Fast head-related transfer function measurement via reciprocity

Dmitry N. Zotkin,[a] Ramani Duraiswami,[b] Elena Grassi,[c] and Nail A. Gumerov[d]
Perceptual Interfaces and Reality Laboratory, Institute for Advanced Computer Studies (UMIACS), University of Maryland at College Park, College Park, MD 20742

(Received 6 October 2004; revised 21 April 2006; accepted 3 May 2006)

An efficient method for head-related transfer function (HRTF) measurement is presented. By applying the acoustical principle of reciprocity, one can swap the speaker and the microphone positions in the traditional (direct) HRTF measurement setup, that is, insert a microspeaker into the subject's ear and position several microphones around the subject, enabling simultaneous HRTF acquisition at all microphone positions. The setup used for reciprocal HRTF measurement is described, and the obtained HRTFs are compared with the analytical solution for a sound-hard sphere and with KEMAR manikin HRTF obtained by the direct method. The reciprocally measured sphere HRTF agrees well with the analytical solution. The reciprocally measured and the directly measured KEMAR HRTFs are not exactly identical but agree well in spectrum shape and feature positions. To evaluate if the observed differences are significant, an auditory localization model based on work by J. C. Middlebrooks [J. Acoust. Soc. Am. **92**, 2607–2624 (1992)] was used to predict where a virtual sound source synthesized with the reciprocally measured HRTF would be localized if the directly measured HRTF were used for the localization. It was found that the predicted localization direction generally lies close to the measurement direction, indicating that the HRTFs obtained via the two methods are in good agreement.
 [DOI: 10.1121/1.2207578]

PACS number(s): 43.66.Qp, 43.66.Pn [AK] Pages: 2202–2215

I. INTRODUCTION

Spatial hearing is one of the fundamental sensory abilities encountered in the animal kingdom. Humans are very good at localizing acoustic sources in the environment, and some animals (such as the barn owl) are even better localizers. It is known (Hartmann, 1999) that various cues participate in spatial sound perception and that some of these cues are created by the process of scattering of the sound off the listeners themselves. Because of such scattering, the signals arriving at our ears are modified in a direction-dependent manner (Batteau, 1967; Wright *et al.*, 1974; Musicant and Butler, 1984).

The scattering process can be modeled as a linear filter applied to the sound emanating from the source. The transfer function, which when applied to the Fourier transform of the source sound converts it to the Fourier transform of the received sound, is called the "head-related transfer function (HRTF)." It captures the scattering behavior of the ear, head, and body of the listener for various locations at which acoustical sources may be. If the head is centered at a given point P and the sound source located at elevation θ, azimuth φ, and distance r in a head-centered spherical coordinate system, then the HRTF $H(r,\theta,\varphi,f)$ is the ratio of the Fourier transform of the signal at the eardrum $F_e(f)$ to the Fourier transform of the signal that would have been received at the point P in free-field $F_P(f)$, where f is the signal frequency. Thus the HRTF depends on four variables and must be measured as a function of these. For relatively distant sources, $r\rightarrow\infty$, the dependence on r is weak and is often ignored. In this case the HRTF may be characterized as a function of just three variables $H(\theta,\varphi,f)$. This assumption has shown to not hold for relatively nearby sources, and some researchers have measured the HRTF at different distances.

The head-related impulse response (HRIR) is the impulse response of the ear-head-body (i.e., the inverse Fourier transform of the HRTF). The HRIR (and therefore the HRTF) is traditionally measured by playing a broad-band test signal from various positions and recording the scattered wave field either at the entrance of the open or blocked ear canal or in the ear canal itself near the eardrum (Shaw and Teranishi, 1968). If the HRTF is treated as a four-dimensional function, then the source should be moved in three spatial dimensions, and the space of locations be sampled. If the dependence on range is neglected, then the source is placed at various locations on the surface of a sphere of radius a, and the HRTF is measured as a function of frequency and direction $H(\theta,\varphi,f)$. This method, characterized by the fact that the source is moved to various locations, is later referred to as the *direct HRTF measurement method*. In this paper, an alternative *reciprocal HRTF measurement method* is proposed, and comparison of the reciprocally measured HRTF with analytically derived HRTF for a sound-hard sphere and comparison between reciprocally and directly measured KEMAR manikin HRTFs are presented.

The paper is organized as follows. In Sec. II, the theoretical background for the method is discussed. In Sec. III, the proposed HRTF measurement method is described and its

[a]Electronic mail: dz@umiacs.umd.edu
[b]Electronic mail: ramani@umiacs.umd.edu Also at Department of Computer Science, University of Maryland, College Park, MD.
[c]Electronic mail: egrassi@umd.edu
[d]Electronic mail: gumerov@umiacs.umd.edu

advantages, limitations, and methods of overcoming these are discussed. In Sec. IV, the apparatus and the signal processing methods for the reciprocal HRTF measurement technique are described. In Sec. V, experimental results for a sphere (modeled using a bowling ball) and for the KEMAR manikin are presented. Section VI concludes the paper.

II. BACKGROUND

A. Existing HRTF measurement methods

Because of interpersonal differences in anatomy, the HRTF varies widely between individuals. Many studies have shown that with individualized HRTF virtual sound sources are indistinguishable from the real ones (Zahorik *et al.*, 1995; Hartmann and Wittenberg, 1996; Kulkarni and Colburn, 1998; Langendijk and Bronkhorst, 2000) and that the localization performance, particularly in the vertical dimension (in elevation), is degraded when a generic (nonindividualized) or distorted HRTF is used (Gardner and Gardner, 1973; Wenzel *et al.*, 1993).

Various approaches exist for HRTF personalization. Traditionally, the far-field ($r \to \infty$) HRTF for a given person is measured by placing miniature microphones or microphone probes into a person's ears and playing (from a relatively large distance) test sounds from each direction for which the HRTF is to be measured. Many papers are devoted to the topic of human HRTF measurement using the direct method (e.g., Shaw and Teranishi, 1968; Blauert, 1969; Mehrgardt and Mellert, 1977; Wightman and Kistler, 1989; Pralong and Carlile, 1994) or include HRTF measurement performed as a part of a larger experiment (Divenyi and Oliver, 1989; Bronkhorst, 1995). Among existing methods of HRTF personalization, the direct measurement procedure is the most accurate and is well developed. Various test signals, such as impulses, white noise, ML sequences (Schroeder, 1979), Golay codes (Zhou *et al.*, 1992), or in fact any broad-band signals with sufficient energy in the frequencies of interest (Grassi *et al.*, 2003), can be used for the measurement. Also, the measurement can be performed with either open or blocked (sealed) ear canals (Møller *et al.*, 1995; Carlile, 1996). In the first case, a miniature microphone or a probe microphone tip is placed inside the ear canal and the measurement includes the ear canal response; in the second case, the ear canal is sealed and a microphone is placed at the entrance to the ear canal, usually in the center of the sealing plug. With a blocked ear canal recording, also called "blocked-meatus" recording, the ear canal response is missing from the measured HRTF. However, in the most common scenario in which HRTFs are used—in a virtual auditory display where sound is delivered through headphones of sufficiently low acoustic impedance—the ear canal response is naturally reintroduced. Because of this reason and due to some dependence of the open-ear-canal measurement on precise microphone placement (Middlebrooks *et al.*, 1989), blocked ear canal measurement is usually performed. It has been shown to capture all location-dependent characteristics of the HRTF by Algazi *et al.* (1999).

In the experimental setup for the direct method numerous choices are involved regarding the type and the length of the signal (which determines the lowest reliably measured frequency and the required reverberant properties of the measurement space), the number of signal repetitions and the level of the room acoustic isolation [which determines the signal-to-noise ratio (SNR) and protects against outliers and incorrect HRTF measurement], the number of simultaneously mounted loudspeakers (which determines the amount of equipment reflections), the spatial HRTF sampling density (which obviously determines the measurement time period length), the equipment cost (which also influences SNR through the quality of the equipment and the availability of an acoustically treated space), and others. However, in any direct measurement setup it is necessary to move the sound emitter (usually, a loudspeaker) between measurement positions, and generally it takes from tens of minutes to 1–2 h to obtain the HRTF for about 300 to 1200 measurement directions. With conservative choices from the list above, HRTF measurement on a dense spatial grid in a regular room with 25 repetitions of the test pulse can take approximately 1.5 h for a single subject [as indicated by the authors of a carefully measured and validated extensive HRTF database (Algazi *et al.*, 2001c) in private communication].

When human listeners are tested in anechoic settings or in virtual presentation, the just noticeable differences (jnd's) in the azimuth vary from about 1° in the region in the front of the listener to 5°–10° to the extreme right and to the extreme left of the listener, whereas the jnd's in elevation are more signal dependent and can vary from 4° (white noise) to 17° (continuous speech by an unfamiliar person) (Blauert, 1997; Carlile *et al.*, 1997). These values and the fact that the precise locations of regions of higher sensitivity vary for subjects suggest that a fine HRTF sampling grid with a large number of points is necessary for applications where the location of sources must be presented with high fidelity. The use of multi-loudspeaker arrays, one per desired position, can introduce inter-reflections between the loudspeakers and contaminate the measurement signal. By using certain tradeoffs in the measurement process and by developing specialized equipment, different laboratories have made significant speed-ups in HRTF measurement. However, despite these advances, the direct measurement method has the fundamental limitation that measurements for different source positions must be made *sequentially* (i.e., the test sounds are emitted one after the other, with some waiting time in between, from the different positions where the HRTF sampling is to be performed).

In this paper, a novel HRTF measurement method based on the reciprocity principle is described and validated. It is based on a method first disclosed by Duraiswami and Gumerov (2003). The method retains most advantages of the direct method while having the potential to significantly decrease the HRTF acquisition time. Once the HRTF is measured with the reciprocal method, it can be used in the regular manner in place of the directly measured HRTF in any application (e.g., in virtual auditory displays).

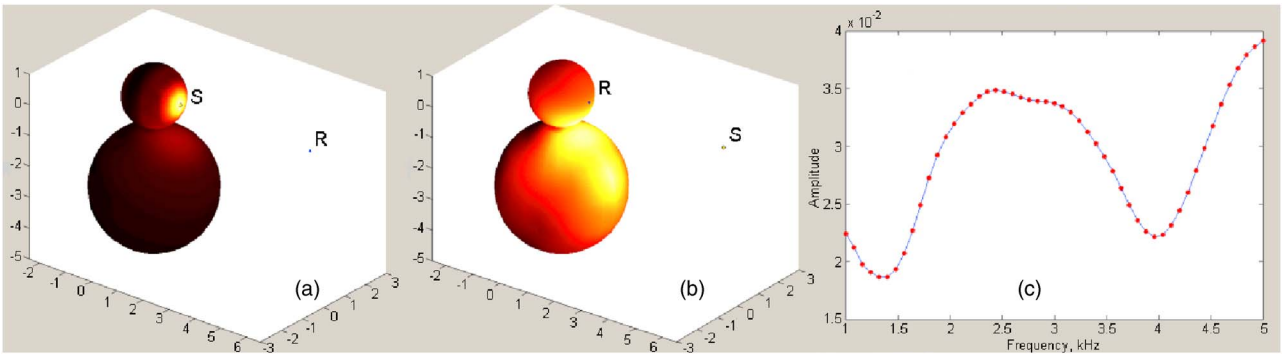

FIG. 1. (Color online) Reciprocity illustration (the acoustic fields are computed with numerical methods). (a) Case A: sound source S in the ear, receiver R at a distance of 4 m. (b) Case B: swapped source and receiver. (c) Dots: intensity versus frequency at the receiver for case (a). Solid line: intensity versus frequency at the receiver for case (b). While the fields in the domain in (a) and (b) are different, reciprocity holds between S and R, as shown in (c).

Other methods for the quick synthesis of a personalized HRTF include frequency scaling in accordance with positions of HRTF features, which were found to be highly correlated with the anthropometry of the subject (Middlebrooks, 1999a, b); use of a morphologically driven parametric HRTF model derived from a large HRTF database (Jin *et al.*, 2000); personalization based on the subjective experience of the user (Runkle *et al.*, 2000); or numerical modeling of acoustic wave propagation (Kahana and Nelson, 2000; Katz, 2001a, b) on a carefully measured mesh of the subject. These methods cannot yet yield the degree of personalization that is achievable when the HRTF is acquired via the direct measurement process.

B. The principle of reciprocity

The reciprocal HRTF measurement method is based on Helmholtz' principle of reciprocity, which states that in an arbitrary complex linear time-invariant acoustic scene "…the pressure at the measurement point **r**, caused by a source at $\mathbf{r}_0$, is equal to the pressure which would be measured at $\mathbf{r}_0$ if the source were placed at **r**" (Morse and Ingaard, 1968). The acoustic field is generally different elsewhere in the scene, but the signal picked up at the receiver is the same if the receiver and the source locations are interchanged. In Fig. 1, the reciprocity principle illustration computed by numerical methods is shown for a monopole (omnidirectional) source.

C. Analytical HRTF computation at low frequency

Accurate HRTF measurement at low frequencies remains a challenge for all measurement methods, mainly because (a) it is hard to produce significant low-frequency output using typical small loudspeakers, (b) a short signal must be used to window out wall and equipment reflections, and (c) the anechoic properties of the acoustic insulation degrade at low frequencies, leading to poor SNR and poorly attenuated wall reflections (Algazi *et al.*, 2001c). However, it is possible to compute the low-frequency HRTF analytically using a simplified model (Algazi *et al.*, 2002b).

The HRTF is a result of the scattering of the acoustic wave on the person's torso, head, and pinnae. These anatomical parts have different characteristic sizes and influence sound waves of wavelengths comparable to their dimensions. Accordingly, the HRTF can be roughly decomposed into parts influenced primarily by the head and torso and by the pinna (Algazi *et al.*, 2001a). Because of the small size of the pinna, the HRTF at low frequencies (longer wavelengths) is mainly a result of interaction of the sound wave with the torso and the head (Algazi *et al.*, 2001b), and this interaction can be well modeled analytically by approximating the head and torso by two spheres (HAT, head-and-torso, or snowman model). The model was first developed and verified by Algazi *et al.* (2002a) using numerical computational methods, and a simplified model, which is referred to as the HAT model, was proposed by Algazi *et al.* (2002b). By replacing the low-frequency part in the measured HRTF by the HAT model HRTF, compensation for the inaccuracies in HRTF measurement at low frequencies is performed.

The HAT model (Algazi *et al.*, 2002b) represents the low-frequency HRTF in terms of certain gross anatomical features of the listener using a "snowman" model. The model consists of two spheres of radii r_t and r_h, modeling the torso and the head, respectively, which are separated by a certain distance h_n, which models the neck height. The parameters r_t, r_h, and h_n are either measured on the subject directly, or are determined from a photograph of the subject that includes a scale, or are fitted to the absolute delays in the measured HRTF. Two pinnaless "ears" are located on the head of the "snowman" diametrically opposing each other on the interaural axis. Two different algorithms are used to model the sound propagation paths in the model and to compute the HAT model HRTF $H_h(f)$ depending on whether the source is located inside or outside of the torso shadow cone with respect to the given ear, as shown in Fig. 2. Sound from a source outside the torso shadow for the given ear arrives through the direct path and through the "shoulder bounce" path. Sound from a source in the torso shadow region is diffracted around the torso to reach the ear. In addition, if the arrival direction for the direct path, the shoulder bounce, or the around-the-torso path falls in the head shadow region, the head shadow is modeled. The HRTF synthesized by the HAT model is minimum phase.

Cross-fading is used on log-magnitudes of the HAT model HRTF $H_h(f)$ and the measured HRTF $H_m(f)$ to obtain the combined HRTF $H_c(f)$:

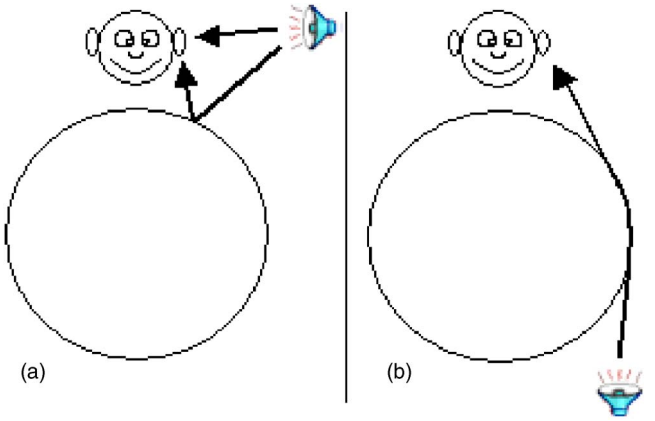

FIG. 2. (Color online) Propagation paths modeled by HAT model: (a) in the case of the source being out of the torso shadow region; (b) in the case of the source being in the torso shadow region.

$$A_c(f) = \begin{cases} A_h(f), & f < f_1, \\ A_h(f) + \dfrac{A_m(f) - A_h(f)}{f_2 - f_1}(f - f_1), & f_1 < f < f_2, \\ A_m(f), & f > f_2, \end{cases}$$

$$A_m(f) = \log|H_m(f)|, \quad A_h(f) = \log|H_h(f)|,$$

$$A_c(f) = \log|H_c(f)|. \tag{1}$$

Thus, the HAT model HRTF magnitude is used for frequencies below f_1; the measured HRTF magnitude is progressively blended in for frequencies from f_1 to f_2; and the measured HRTF magnitude is used for frequencies above f_2. Finally, the phase of the combined HRTF $H_c(f)$ is set to the phase of $H_h(f)$ at all frequencies. Such an approach to modeling HRTF phase is also justified by the results of Kulkarni *et al.* (1999), where it was shown that human listeners are sensitive almost exclusively to the HRTF magnitude.

III. PROPOSED APPROACH

A. Description and advantages

A new method for HRTF measurement is proposed here. It is referred to as the reciprocal method in this paper. The method is based on the reciprocity principle, which has already been used in numerical simulations of problems related to spatial hearing (Kahana, 2000). Application of reciprocity to the HRTF acquisition problem reveals an elegant modification to the HRTF measurement setup, which is to literally exchange the loudspeaker and the microphone, that is, to put a (miniature) loudspeaker in the person's ear and the microphones at the positions where HRTF is to be measured. The reciprocity principle implies that if the loudspeaker and microphone positions were exchanged exactly and no other changes were made in the setup, the measurements obtained would be identical to the ones obtained with the direct method. Inevitable slight microphone, loudspeaker, and subject positioning differences imply that a perfect match between recordings is unlikely to be achieved. Still, according to the reciprocity principle, the HRTF obtained by the reciprocal method should be in reasonable agreement with that measured with the direct method.

The reciprocal measurement method has several advantages over the direct method. First, because in practice microphones are much smaller than loudspeakers, it is possible to build an array of microphones around the person and have one microphone per HRTF measurement direction without artifacts arising due to interequipment reflections. (Arguably, if one uses microloudspeakers instead of regular ones, one can also mount as many of them as the number of the measurement directions without causing significant reflections. However, tiny speakers are expensive, and the measurement process is still sequential, as discussed below.) Second, HRTF sampling for many positions can be done *in parallel* by playing the test sound using an in-ear loudspeaker and recording the received sound simultaneously at all microphones. Thus, the HRTF for all recording positions is acquired at the same time (for one ear), unlike in the direct method. In essence, with an appropriate number of microphones, HRTF acquisition (for two ears) over the whole sphere of directions with the reciprocal method can be done in the same time that it takes for the direct method to perform HRTF acquisition at two positions. In principle, all technological and signal processing innovations that have been used to speed up direct measurements can also be employed in the reciprocal method. Finally, sampling of the HRTF at an additional set of directions can be quickly performed by rotating either the person or the measurement microphone array to a nonoverlapping configuration and performing another pair of measurements.

B. Issues that must be addressed in the reciprocal method

A possible disadvantage of the reciprocal HRTF measurement method is a narrower effective frequency band. The loudspeaker used for reciprocal measurement must obviously fit into the ear canal opening. Therefore, it must be physically small, leading to possibly poor low-frequency output. However, as mentioned above, measurement of the HRTF at low frequencies is always problematic, so this issue is not unique to the reciprocal method, though it is more significant for it. A solution to the problem is offered by augmentation of the measured HRTF by an analytical solution at low frequencies. Such an analytical solution may be obtained by using the HAT model discussed earlier. In fact, the HAT model is a simplified version of the low-frequency HRTF computation technique based on numerical methods, which was developed by Algazi *et al.* (2002a) and could be used as an alternative. No differences in localization performance were observed by Algazi *et al.* (2001b) with the bandlimited (up to 3 kHz) sound source signal when the measured HRTF was replaced by the analytically computed HRTF based only on the spherical-head shadow and torso reflection.

Another possible concern is the safety of the sound level of the in-ear speaker. It is necessary to keep the sound volume at a comfortable and safe level for the subject. An obvious solution is to provide acoustic insulation between the speaker and the eardrum. In the setup used in the experi-

ments, a plug made from a soft material (a silicone compound commonly used by swimmers to seal their ears) is used. The plug is made by fully wrapping the microspeaker in the silicone so that all surfaces of the microspeaker, except the frontal surface where the sound-emitting aperture is located, are covered. In this way, the back (eardrum-facing) side of the microspeaker, where two thin wires are attached, is isolated from the eardrum by the silicone layer, and the plug simultaneously performs the functions of acoustic insulation, blocking the ear canal to achieve a blocked-meatus configuration, and holding the speaker in place.

In human HRTF measurement, it would be also advisable for subject comfort to increase the test signal volume gradually during the first few seconds of the signal to allow gradual habituation of the middle ear mechanisms to the signal level (in particular, adaptation of the amplification controlled by contraction of the stapedius muscle). However, even with such gradual increase the sound volume can be raised no further than certain safety (and comfort) limit for the participant. Because of such a limit on the sound level, a lower SNR than the one achieved in the direct method may be expected. The contralateral side HRTF directions are particularly problematic, and more signal repetitions can be necessary to achieve results of good quality. To identify a safe sound level for the reciprocal method in human subjects, experiments to compare eardrum sound levels observed in the direct and in the reciprocal HRTF measurement methods were carried out using the KEMAR manikin. These results are reported in Sec. V F.

IV. EXPERIMENTAL SETUP

To verify the feasibility of the proposed reciprocal method, a set of experiments corresponding to those reported with the direct method in the literature was carried out, and the results are described in Sec. V. The new measurement method was first evaluated on a sound-hard sphere (bowling ball) and the results compared against the analytical solution obtained by Lord Rayleigh (Strutt, 1904), modified for finite-distance sources by Rabinowitz *et al.* (1993), and presented recently together with experimental verification by Duda and Martens (1998), who also used a bowling ball as a sound-hard sphere model. Then, the HRTF of the KEMAR manikin was measured using both the direct and the reciprocal HRTF measurement methods, and measurements were compared to each other with the help of a sound localization model (based on Middlebrooks, 1992). The experimental setup and experimental methods used to obtain these measurements are described below.

A. Reciprocal method apparatus

The setup used for validation consisted of a spherical mesh structure constructed with parts from the ZomeTool construction kit, which is sold as a toy (shown in Fig. 3). The mesh was made up of struts (sticks) and nodes (balls) as shown in the picture. The total number of nodes in the constructed spherical structure of radius 0.70 m is 131. Thirty-two microphones were mounted at the nodes of the structure in a symmetrical and roughly equispaced manner. (More microphones could have been attached at intermediate locations; the limiting factor in the experiment was the acquisition hardware available at the time the experiments were conducted, and in principle there is no limitation on the number of channels that can be acquired simultaneously.) The microphones were affixed with scotch tape to short struts, which in turn were attached to the nodes. The bottom part of the mesh was left open in order to accommodate a subject. The spherical mesh was suspended from the ceiling with several wires in a large office room ($4.7\times5.6\times2.7$ m^3) with acoustically untreated walls. The mesh was positioned at the center of the room to maximally delay the arrival of interfering wall reflections. The earliest reverberant reflections arrived well after the head response time and were windowed out in the data processing step. The measurements were done mostly in the evening to ensure that noise from neighboring offices was minimized. In addition, the computer used for measurements was placed outside the room during the recording with the microphone connecting cables running under the door, other computers in the room were turned off, and incandescent lighting was used instead of fluorescent lights.

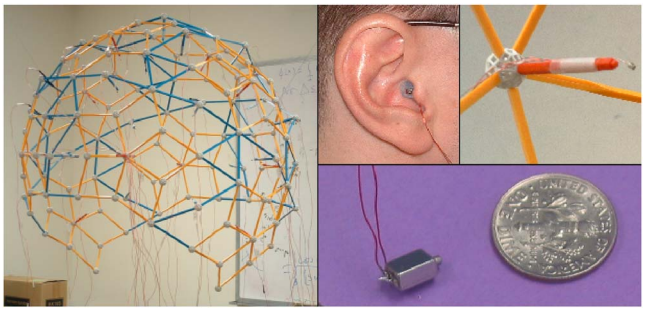

FIG. 3. (Color online) Left: The measurement mesh. Bottom right: The microspeaker. Top middle: The microspeaker inserted into the person's ear. Top right: An enlargement of the one node of the measurement mesh.

To minimize the load on the constructed spherical mesh, thin (gauge 30) enamel-coated wires were used to connect the microphones to two custom-made preamplification cards, each handling 16 channels. The outputs of the preamplifiers were connected to two National Instruments PCI-6071E data acquisition boards plugged into a Pentium Xeon 1.7 GHz Dell Dimension 8100 PC running Windows XP.

The microphone array was calibrated with an active tracking device (Polhemus 3D FASTRAK system). The tracking device has the ability to determine the three-dimensional coordinates of an associated receiver (a small sensor measuring about $12\times12\times6$ mm^3) with respect to a transmitter (a source box measuring about $51\times51\times51$ mm^3, with one corner marked as the origin). To perform the calibration, the transmitter was positioned so that its origin coincided with the center of the measurement sphere and leveled. Then, the coordinates of each of the 32 microphones were measured by putting the sensor next to the corresponding microphone and recording the observed sensor coordinates. The azimuth and the elevation of all microphones with respect to the center of the sphere were then computed from the obtained measurements. The coordinate system adopted for use with the reciprocal measurement pro-

cedure was a standard vertical-polar coordinate system with the azimuth varying from −180° to 180°, with 0° azimuth being in front of the subject and positive values going to the right, and with the elevation varying from −90° to 90°, with 0° elevation being in front of the subject and positive values going up.

B. Direct method apparatus (I)

Two methods were used for direct HRTF acquisition. The first setup was obtained by exchanging the microspeaker and the microphone positions in the reciprocal HRTF measurement setup described above, so that the microphone was inserted in the blocked ear canal of the subject (or in the hole of the sphere) as it is normally done in the direct measurement method, and the microspeaker was attached to the strut that was placed at the node of the spherical mesh structure. The advantage of this method is that the comparison of the direct and the reciprocal measurements performed at the same distance and with the same physical setup can be made. However, the microspeaker needed to be moved manually between measurement positions, slowing down the procedure significantly, and the number of measurement positions was limited in any case by the grid structure. Accordingly, the direct HRTF measurement in this setup was performed only at 32 positions (the same 32 positions at which the reciprocal HRTF was sampled). The HRTF measured in this setup is referred to as a sparse-grid direct HRTF. These measurements were performed at the same distance (0.70 m) as the reciprocal ones.

C. Direct method apparatus (II)

A direct (traditional) HRTF measurement apparatus was also used to obtain the directly measured HRTF with higher spatial resolution, which could not be done on the spherical mesh due to equipment and time constraints. As the focus of the current paper is on the validation of the reciprocal HRTF measurement method, the direct measurement setup and the corresponding measurement procedures are described only briefly here. A more detailed description may be found in Grassi *et al.* (2003). In the direct measurement, a set of loudspeakers (8-Ω Realistic 3 in. midrange tweeter, 700–20 000 Hz), mounted on a semicircular hoop rotating around the horizontal axis, emitted the acoustic signals used in the measurements. The radius of the hoop was 0.90 m, and the HRTF measurement distance (from the speaker membrane to the center of the hoop where the subject is placed) was 0.84 m. To avoid excessive interequipment reflections, only six loudspeakers were placed simultaneously on the hoop and recordings were taken in several sets to cover all desired azimuths. For each configuration of loudspeakers, the hoop stepped through all requested elevations, automatically controlled by the computer. Data collection was performed in a regular-sized office room ($3.2\times3.8\times2.4$ m^3). The room walls were coated with dispersive foam (4.5 cm egg crate foam) to dampen sound reflections. The HRTF was measured at 1149 directions covering the full sphere without the bottom part in approximately 5° steps, forming a dense-grid direct HRTF. The exact arrangement of directions in the direct grid can be found in Grassi *et al.* (2003). For most directions (elevation-azimuth pairs) in the reciprocity grid there was no exactly corresponding direction in the dense direct grid. However, the nearest direction was within 1.6° on average over all 32 reciprocal measurement directions and within 5° at most.

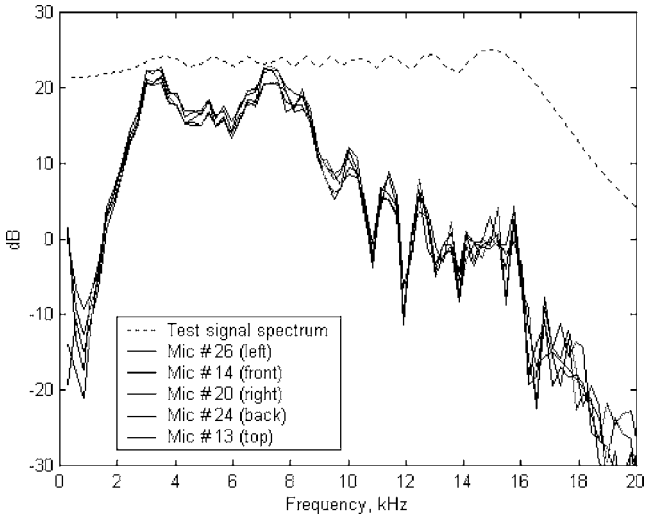

FIG. 4. Comparison between spectra of the test signal recorded at five microphones on the reciprocal measurement mesh.

D. Microphones and microspeaker

The microphones used for the measurements of the HRTF in the reciprocal setup were of type Knowles Electronic FG-3629 (omnidirectional microphone encased in a cylindrical case with a diameter of 2.57 mm and a height of 2.57 mm). The microspeaker used was of type Knowles Electronics ED-9689 (physical dimensions $6.33\times4.31\times2.97$ mm^3). The microspeaker is shown in the bottom right part of Fig. 3. An important question for the HRTF measurement is the microspeaker directivity pattern. An experiment was therefore performed to determine if the microspeaker behaves sufficiently closely to an omnidirectional point source. In the experiment, the same test signal that was later used for reciprocal HRTF measurement was employed. The signal was a frequency sweep with a roughly flat spectrum from 1 kHz up to 16 kHz. The upper frequency limit was chosen to be consistent with upper frequency HRTF measurement limit used by other researchers (Pralong and Carlile, 1994; Bronkhorst, 1995; Langendijk and Bronkhorst, 2000). The microspeaker was placed at the center of the reciprocity measurement apparatus with the speaker opening facing the "front" microphone (the microphone that was located in the front of the subject during HRTF measurement), the test signal was played through the speaker, and the recording was performed simultaneously at 32 microphones. In Fig. 4, the recorded signal spectra are shown for five microphones ("front," "left," "right," "back," and "top") along with the spectrum of the test signal. It can be seen that the microspeaker put out a smaller amount of energy at frequencies above 9 kHz, but the output was sufficient through the frequency band of interest. More importantly, the recordings were close to identical at all five microphones, showing the

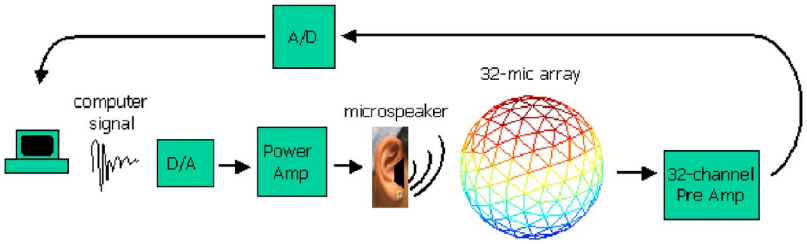

FIG. 5. (Color online) A schematic of the system for the reciprocal HRTF measurement.

omnidirectionality of the microspeaker. Therefore, no equalization to account for a speaker beampattern was necessary.

E. Signal acquisition

The schematic of the experiment and associated signal processing in the reciprocal HRTF measurement is depicted in Fig. 5. The excitation signal was generated in MATLAB, output through the D/A port of a National Instruments PCI-6071E card, power amplified, and played via the microphone inserted in the manikin ear or in a hole in the sound-hard sphere (bowling ball). The manikin/sphere was placed at the center of the spherical mesh. The received signal data were collected simultaneously from all 32 microphones in the spherical mesh. The signal recorded by the microphones was decoupled, amplified, and low-pass filtered with custom preamplifiers (fourth-order Bessel filter, cutoff frequency of 18 kHz, gain 100) to avoid aliasing. Signals were sampled at a rate of 39.0625 kHz and were acquired through the A/D port of the NI card. The D/A channel, which drives the microspeaker, and all A/D channels, which receive input from microphones, were triggered by a common hardware trigger and were running off a common clock source, ensuring that the synchrony necessary for time averaging is achieved.

F. Signal processing and HRTF computation

The test signal for the reciprocal HRTF measurement was a 96-sample-long (2.45 ms at 39.0625 kHz) linear frequency upsweep pulse with a maximum frequency of 16 kHz. The pulse duration was chosen to prevent an overlap of reflections from the experimental equipment and surrounding walls. One test session consisted of an upsweep signal repeated 48 times with a 420-ms pause between repetitions (dictated by an experimentally determined time for reverberant echoes to die out) for a total recording length of about 20 s. At the end of each test session, the manikin/sphere was removed from the spherical mesh and the microspeaker was placed at the center of the mesh on a thin stick in order to measure the reference (calibration) signal for each microphone. In this way any mismatch between microphones or preamplifier channels was captured in the reference signal and was removed during normalization.

The signal processing for the reciprocal HRTF measurement was performed as follows. The raw recorded pulses were averaged coherently in time to decrease the random noise first (the noise magnitude decreases proportionally to the number of pulses). Then the start of the pulse was detected using thresholding and a time window of 144 subsequent samples (3.7 ms) was kept. The measurement window was thus about 1.25 ms longer than the length of the pulse so that additional time was allocated after the end of the pulse to accommodate the system impulse response. The window length was chosen with knowledge of the room geometry and the relative placement of the measurement apparatus to ensure that all wall reflections arrive after the end of the measurement window and thus are discarded.

The HRTF magnitude for a particular direction (θ,φ) was then computed by taking the ratio of the Fourier transform of the signal acquired from the given microphone to the Fourier transform of the reference signal for the same microphone, i.e., $H(\theta,\varphi,f)=F(y(\theta,\varphi,t))/F(u(\theta,\varphi,t))$, where F is a discrete-time Fourier transform operator (i.e., a 144-point FFT), $y(\theta,\varphi,t)$ is the time-averaged pulse recorded at the microphone with the elevation θ and azimuth φ, and $u(\theta,\varphi,t)$ is the time-averaged reference pulse for the same microphone. A resolution of about 270 Hz was thus obtained in the resultant HRTF. Additional frequency smoothing was then performed on the magnitude of the obtained $H(\theta,\varphi,f)$ using a second-order Butterworth filter with cutoff $0.5\times$ normalized frequency.

Sparse-grid direct HRTF acquisition was performed with exactly the same test signal and signal processing procedures as described here for the reciprocal HRTF acquisition. For the dense-grid direct HRTF acquisition, a similar test signal was used and a similar experimental procedure was followed, with the pulse length being 150 samples (1.8 ms at 83.333 kHz) and the window length being 225 samples (2.7 ms). Smoothing was also performed similarly, but with a different normalized cutoff frequency so that the same frequency resolution was achieved in the final computed HRTF.

G. Sphere

A bowling ball was used as a physical model of a sound-hard sphere. The ball (polyester Ebonite Mirage) had a diameter of 21.84 cm, weighed 5.45 kg, and had a hard-coated surface with no finger holes. A 9.5-mm-diameter hole was drilled in the ball to accommodate the microspeaker for the reciprocal measurement experiment. No more holes were drilled. For the reciprocal HRTF measurement experiment, the microspeaker was inserted into the hole together with the silicone plug holding it in place so that the microspeaker was centered in the hole with the sound-emitting aperture facing outwards. The silicone plug surface and the microspeaker in the plug were both flush with the sphere surface.

The sphere was placed at the center of the reciprocal HRTF measurement mesh on a photographic tripod and was centered within the mesh using the mesh structure itself as a visual guide, as described below. To minimize tripod reflec-

tions, the tripod legs were covered with a soft fabric, and all tripod handles located just below the sphere were removed.

H. KEMAR manikin

An object traditionally used as a reference for HRTF measurement is the Knowles Electronics Manikin for Acoustic Research (KEMAR) (Burkhard and Sachs, 1975). Measurements of the HRTF for KEMAR have been performed by several researchers (e.g., Gardner and Martin, 1995; Algazi *et al.*, 2001c) and are widely available to the research community on the Internet. The KEMAR model used in the experiments described in this paper was a DB-4004, configured with two neck rings and a torso. Pinnae used were of type DB-060 and DB-061 (left and right "small" pinna, respectively). They were mounted on the DB-050 ear canal extensions with DB-100 occluded ear simulators (Zwislocki couplers). In both ears, a microphone (Knowles FG-3629) was placed into the opening of the Zwislocki coupler and sealed with the silicone. The signals recorded at these microphones were used to measure "eardrum" sound pressure levels during the reciprocal HRTF measurement.

The direct measurement of the KEMAR HRTF was done in the blocked-meatus setup. The microphone (Knowles FG-3629) was wrapped in a silicone plug and was inserted into the ear canal so that the ear canal was sealed and the microphone was located in the center of the seal, flush with the seal surface. The microspeaker insertion for the reciprocal measurement was done in the same manner. The KEMAR was mounted on a long metal pipe that was screwed into the heavy metal stand at the bottom end and into the mounting aperture on the KEMAR bottom plate at the top end so that any interference created by the mounting hardware was negligible.

I. KEMAR alignment procedures

In the direct measurement apparatus, the following procedure was used to center the KEMAR manikin in the measurement hoop. Two calibrated lasers were placed at the pivotal points of the hoop so that the beam of each laser was pointing towards the opposite pivotal point within a few millimeters of the other laser's aperture. A third laser was placed on the wall in front of the setup and was calibrated so that it was pointing to the 0° azimuth mark of the measurement hoop both when the hoop was at 0° elevation (in front of the subject) and at 180° elevation (at the back of the subject). The fourth laser, placed on the ceiling, projected a beam vertically downwards and was calibrated to point to the premarked center of the setup (the location midway between the hoop pivotal points) on the ground and to the 0° azimuth mark on the hoop when the hoop was at 90° elevation (up). The KEMAR was placed in the setup so that the beams of lasers located at pivotal points produced spots at the ear canal openings and was further aligned so that the spot from the frontal laser was located on the tip of the KEMAR's nose and the spot from the overhead laser was projected on the center of the cross etched on the head of the KEMAR.

In the reciprocal setup, a natural coordinate system for alignment was provided by the regular structure of the mesh itself. Four microphones in the mesh were located directly in the front, in the back, on the left, and on the right of the subject. They were mounted at the centers of four crosses so that it was possible to align the manikin in the center of the mesh in the same manner as for the direct measurement by visual inspection. For example, when viewing the mesh from the front (with the KEMAR inside), the front microphone vertical line of mounting, the KEMAR nose, and the back microphone vertical line of mounting should all lie in the same plane. This condition was easy to check visually. In the same manner the left and right microphones were aligned with the KEMAR ear canals and the height of the mounting was adjusted so that the tip of the manikin's nose was in the equatorial plane of the measurement mesh. The alignment achieved was verified for each measurement set to ensure no deviations from centered position.

J. Augmentation by analytical solution

The validity of measured HRTF is limited approximately to the range from 1.5 to 16 kHz, determined by the frequency limitations of the microspeaker used in the experiments. Therefore, it is necessary to extrapolate the HRTF to reasonable values outside this range (especially at low frequencies) to make them suitable for sound rendering. The following approximations are applied:

(i) The high-frequency end is tapered to zero using a half Blackman window, starting at 16 kHz and reaching zero at 22.05 kHz.
(ii) At low frequencies the HRTF is approximated by the analytical solution obtained for the head-and-torso (HAT) model described in Sec. II C [Eq. (1)] with cutoff frequencies f_1 and f_2 being 1 and 3 kHz, respectively.

Augmentation of the measured HRTF with a HAT model compensates for the weak low-frequency response of the microspeaker and is necessary for accurate reproduction of virtual auditory scenery as most of the signals to be reproduced (e.g., speech or music) contain significant energy at lower frequencies.

V. RESULTS AND DISCUSSION

A. SNR estimation

The reciprocal experiments were conducted in the large office room described above. The SNR was estimated for the reference signal obtained by putting the microspeaker at the center of the reciprocity measurement mesh and was found to be 23.7 dB. After averaging with 48 pulses, the SNR increased to 37.4 dB. As a reference, the SNR achieved in the direct measurement setup was 32.5 dB and was improved to 41.5 dB by averaging.

B. Sphere

The plots presented in Fig. 6 show the results of the reciprocal HRTF measurement of the sphere together with the analytical solution for the sphere (Rabinowitz *et al.*,

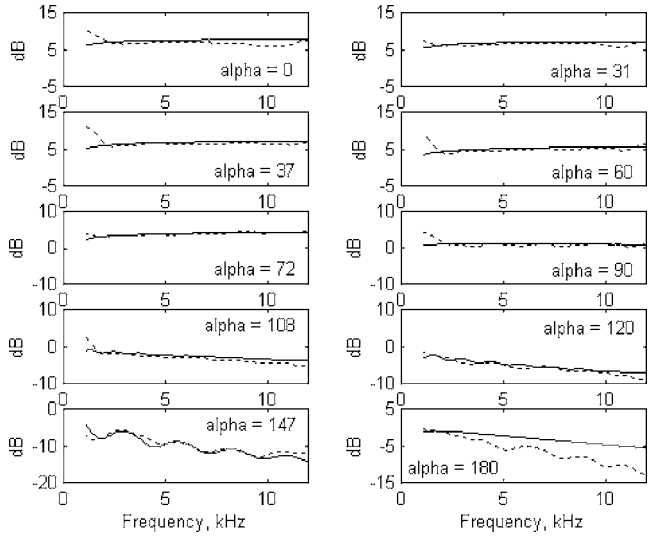

FIG. 6. Comparison between the analytically computed HRFT (solid line) and reciprocally measured HRFT (dotted line) for the sphere. Alpha is the incidence angle.

1993) for ten incident angles (indicated in the plots). The solid line represents the analytical HRTF solution, and the dotted line shows the measured HRTF.

It can be deduced from the plots that the measured HRTF matches the analytical solution with good precision above approximately 1.5 kHz, as can be expected based on the microspeaker effective frequency band. The 6-dB gain at normal incidence is well reproduced, suggesting that the microspeaker is sufficiently small to avoid disturbing the sound field and that the silicone plug surface acts as a sound-hard material (the gain is as expected a little more than 6 dB because the source distance is 0.70 m instead of infinity), and even for the weakest signal presented in the plots (at 147°), the characteristic ridges in the analytical solution are followed very well by the experimental HRTF. There are also no systematic differences between the analytically computed HRTF and the measured HRTF at most positions. However, a significant difference is observed at high frequencies for the 180° incidence angle. The ridge pattern exhibited by the measured HRTF suggests a small angular error in the positioning of the sphere, which brings the measurement position out of the bright spot. All other discrepancies are limited by approximately 2 dB and are similar to the experimental errors observed by Duda and Martens (1998) in the comparison between the sphere HRTF (measured with the direct method) and the analytical HRTF expression for the source at various distances.

C. KEMAR manikin

The second group of plots shown in Fig. 7 includes 32 pairs of direct and reciprocal HRTF measurements of the KEMAR manikin (left ear). The reciprocal measurement was done using the reciprocal method apparatus described above, and all available data (i.e., 32 measurement positions) from one measurement set are presented. The direct measurement was performed in the sparse-grid direct setup described above so that the results presented in this figure were obtained literally by application of the reciprocity principle without changing anything else in the acoustic scene, in the test signal, and in the signal processing. No additional HRTF smoothing was performed beyond the light smoothing described in Sec. IV F. Annotations in each plot show the (elevation, azimuth) pair of the measurement position.

The direct and the reciprocal measurements presented in the plots were done on two different days, and the KEMAR manikin was moved to the different room and then brought back and recentered in the spherical mesh for the second experiment. Nevertheless, the agreement between measurements is quite good within the effective frequency band of the reciprocal HRTF measurement. Positions of HRTF features (such as peak and notches) are matched well in all plots, although at some positions the notch depth differs between the direct and the reciprocal measurement. Also, a level disagreement is observed at the contralateral (0, 90) ("bright spot") position. As in the sphere case, the reason for

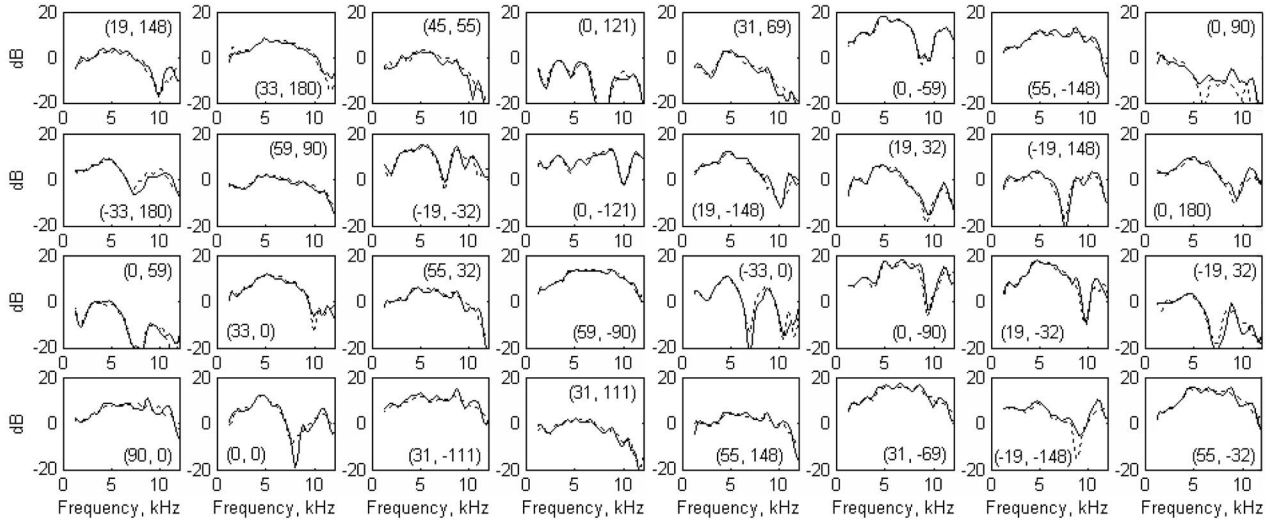

FIG. 7. Comparison between KEMAR left ear HRTF measurements using the direct method at 0.70 m (solid line) and the reciprocal method at 0.70 m (dotted line). The measurement directions are annotated on the plots as (elevation, azimuth) pairs.

the disagreement is likely to be a small positioning error, which moves the measurement position out of the bright spot.

D. Sound localization model

To further validate the reciprocal measurement technique, a simulated sound localization experiment was performed with the help of a localization model that uses both temporal and spectral cues and is based on the narrow-band sound localization model proposed by Middlebrooks (1992). It was chosen to modify Middlebrooks (1992) model somewhat by using the rms difference between HRTF magnitude spectra as the spectral similarity metric, which was used more recently by other researchers to quantify differences between HRTFs (Blommer and Wakefield, 1997; Kulkarni and Colburn, 2004), and by omitting the ILD-based similarity term (due to the unavailability of right ear reciprocal HRTF data). In the experiment, the HRTF obtained with the direct method was taken as the baseline HRTF. Then, for the HRTF measured reciprocally at the direction p, the degree of similarity to the baseline HRTF for each direction q in the baseline HRTF was computed according to the localization model, and the similarity-maximizing direction q' was chosen, corresponding to the sound being "localized" at the direction q'. The validation goal was to have q' at or sufficiently close to p for all p comprising the reciprocally measured HRTF.

Ignoring the range dependence of the HRTF, let the reciprocal HRTF for the direction p be $H_p=H(\theta_p,\varphi_p,f)$ and the direct HRTF for the direction q be $\hat{H}_q=\hat{H}(\theta_q,\varphi_q,f)$. The spectral similarity coefficient $b_{pq}^{(s)}$ was computed for the pair $(H_p,\hat{H}_q)$ as

$$b_{pq}^{(s)}(H_p,\hat{H}_q)=\left[\frac{1}{2\pi(f_2-f_1)}\int_{2\pi f_1}^{2\pi f_2}(20\log_{10}|H(\theta_p,\varphi_p,f)| - 20\log_{10}|\hat{H}(\theta_q,\varphi_q,f)|)^2\,df\right]^{1/2} \quad (2)$$

in the continuous frequency case, or the same with the integral replaced by a summation in the discrete frequency case. For a given reciprocal measurement direction p, the coefficients $b_{pq}^{(s)}$ for all direct measurement directions q were computed. Then, they were normalized to have a zero mean and unit variance, as suggested by Middlebrooks (1992).

It was also desirable to include temporal information in the HRTF comparisons, at least to compensate for the missing ILD term. The natural choice for the temporal cue would be the interaural time difference (ITD). However, ITD information was not immediately available in the reported measurements because the reciprocal KEMAR HRTF measurement was performed only for the left ear. Therefore, an estimation of ITD for reciprocal measurement was constructed by utilizing the symmetry of the setup. The ITD for a given direction $p=(\theta,\varphi)$ is usually computed as the difference in time of arrival (TOA) of the signal to the left and to the right ears. If the subject and the setup are sufficiently symmetric, then the TOA for the right ear for the direction p should be the same as the TOA for the left ear for the direction $p^*=(\theta,-\varphi)$. The ITD τ_p for the reciprocal measurement was therefore constructed as $\tau_p=t_{p^*}-t_p$, where t_p and t_{p^*} were the TOAs (for the left ear) of the signal for the directions p and p^*, respectively. The actual values of the TOAs were determined from the experimental data via thresholding.

Because of the symmetric arrangement of the measurement directions, the measurement grid included the direction p^* for 31 of 32 directions p [e.g., for $(-19,-32)$ the symmetric direction was $(-19,32)$, and some directions such as (33, 180) were symmetric pairs to themselves so the ITD there was forced to be zero]. Only one direction (45, 55) did not have a symmetric counterpart; therefore, t_{p^*} for it was computed as average TOA for two directions closest to $(45, -55)$, which were $(31,-69)$ and $(55,-32)$.

The ITDs $\hat{\tau}_q$ for all positions q in the direct setup were computed using data from both ears (as both sparse-grid and dense-grid direct HRTF measurements were performed on both ears). Temporal similarity coefficients $b_{pq}^{(t)}$ were then computed to be simply

$$b_{pq}^{(t)}=|\tau_p-\hat{\tau}_q|. \quad (3)$$

The coefficients $b_{pq}^{(t)}$ were computed for all directions q and then were normalized to have zero mean and unit variance. After extracting the pure time delay, the remaining phase information in the HRTF was not used for matching. This decision was based on the observation by Kulkarni *et al.* (1999) that humans are largely insensitive to the HRTF phase. Then, finally, the direction $q'=\arg\min_q(b_{pq}^{(s)}+b_{pq}^{(t)})$ was taken to be the localization direction predicted by the model. If q' is at or close to p, then it can be concluded that the reciprocal measurement method provides sufficiently feasible HRTF.

E. Analysis of KEMAR manikin results using the auditory model

The sound localization model was used first with the sparse-grid direct left ear HRTF as the baseline HRTF and with $f_1=1.5$ kHz, $f_2=16$ kHz, which was the validity range of reciprocal HRTF measurement based on the microspeaker characteristics and on the test signal used (see Sec. IV D). For each direction p in the reciprocally measured left ear HRTF, the model was applied to predict the localization direction q' for the sound if it were filtered with the reciprocal HRTF for the direction p. It was found that the matching obtained was, in fact, perfect, and that the model output q' was equal to p for all 32 directions p. For reference, in Table I the values of the spectral similarity coefficient (prenormalization) between reciprocal and direct measurements taken at the same direction p (in other words, rms difference between magnitude spectra of HRTF taken by direct and by reciprocal method) are shown for all p. This table essentially quantifies the data shown in Fig. 7.

However, as the baseline HRTF in this case was sampled only at 32 directions, the perfect matching was perhaps not surprising. For further testing, the sound localization model was applied using the dense-grid direct left ear HRTF as the baseline. For each direction p of the reciprocal HRTF mea-

TABLE I. Root-mean-square difference (RMSD) between the sparse-grid direct and the reciprocal HRTF spectra magnitude shown in Fig. 7 for each measurement direction p. Direction are listed as (elevation, azimuth) pairs in a vertical-polar coordinate system.

p	RMSD (dB)	p	RMSD (dB)	p	RMSD (dB)	p	RMSD (dB)
(19, 148)	2.5	(0, 59)	3.0	(31, 69)	2.8	(−33,0)	3.3
(33, 180)	1.8	(33, 0)	1.9	(0,−59)	1.1	(0,−90)	2.1
(45, 55)	4.4	(55, 32)	2.8	(55,−148)	2.0	(19,−32)	2.1
(0, 121)	3.3	(59,−90)	1.2	(0, 90)	5.0	(−19,32)	2.9
(−33,180)	1.8	(90, 0)	1.9	(19,−148)	2.2	(55, 148)	1.9
(59, 90)	1.5	(0, 0)	3.0	(19, 32)	1.9	(31,−69)	1.7
(−19,−32)	2.4	(31,−111)	2.1	(−19,148)	2.4	(−19,−148)	3.2
(0,−121)	1.0	(31, 111)	2.3	(0, 180)	2.0	(55,−32)	2.2

surement, the predicted localization direction q' from the dense-grid direct HRTF is shown in Table II.

It should be noted here that most directions in the reciprocal grid were not present in the dense direct grid, and for such p the closest possible q' was several degrees off. Also, the localization model was basically "monaural" (operating on left ear data only) so its performance can be expected to be worse than the performance obtained with normal, binaural localization. Finally, the dense-grid direct HRTF was measured here at 0.84 m and the reciprocal HRTF was measured at 0.70 m, which may create slight discrepancies in HRTF. It was shown, among others, by Brungart and Rabinowitz (1999) and by Shinn-Cunningham *et al.* (2000) that the difference in the measurement distances creates changes in the measured HRTF that can not be represented by a simple intensity difference. A first reason for this is the fact that the ear may lie in the head shadow for a given source direction and distance but be out of the shadow for a source in the same direction but at the larger distance. Further, the pinna-related HRTF features may shift due to the parallax effect as the source distance decreases and the difference in source directions with respect to the head center and with respect to the pinna becomes more pronounced.

The results shown in Table II demonstrate that in most cases the predicted localization direction q' fell very close to the actual measurement direction p. The localization error, measured as an angle between vectors p and q', was within 5° for 22 (69%) reciprocal measurement directions, within 10° for 29 (91%) reciprocal measurement directions, and the maximum observed error was 19.5°. The average value of error over 32 measurement directions was 4.7°. Two measurement directions showing the largest errors were located in the overhead-back region of space in the contralateral hemisphere. It is known that in this region the HRTF is relatively featureless and does not vary much with the source direction, which explains the relatively large errors observed there. In fact, a study of errors made by humans in sound localization experiments (Carlile *et al.*, 1997) also suggested larger natural localization errors in the area over the listener and to the back. Because of the head shadowing, the recorded microphone signal for these two directions was also weak and noise-prone. However, even for these two directions the model-predicted localization direction lay on the correct cone of confusion (i.e., p and q' had very close azimuth values in the interaural-polar coordinate system; this is

TABLE II. Comparison between reciprocal HRTF measurement direction p and model-predicted localization direction q' (selected from the dense-grid direct HRTF set) for all 32 p. Angle between p and q' is shown as ang(p,q'). Directions are listed as (elevation, azimuth) pairs in a vertical-polar coordinate system.

No.	p	q'	ang(p,q')	No.	p	q'	ang(p,q')
1	(19, 148)	(20, 148)	1.0	17	(31, 69)	(35, 69)	4.0
2	(33, 180)	(35, 180)	2.0	18	(0,−59)	(0,−55)	4.0
3	(45, 55)	(50, 51)	5.4	19	(55,−148)	(60,−137)	7.7
4	(0, 121)	(5, 120)	5.1	20	(0, 90)	(5, 81)	10.3
5	(−33,180)	(−30,180)	3.0	21	(19,−148)	(20,−153)	4.8
6	(59, 90)	(65, 90)	6.0	22	(19, 32)	(20, 32)	1.0
7	(−19,−32)	(−20,−27)	4.8	23	(−19,148)	(−15,149)	4.1
8	(0,−121)	(0,−120)	1.0	24	(0, 180)	(0, 180)	0.0
9	(0, 59)	(0, 55)	4.0	25	(−33,0)	(−35,0)	2.0
10	(33, 0)	(30, 0)	3.0	26	(0,−90)	(0,−85)	5.0
11	(55, 32)	(60, 31)	5.0	27	(19,−32)	(20,−32)	1.0
12	(59,−90)	(50,−90)	9.0	28	(−19,32)	(−15,31)	4.1
13	(90, 0)	(90, 0)	0.0	29	(55, 148)	(70, 130)	17.0
14	(0, 0)	(0, 0)	0.0	30	(31,−69)	(30,−71)	2.0
15	(31,−111)	(30,−109)	2.0	31	(−19,−148)	(−15,−154)	7.0
16	(31, 111)	(25, 90)	19.5	32	(55,−32)	(50,−32)	5.0

not immediately obvious from Table II as angles in Table II are listed in vertical-polar coordinate system).

The fact that the HRTF measurements obtained at two ranges (0.70 m for the reciprocal setup and 0.84 m for the direct setup) compared well could have been expected from the work of Brungart and Rabinowitz (1999), where the effects of source distance on measured HRTF were studied. Only subtle differences were noticed when the HRTF measurements at 1.0 m and at 0.50 m were compared. The difference between the far-field HRTF and the HRTF taken at 0.70 m can be expected to be even smaller.

The simulated localization experiment was also performed using the sound localization model operating only with the spectral similarity coefficient given by Eq. (2) [i.e., ignoring the temporal similarity coefficient given by Eq. (3) altogether]. This was found to cause only a small change in the results. This is consistent with the observations made by Middlebrooks (1992) that subjects localized sounds consistently in the direction for which the directional transfer function magnitude spectrum most closely resembled the stimulus spectrum. This further confirmed the validity of the model used in the current paper for prediction of the localization direction. Furthermore, even with the model operating only on left ear data and with lack of exactly matching direct measurement directions for most of the reciprocal measurement directions, the observed predicted localization errors were small (except for the region of comparatively large natural human localization errors). The error could be expected to decrease further if HRTF for both ears were used in the model.

F. Eardrum SPL measurement

The sound intensity level produced at the eardrum by the in-ear microspeaker embedded within a silicone plug was then evaluated to ensure safety of the technique when used with a human subject. The KEMAR used in the experiment was equipped with a microphone located inside the head in the opening of the occluded ear simulator at the termination of the ear canal (i.e., at the simulated eardrum). Four signal recordings were performed at this location.

In the first recording, the KEMAR was placed in the direct measurement setup, and the ear canal was left open. The direct setup loudspeaker broadcast the test pulses used in the direct HRTF measurement technique. This recording thus provided an estimation of the sound level that would be observed at the eardrum during the direct HRTF measurement if the ear canal were not blocked. In the second recording, the setup was unchanged, but the ear canal was sealed with a silicone plug with an embedded microphone. In this way, the sound level in the direct measurement procedure with a blocked ear canal was evaluated. In the third recording, the sound level at the simulated eardrum was measured during the course of a reciprocal measurement of KEMAR HRTF. Finally, the fourth recording consisted of the voice of a person talking in a normal voice approximately 1 m in front of the KEMAR. Table III shows the sound pressure levels evaluated over different time windows for all four signals. The first column is the dB level corresponding to the peak amplitude of the signal, the second column is the rms dB SPL over the window that begins at the start of the recorded pulse and is twice as long as the pulse (i.e., 4.9 ms for the reciprocal HRTF measurement method and 3.6 ms for the direct HRTF measurement method), and the third column is the rms dB SPL taken over the whole signal. (All dB levels are relative to 20 μPa sound pressure.) It can be seen that the peak SPL for the reciprocal measurement method was about 14 dB higher than the peak speech SPL; however, it was well below tolerance threshold shown in OSHA (1974). Also, the average SPL for the reciprocal measurement method was only 11 dB higher than the average speech SPL due to the long interpulse interval.

TABLE III. Comparison of the SPL observed in the KEMAR ear canal for the direct method, for the reciprocal method, and for a speech signal.

	Peak SPL (dB)	rms SPL, windowed (dB)	rms SPL, whole signal (dB)
Direct method	80.1	67.1	50.0
Direct w/blocked ear canal	57.6	44.8	28.5
Reciprocal method	89.4	78.9	63.3
Speech recording	75.8	N/A	52.2

G. Discussion

The main issue to be decided is whether the directly measured HRTF and the reciprocally measured HRTF are perceptually identical despite the small differences observed (see Fig. 7). The perceptual identity question is impossible to answer without doing an actual perceptual experiment with a set of human subjects. In this paper, a sound localization model [an extension of the model presented in Middlebrooks (1992)] was used instead. The results of this comparison showed that the average localization error predicted by the model is less than the spacing between HRTF measurement directions typically used now in the state-of-the-art HRTF measurement facilities. In fact, this result is all the more remarkable because the comparison was performed monaurally and because the reciprocal and the direct HRTF measurements were obtained under somewhat different conditions.

Based on the results presented, it can be reasonably expected that the reciprocally measured HRTF can be used interchangeably with the directly measured HRTF in virtual audio synthesis and that the errors introduced by such exchange would lie within the errors caused by the discreteness of the measurement grid, by the natural errors in sound localization by humans, and by the natural variability in HRTF between repetitive experimental HRTF measurements due to subject positioning variability, microphone/speaker positioning variability, and noise.

VI. CONCLUSIONS

A new experimental method of measuring the HRTF rapidly is presented. The method is based on the reciprocity principle and exchanges the positions of the speaker and the microphone in the direct HRTF measurement setup. The

method is validated in two setups (a sphere and a manikin head), and it is shown that the transfer functions measured with the proposed method and with the direct measurement method are in good agreement.

ACKNOWLEDGMENTS

Partial support of NSF Award Nos. 0086075 and 0205271 is gratefully acknowledged. We thank Dr. Kenneth W. Grant of the Army Audiology and Speech Center at Walter Reed Army Medical Center, Washington DC, for allowing us to use the center's KEMAR manikin in these experiments. We would also like to acknowledge helpful discussions with Dr. R. O. Duda, Dr. V. R. Algazi, and Dr. B. G. Shinn-Cunningham regarding KEMAR HRTF measurement procedure and to thank Emkay electronics for providing us with the microspeakers used in the experiments. Finally, we would like to thank Dr. Armin Kohlrausch, associate editor, and three anonymous reviewers who provided a very careful review that helped us to significantly improve the manuscript.

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

Algazi, V. R., Duda, R. O., Morrison, R. P., and Thompson, D. M. (**2001a**). "Structural composition and decomposition of HRTFs," Proc. IEEE WASPAA 2001, New Paltz, NY, pp. 103–106.

Algazi, V. R., Avendano, C., and Duda, R. O. (**2001b**). "Elevation localization and head-related transfer function analysis at low frequencies," J. Acoust. Soc. Am. **109**, 1110–1122.

Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C. (**2001c**). "The CIPIC HRTF database," Proc. IEEE WASPAA 2001, New Paltz, NY, pp. 99–102.

Algazi, V. R., Duda, R. O., Duraiswami, R., Gumerov, N. A., and Tang, Z. (**2002a**). "Approximating the head-related transfer function using simple geometric models of the head and torso," J. Acoust. Soc. Am. **112**, 2053–2064.

Algazi, V. R., Duda, R. O., and Thompson, D. M. (**2002b**). "The use of head-and-torso models for improved spatial sound synthesis," Proc. AES 113th Convention, Los Angeles, CA, preprint 5712.

Batteau, D. W. (**1967**). "The role of the pinna in human localization," Proc. R. Soc. London, Ser. B **168**, 158–180.

Blauert, J. (**1969**). "Sound localization in the median plane," Acustica **22**, 205–213.

Blauert, J. (**1997**). *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT, Cambridge, MA).

Blommer, M. A. and Wakefield, G. H. (**1997**). "Pole-zero approximation for head-related transfer function using a logarithmic error criterion," IEEE Trans. Speech Audio Process. **5**, 278–287.

Bronkhorst, A. W., (**1995**). "Localization of real and virtual sound sources," J. Acoust. Soc. Am. **95**, 2542–2553.

Brungart, D. S. and Rabinowitz, W. M. (**1999**). "Auditory localization of nearby sources. Head-related transfer functions," J. Acoust. Soc. Am. **106**, 1465–1479.

Burkhard, M. D. and Sachs, R. M. (**1975**). "Anthropometric manikin for acoustic research," J. Acoust. Soc. Am. **58**, 214–222.

Carlile, S., (ed.) (**1996**). *Virtual Auditory Space: Generation and Applications* (Landes, Austin).

Carlile, S., Leong, P., and Hyams, S. (**1997**). "The nature and distribution of errors in the localization of sounds by humans," Hear. Res. **114**, 179–196.

Divenyi, P. L. and Oliver, S. K. (**1989**). "Resolution of steady-state sounds in simulated auditory space," J. Acoust. Soc. Am. **85**, 2042–2052.

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.

Duraiswami, R. and Gumerov, N. A. (**2003**). "Method for measurement of head related transfer functions," United States Patent Application 20040091119, Serial No. 10/702465, filed 7 November 2003.

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, W. G. and Martin, K. D. (**1995**). "HRTF measurements of a KEMAR," J. Acoust. Soc. Am. **97**, 3907–3908.

Grassi, E., Tulsi, J., and Shamma, S. A. (**2003**). "Measurement of head-related transfer functions based on the empirical transfer function estimate," Proc. 2003 Intl. Conf. on Auditory Displays (ICAD 2003), Boston, MA, pp. 119–121.

Hartmann, W. M. (**1999**). "How we localize sound," Phys. Today **1999**, 24–29.

Hartmann, W. M. and Wittenberg, A. (**1996**). "On the externalization of sound images," J. Acoust. Soc. Am. **99**, 3678–3688.

Jin, C., Leong, P., Leung, J., Corderoy, A., and Carlile, S. (**2000**). "Enabling individualized virtual auditory space using morphological measurements," Proc. of the 1st IEEE Pacific-Rim Conf. on Multimedia (2000 Intl. Symposium on Multimedia Information Processing), Sydney, Australia, pp. 235–238.

Kahana, Y. (**2000**). "Numerical modeling of the head-related transfer function," Ph.D. thesis, ISVR, University of Southampton, UK.

Kahana, Y. and Nelson, P. A. (**2000**). "Spatial acoustic mode shapes of the human pinna," Proc. AES 109th Convention, Los Angeles, CA, preprint 5218.

Katz, B. F. G. (**2001a**). "Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation," J. Acoust. Soc. Am. **110**, 2440–2448.

Katz, B. F. G. (**2001b**). "Boundary element method calculation of individual head-related transfer function. II. Impedance effects and comparisons to real measurements," J. Acoust. Soc. Am. **110**, 2449–2455.

Kulkarni, A. and Colburn, H. S. (**1998**). "Role of spectral detail in sound-source localization," Nature (London) **396**, 747–749.

Kulkarni, A. and Colburn, H. S. (**2004**). "Infinite-impulse-response models of the head-related transfer function," J. Acoust. Soc. Am. **115**, 1714–1728.

Kulkarni, A., Isabelle, S. K., and Colburn, H. S. (**1999**). "Sensitivity of human subjects to head-related transfer-function phase spectra," J. Acoust. Soc. Am. **105**, 2821–2840.

Langendijk, E. H. A. and Bronkhorst, A. W. (**2000**). "Fidelity of three-dimensional-sound re-production using a virtual auditory display," J. Acoust. Soc. Am. **107**, 528–537.

Mehrgardt, S. and Mellert, V. (**1977**). "Transformation characteristics of the external human ear," J. Acoust. Soc. Am. **61**, 1567–1576.

Middlebrooks, J. C. (**1992**). "Narrow-band sound localization related to external ear acoustics," J. Acoust. Soc. Am. **92**, 2607–2624.

Middlebrooks, J. C. (**1999a**). "Individual differences in external-ear transfer functions reduced by scaling in frequency," J. Acoust. Soc. Am. **106**, 1480–1492.

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

Middlebrooks, J. C., Makous, J. C., and Green, D. M. (**1989**). "Directional sensitivity of sound-pressure levels in the human ear canal," J. Acoust. Soc. Am. **86**, 89–108.

Møller, H., Sørensen, M. F., Hammershøi, D., and Jensen, C. B. (**1995**). "Head-related transfer functions of human subjects," J. Audio Eng. Soc. **43**, 300–321.

Morse, P. M. and Ingard, K. U. (**1968**). *Theoretical Acoustics* (Princeton U.P., Princeton, NJ).

Musicant, A. D. and Butler, R. A. (**1984**). "The influence of pinnae-based spectral cues on sound localization," J. Acoust. Soc. Am. **75**, 1195–1200.

Occupational Safety and Health Administration, U. S. Dept. of Labor (**1974**). "Occupational Noise Exposure," Regulation 1910.95, Standards 29 CFR.

Pralong, D. and Carlile, S. (**1994**). "Measuring the human head-related transfer functions: A novel method for the construction and calibration of a miniature in-ear recording system," J. Acoust. Soc. Am. **95**, 3435–3444.

Rabinowitz, W. M., Maxwell, J., Shao, Y., and Wei, M. (**1993**). "Sound localization cues for a magnified head: Implications from sound diffraction about a rigid sphere," Presence **2**, 125–129.

Runkle, P., Yendiki, A., and Wakefield, G. H. (**2000**). "Active sensory tuning for immersive spatialized audio," Proc. 2000 Intl. Conf. on Auditory Displays (ICAD 2000), Atlanta, GA, pp. 141–144.

Schroeder, M. R. (**1979**). "Integrated-impulse method measuring sound decay without using impulses," J. Acoust. Soc. Am. **66**, 497–500.

Shaw, E. A. G. and Teranishi, R. (**1968**). "Sound pressure generated in an

external-ear replica and real human ears by a nearby point source," J. Acoust. Soc. Am. **44**, 240–249.
Shinn-Cunningham, B. G., Santarelli, S. G., and Kopco, N. (**2000**). "Tori of confusion: Binaural localization cues for sources within reach of a listener," J. Acoust. Soc. Am. **107**, 1627–1636.
Strutt, J. W., (Lord Rayleigh) (**1904**). "On the acoustic shadow of a sphere," Philos. Trans. R. Soc. London, Ser. A **203**, 87–89.
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. J. (**1989**). "Headphone simulation of free-field listening. I. Stimulus synthesis," J. Acoust. Soc. Am. **85**, 858–867.
Wright, D., Hebrank, J. H., and Wilson, B. (**1974**). "Pinna reflections as cues for localization," J. Acoust. Soc. Am. **56**, 957–962.
Zahorik, P. A., Wightman, F. L., and Kistler, D. J. (**1995**). "On the discriminability of virtual and real sound sources," Proc. IEEE WASPAA 1995, New Paltz, NY, pp. 76–79.
Zhou, B., Green, D. M., and Middlebrooks, J. C. (**1992**). "Characterization of external ear impulse responses using Golay codes," J. Acoust. Soc. Am. **92**, 1169–1171.