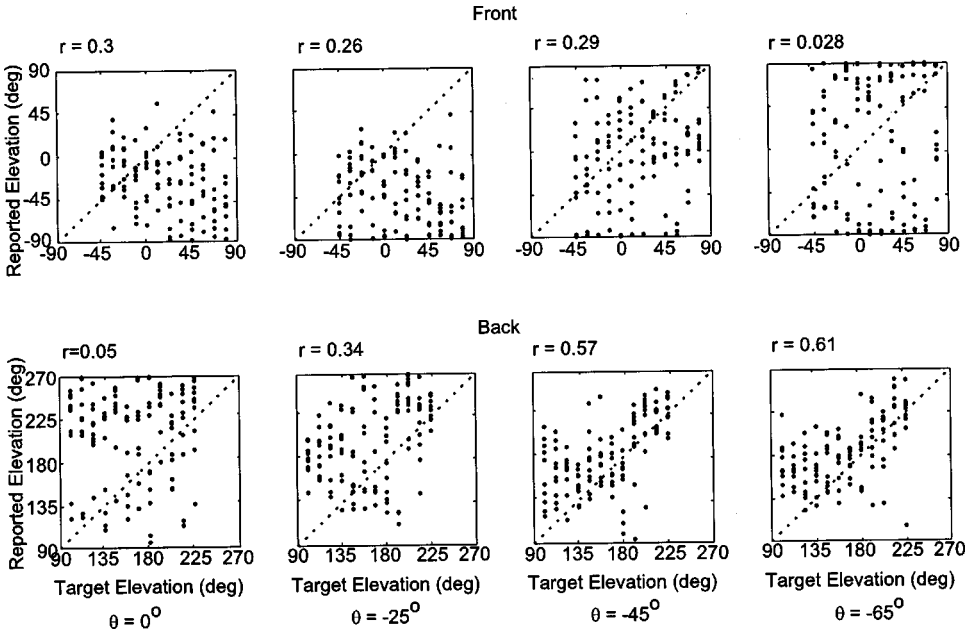

FIG. 14. Scatterplots for the HAT model, 3-kHz bandwidth, Subject S1. A comparison with Fig. 4 where the measured HRTF was used shows very similar results. The greatest difference occurs at large azimuth in front, where very few low elevations were reported with the measured HRTF. However, elsewhere the results are quite comparable.

while the pinnae do not contribute significantly at these low frequencies. The torso reflection effects are stronger on the ipsilateral side, while the head diffraction effects are stronger on the contralateral side where the direct sound is attenuated by the head. Further, it was shown that simple geometric models for the head and the torso provide strong corroboration of the physical basis for low-frequency elevation cues. The parameters of these models can be estimated from anthropometry to account for individual differences. A simple head-and-torso (HAT) geometric model was used to synthesize approximate HRTFs. Below 3 kHz, the synthetic HRTF was basically similar to the measured HRTF. Psychoacoustic experiments were conducted with an individualized HAT approximation of low-frequency HRTF data. It was observed that the approximate HRTFs provided low-frequency elevation cues that were just as effective as those provided by the measured HRTFs.

This study did not systematically examine other possible sources of low-frequency elevation cues. We now discuss these briefly and speculate on their importance on the basis of the results of this work. First, the changes of the ITD with elevation that were discussed in Sec. IV E could provide elevation cues. However, these ITD deviations are significant in only a fairly small range of spatial locations, and could not by themselves explain the full range of low-frequency effects observed. Second, timbre and loudness are monaural spectral properties that vary with elevation. Based on the results reported for the median plane in this and previous studies, these physical variations are clearly ineffective as low-frequency elevation cues. Finally, there are other larger anatomical structures (such as the legs) that effect the HRTF at low frequencies. Although not included in this paper, other HRIR measurements with seated subjects reveal knee reflections at low elevations and in the front, but they vanished at about $-35°$ and occurred only in the front where low-frequency elevation cues are weak. Thus, we believe that knee reflections can at best provide very limited elevation cues. An interesting unanswered question is the general effect of posture or of head rotation on low-frequency elevation cues.

The existence of low-frequency cues has implications for the binaural simulation of virtual sources. Spherical head models are commonly used to estimate the low-frequency behavior of the HRTF; this work suggests that the torso provides additional cues that also should be taken into account. Finally, recognition of the presence of low-frequency cues provides a possible opportunity for enhancing elevation cues for listeners with hearing loss at higher frequencies.

ACKNOWLEDGMENTS

The authors would like to thank Dennis Thompson for his help with much of the experimental work, and the Editor and anonymous reviewers for their very thorough, thoughtful, and helpful suggestions. Support of the research was provided by the University of California DiMI University-Industry collaborative program, by the Creative Advanced Technology Center, by Aureal, by the Interval Research Corporation and by the Hewlett-Packard Research Laboratories. Support was also provided by the National Science Foundation under Grants No. NSF IRI-96-19339 and NSF 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: THE ELLIPSOIDAL TORSO MODEL

This Appendix explains the algorithm used to compute the time delay $D(\theta,\phi)$ for the torso reflection as a function of the azimuth θ and elevation ϕ of the sound source. The geometry for the ellipsoidal torso model is shown in Fig. A1, which identifies the following anthropometric parameters:

a_1—head radius;
a_2—ear-canal offset down;
a_3—ear-canal offset back;

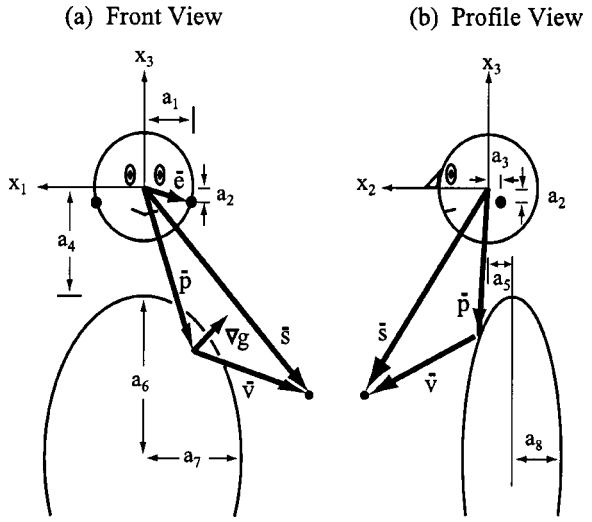

FIG. A1. Anthropometry for the torso model and related geometry.

a_4—distance from the center of the head to the top of the torso;
a_5—displacement of the head in front of the torso;
a_6—torso half-height;
a_7—torso half-width;
a_8—torso half-depth.

In contrast to the spherical-head model, we do not attempt to solve the wave equation for the ellipsoid, for which there is no simple analytical solution. Instead, we assume that the ellipsoid is a rigid surface and a specular reflector for sound with suitably short wavelengths. This approach is justified by the data, which exhibit a strong isolated reflection due to the torso. Thus, a ray-tracing algorithm is used to compute the time delay $D(\theta,\phi)$ of the torso reflection.

The algorithm can be outlined as follows. Given a sound source at the point $\bar{s}$, the problem is to compute the point $\bar{p}$ on the surface of the ellipsoid where the reflection will occur, and use $\bar{p}$ to calculate the difference in path lengths to the ears $\bar{e}$ for the direct and the reflected sound waves. The calculation makes use of the vector $\bar{v}=\bar{s}-\bar{p}$ from the reflection point $\bar{p}$ to the source $\bar{s}$. Once $\bar{p}$ is determined, the torso reflection delay is obtained by first computing the difference between the path length for the direct and the reflected sound from the source to the center of the head, $d=\|\bar{p}\|+\|\bar{v}\|-\|\bar{s}\|$, where $\|\bar{p}\|$ is the length of $\bar{p}$. A correction based on Woodworth's formula (Blauert, 1997) is then applied to account for the additional distance of each component to the ear position $\bar{e}$. The total delay is obtained as $D(\theta,\phi)=(d+d_r+d_s)/c$ where c is the speed of sound in air (340 m/s) and d_r and d_s are the corrections for the diffraction around the head for the reflection and the source, respectively. For example, the correction for the direct sound can be computed as $d_s=a_1\sin(\psi_s-\pi/2)$, where ψ_s is the angle between the source vector $\bar{s}$ and the ear vector $\bar{e}$. This formula gives positive values for angles of incidence greater than 90°, and negative otherwise. The same formula is applied to correct the path length of the reflection.

The main problem is to compute the reflection point $\bar{p}$ on the surface of the ellipsoid for a given source location $\bar{s}$. Our approach is to work backwards, stepping systematically across the surface of the ellipsoid at points $\bar{p}_i$ to find the source direction $\bar{s}_i$ that would cause a reflection at that point. For a given $\bar{p}=\bar{p}_i$ we apply Snell's law to determine the direction $\bar{u}$ of the incident sound vector $\bar{v}=\alpha\bar{u}$. To obtain $\bar{u}$ we first compute the normal to the ellipsoid surface ∇g at point $\bar{p}$, where the equation for the ellipsoid is written as

$$g(x_1,x_2,x_3)=\left(\frac{x_1}{a_7}\right)^2+\left(\frac{x_2+a_5}{a_8}\right)^2+\left(\frac{x_3+a_4+a_6}{a_6}\right)^2=1,$$

and thus the normal vector is

$$\nabla g=2\left[\frac{x_1}{a_7^2}\ \frac{x_2+a_5}{a_8^2}\ \frac{x_3+a_4+a_6}{a_6}\right]^T,$$

where T is the transposition operator. We use ∇g to resolve $\bar{p}$ into its normal and tangential components. The mirror source about the tangential plane will have the same tangential component as $\bar{p}$, while its normal component will be opposite in direction. Thus

$$\bar{u}=\bar{p}-2\frac{\bar{p}^T\nabla g}{\|\nabla g\|^2}\nabla g.$$

Once the direction vector $\bar{u}$ is found, the source location can be obtained by noting that $\bar{s}=\bar{p}+\alpha\bar{u}$. To compute α we use the constraint that the range of the source is known. In this case we assume that all source locations are on the surface of a sphere with radius $r=1$ m (which is the case in our measurements). Thus, the constraint can be written as $\|\bar{p}+\bar{v}\|=\|\bar{p}+\alpha\bar{u}\|=1$, and the value of α is computed as the positive root of

$$\alpha=\frac{-\bar{p}^T\bar{u}\pm\sqrt{(\bar{p}^T\bar{u})^2-\|\bar{u}\|^2(\|\bar{p}\|^2-1)}}{\|\bar{u}\|^2}.$$

With values of vectors $\bar{p}$, $\bar{v}$, and $\bar{s}$, we can now compute the torso delay $D(\theta,\phi)$.

This procedure yields the values of the torso delay for source locations which do not lie on a regular spatial grid and that usually do not coincide with our measurement points. We solve this final problem by applying an interpolation procedure based on a spherical harmonic expansion.

[1]Only static localization cues are considered in this paper. Low-frequency dynamic cues are also important. Perrett and Noble (1997) verified Wallach's hypothesis that horizontal head rotation can be used to resolve front/back confusion as well as to determine the magnitude of the elevation angle. Moreover, they showed that this dynamic cue requires the presence of acoustic energy below 2 kHz. They observed in passing that, although horizontal head rotation cannot resolve an up/down ambiguity in elevation, their subjects were nonetheless able to tell if the source was above or below the horizontal plane; they speculated that spectral cues created by the shoulders and torso were responsible.

[2]Note that these angles are different from the angles in a conventional vertical–polar coordinate system. In particular, a surface of constant interaural–polar azimuth is a horizontal cone, while a surface of constant vertical–polar azimuth is a vertical plane. The advantages of interaural–polar coordinates were pointed out by Searle *et al.* (1976), and they have also been used by Morimoto and Aokata (1984) and by Middlebrooks (1999). However, these authors have named the angles differently. Morimoto and Aokata call 90°-θ the ''lateral angle'' and ϕ the ''rising angle,'' while Middlebrooks calls θ the ''lateral angle'' and ϕ the ''polar angle.'' At the risk of some confusion, we have chosen to retain conventional terminology.

[3]As expected, front/back confusion was greater for low-pass stimuli than for

full-bandwidth stimuli (see Carlile and Pralong, 1994). For some subjects, the location of the low-pass-filtered sound always appeared to be in the back.

[4]Some caution must be exercised in computing statistics for directional data because of the 360° ambiguity (Mardia, 1972) and the possibility of up/down as well as front/back confusion (Wenzel *et al.*, 1993). However, because we separated front and back stimuli, and because the reported data were confined to a semicircle, we computed the bias and rms error using the target and reported angles as if they were rectangular coordinates. In particular, the bias was computed as the average signed error, and the rms error as the square root of the average of the squared error. The probable presence of up/down confusion makes the resulting values a bit more pessimistic than necessary, but does not change the conclusion that the angular errors are large.

[5]Mathematically, the HRTF is defined as the ratio of two transfer functions, one from the source to the ear with the subject present, and the other from the source to the location of the center of the head under free-field conditions. For an infinitely distant source, these transfer functions become identical at very low frequencies, and the HRTF approaches 1 (unity DC gain). However, at close ranges, the inverse square law results in a higher DC gain for the ipsilateral ear and a lower DC gain for the contralateral ear (Duda and Martens, 1998). In our HAT model, these small differences are ignored.

[6]Avendano, Algazi, and Duda (1999) describe a more elaborate torso model in which the reflection coefficient varied with azimuth, elevation, and frequency. The torso model used in this paper seems to produce similar elevation perceptions, and was chosen for its simplicity.

Algazi, V. R., Avendano, C., and Thompson, D. (**1999**). ''Dependence of subject and measurement position in binaural signal acquisition,'' J. Audio Eng. Soc. **47**(11), 937–947.

Asano, F., Suzuki, Y., and Sone, T. (**1990**). ''Role of spectral cues in median plane localization,'' J. Acoust. Soc. Am. **88**, 159–168.

Avendano, C., Algazi, V. R., and Duda, R. O. (**1999**). ''A head-and-torso model for low-frequency binaural elevation effects,'' in Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, pp. 179–182.

Avendano, C., Duda, R. O., and Algazi, V. R. (**1999**). ''Modeling the contralateral HRTF,'' in *Proceedings of the AES 16th Conference on Spatial Sound Reproduction* (Rovaniemi, Finland), pp. 313–318.

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

Brown, C. P., and Duda, R. O. (**1998**). ''A structural model for binaural sound synthesis,'' IEEE Trans. Speech Audio Process. **6**(5), 476–488.

Buell, T. N., and Hafter, E. R. (**1988**). ''Discrimination of interaural differences of time in the envelopes of high-frequency signals: Integration times,'' J. Acoust. Soc. Am. **84**, 2063–2066.

Butler, R. A. (**1986**). ''The bandwidth effect on monaural and binaural localization,'' Hear. Res. **21**, 67–73.

Carlile, S., and Pralong, D. (**1994**). ''The location-dependent nature of perceptually salient features of the human head-related transfer functions,'' J. Acoust. Soc. Am. **95**, 3445–3459.

Carlile, S., editor (**1996**). *Virtual Auditory Space: Generation and Applications* (R. G. Landes, Austin, TX).

Cramer, H. (**1946**). *Mathematical Methods of Statistics* (Princeton University Press, Princeton, NJ), pp. 397–400.

Duda, R. O., and Martens, W. L. (**1998**). ''Range dependence of the response of a spherical head model,'' J. Acoust. Soc. Am. **104**, 3048–3058.

Duda, R. O., Avendano, C., and Algazi, V. R. (**1999**). ''An adaptable ellipsoidal head model for the interaural time difference,'' in Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing ICASSP'99, II-965–968.

Gardner, M. B., and Gardner, R. S. (**1973**). ''Problem of localization in the median plane: Effect of pinna cavity occlusion,'' J. Acoust. Soc. Am. **53**, 400–408.

Gardner, M. B. (**1973**). ''Some monaural and binaural facets of median plane localization,'' J. Acoust. Soc. Am. **54**, 1489–1495.

Genuit, K., and Platte, H. J. (**1981**). ''Untersuchungen zur Realisation einer richtungsgetreuen Übertragung mit elektroakustischen Mitteln (Investigations on the implementation of directionally faithful transmission by electroacoustical means),'' *Fortschritte der Akustik,* FASE/DAGA '81, Berlin (VDE-Verlag, Berlin), pp. 629–632.

Genuit, K. (**1984**). ''Ein Modell zur Beschreibung von Außenohrübertragungseigenschaften (A model for the description of the outer-ear transfer function),'' Doctoral dissertation, Dept. of Elec. Engr., Rheinisch-Westfälischen Technichen Hochschule Aachen, Aachen, Germany.

Hanson, W. W. (**1944**). ''The baffle effect of the human body on the response of a hearing aid,'' J. Acoust. Soc. Am. **16**, 60–62.

Kuhn, G. F. (**1977**). ''Model for the interaural time differences in the azimuthal plane,'' J. Acoust. Soc. Am. **62**, 157–167.

Kuhn, G. F., and Guernsey, R. M. (**1983**). ''Sound pressure distribution about the human head and torso,'' J. Acoust. Soc. Am. **73**, 95–105.

Kuhn, G. F. (**1987**). ''Physical acoustics and measurements pertaining to directional hearing,'' in *Directional Hearing*, edited by W. A. Yost and G. Gourevitch (Springer, New York), pp. 3–25.

Mardia, K. V. (**1972**). *Statistics of Directional Data* (Academic, London).

Middlebrooks, J. C., and Green, D. M. (**1991**). ''Sound localization by human listeners,'' Annu. Rev. Psychol. **42**, 135–159.

Middlebrooks, J. C. (**1999**). ''Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency,'' J. Acoust. Soc. Am. **106**, 1493–1510.

Møller, H. (**1992**). ''Fundamentals of binaural technology,'' Appl. Acoust. **36**(5), 171–218.

Morimoto, M., and Aokata, H. (**1984**). ''Localization cues of sound sources in the upper hemisphere,'' J. Acoust. Soc. Jpn. (E) **5**(3), 165–173.

Perrett, S., and Noble, W. (**1997**). ''The effect of head rotations on vertical plane sound localization,'' J. Acoust. Soc. Am. **102**, 2325–2332.

Roffler, S. K., and Butler, R. A. (**1967**). ''Factors that influence the localization of sound in the vertical plane,'' J. Acoust. Soc. Am. **43**, 1255–1259.

Searle, C. L., Braida, L. D., Davis, M. F., and Colburn, H. S. (**1976**). ''Model for auditory localization,'' J. Acoust. Soc. Am. **60**, 1164–1175.

Shaw, E. A. G. (**1997**). ''Acoustical features of the human external ear,'' in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Erlbaum, Mahwah, NJ), pp. 25–47.

Theile, G., and Spikofski, G. (**1982**). ''Die Bedeutung des menschlichen Rumpfes für die Lokalisation in der Medianbene (The importance of the human torso for localization in the median plane),'' *Fortschritte der Akustik*, FASE/DAGA '82, Göttingen (DPG-Verlag, Bad Honnef), pp. 1181–1186.

Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. (**1993**). ''Localization using nonindividualized head-related transfer functions,'' J. Acoust. Soc. Am. **94**, 111–123.

Wightman, F. L., and Kistler, D. L. (**1997**). ''Factors effecting the relative salience of sound localization cues,'' in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Erlbaum, Mahwah, NJ), pp. 1–23.