

# Headphone simulation of free-field listening. I: Stimulus synthesis

Frederic L. Wightman

*Department of Psychology and Waisman Center, University of Wisconsin—Madison, Madison, Wisconsin 53705*

Doris J. Kistler

*Waisman Center, University of Wisconsin—Madison, Madison, Wisconsin 53705*

(Received 14 May 1988; accepted for publication 12 October 1988)

This article describes techniques used to synthesize headphone-presented stimuli that simulate the ear-canal waveforms produced by free-field sources. The stimulus synthesis techniques involve measurement of each subject's free-field-to-eardrum transfer functions for sources at a large number of locations in free field, and measurement of headphone-to-eardrum transfer functions with the subject wearing headphones. Digital filters are then constructed from the transfer function measurements, and stimuli are passed through these digital filters. Transfer function data from ten subjects and 144 source positions are described in this article, along with estimates of the various sources of error in the measurements. The free-field-to-eardrum transfer function data are consistent with comparable data reported elsewhere in the literature. A comparison of ear-canal waveforms produced by free-field sources with ear-canal waveforms produced by headphone-presented simulations shows that the simulations duplicate free-field waveforms within a few dB of magnitude and a few degrees of phase at frequencies up to 14 kHz.

PACS numbers: 43.66.Qp, 43.66.Pn, 43.66.Yw, 43.88.Si [WAY]

## INTRODUCTION

Research on the acoustical and psychological bases of human sound localization suggests that the primary acoustical cues used to determine sound source position are the interaural differences in time of arrival ( $\Delta T$ ) of a sound wave at a listener's two ears, the interaural differences in overall intensity ( $\Delta I$ ) of the sound, and the position-dependent filtering caused by the interaction of an incoming sound wave with the folds of the pinnae (for an excellent review of this literature, see Butler, 1975). Lord Rayleigh's classic "duplex theory" (Strutt, 1907), which has motivated most of the modern research on sound localization, ignored pinna filtering and held that apparent sound source position was determined entirely by  $\Delta T$  cues at low frequencies and  $\Delta I$  cues at high frequencies. In the past quarter century, many studies have shown that cues provided by pinna filtering are more important than previously believed, especially for localizing sounds on the median plane (where  $\Delta T$  and  $\Delta I$  cues are minimized), and for establishing the "externalized" (out-of-head) character of sounds in the natural environment (Blauert, 1969; Butler and Belendiuk, 1977; Butler and Planert, 1976; Gardner and Gardner, 1973; Hebrank and Wright, 1974; Plenge, 1974). A recent model of sound localization also gives recognition to the importance of pinna cues (Searle *et al.*, 1976).

Even though the recent research recognizes the importance of pinna cues, there have been relatively few experiments in which these cues have been manipulated systematically. Most of the available data have come from experiments in which localization performance was measured before and after an experimental manipulation designed to reduce or remove pinna cues. These manipulations

have included filling the pinna folds with putty (Gardner and Gardner, 1973; Oldfield and Parker, 1984), covering the pinnae with blocks (Gardner and Gardner, 1973), and inserting tubes into the ear canals (Jongkees and Groen, 1946; Fisher and Freedman, 1968). All of the studies reported decrements in localization acuity following pinna deformation. However, since the techniques used did not allow precise, systematic control over the stimulus, the experiments provided only limited information about the role of these cues.

Presenting stimuli over headphones allows complete specification of the stimuli at a listener's ears and thus solves the stimulus control problem. Most of what we have learned about processing of  $\Delta T$  and  $\Delta I$  cues has come from experiments in which the stimuli were delivered by headphones. However, the extent to which the results of these experiments could be generalized to free-field listening conditions has been questioned, mostly because of the unnatural quality of sounds heard over headphones (e.g., they are typically heard as originating inside the listener's head). The fact that these experiments have been called "lateralization" rather than "localization" experiments represents explicit recognition of this lack of generalizability. Nevertheless, there are undeniable advantages to headphone stimulus delivery. Several investigators have attempted to bring these advantages to bear on various questions about human sound localization other than those relating to processing of  $\Delta T$  and  $\Delta I$  cues. For example, Bloom (1977) and Watkins (1978) attempted to simulate source elevation changes by altering the spectrum of headphone-delivered stimuli in a manner analogous to pinna filtering. Other researchers have studied the apparent locations of headphone-presented sounds which had been "binaurally" recorded, with microphones placed in the ears of a dummy head (e.g., Plenge, 1974) or at the ear-canal

entrance of a human listener (e.g., Butler and Belendiuk, 1977). These pioneering experiments have produced important, suggestive data. However, the results are not readily generalized to free-field localization conditions, since the experiments did not directly assess the degree to which the headphone stimuli reproduced either the acoustical or the psychological features of a free-field stimulus.

We have attempted to produce a veridical simulation of the free-field listening experience by using digital techniques to synthesize headphone-presented stimuli. The synthesis techniques and objective tests of their acoustical adequacy are described here; psychophysical tests of the perceptual adequacy of the simulation are described in the companion article (Wightman and Kistler, 1989). The basic assumption that guides our approach is that, if the acoustical waveforms at a listener's eardrums are the same under headphones as in free field, then the listener's experience should also be the same. This assumption is an obvious oversimplification, in that it denies the relevance of head movements, visual cues, and other localization cues. However, the success we have had in the psychophysical validation experiments (Wightman and Kistler, 1989) indicates that within a limited range of stimulus conditions, the assumption may be warranted.

## 1. METHOD

Our approach is based on well-understood linear filtering principles. Let  $x_1(t)$  represent an electrical signal that drives a loudspeaker in free field, and let  $y_1(t)$  represent the resultant electrical signal from a probe microphone positioned at a listener's eardrum. Similarly, let  $x_2(t)$  represent an electrical signal that drives a headphone, with  $y_2(t)$  the resultant microphone response. Given  $x_1(t)$ , our goal is to produce  $x_2(t)$  such that  $y_2(t)$  equals  $y_1(t)$ . We do this by designing a linear filter that transforms  $x_1(t)$  into the desired  $x_2(t)$ .

The design of the appropriate filter is best described in the frequency domain. Thus  $X_1(j\omega)$ , or simply  $X_1$ , is the Fourier transform of  $x_1(t)$ ,  $Y_1$  is the transform of  $y_1(t)$  and so forth. The probe microphone's response to  $x_1(t)$  can be written:

$$Y_1 = X_1 LFM, \quad (1)$$

where  $L$  is the loudspeaker transfer function,  $F$  the free-field-to-eardrum transfer function (sometimes called the head-related transfer function, or HRTF), and  $M$  the microphone transfer function. The probe microphone's response to  $x_2(t)$  can be written

$$Y_2 = X_2 HM, \quad (2)$$

where  $H$  represents the headphone-to-eardrum transfer function. Setting  $Y_1 = Y_2$  and solving for  $X_2$  yield

$$X_2 = X_1 LF/H. \quad (3)$$

This equation shows that the desired filter transfer function  $T$  is given by

$$T = LF/H. \quad (4)$$

Thus, if the signal  $x_1(t)$  is passed through this filter and the resultant  $x_2(t)$  is transduced by the headphone, the signal

recorded by the probe microphone at the eardrum will be  $y_1(t)$ , the same signal produced by the loudspeaker in free field. This is represented in the frequency domain by substituting the right side of Eq. (3) for  $X_2$  in Eq. (2).

The filter described in (4) applies only to a single free-field loudspeaker position and one ear. To synthesize each stimulus, then, we must design a pair of filters (one for each ear) for each desired free-field source position.

The first phase of our synthesis procedure involves measurement of the free-field-to-eardrum transfer function (HRTF) for each ear of a subject, for a large number of sound source positions. In practice, what we actually measure is a quantity like  $Y_1$  in Eq. (1), which includes not only the free-field-to-eardrum characteristics ( $F$ ), but also the characteristics of the test signal ( $X_1$ ), loudspeaker ( $L$ ), and microphone ( $M$ ). A headphone-to-eardrum transfer function [like  $Y_2$  in Eq. (2)] is also measured for each ear of the same subject. In the second phase of the synthesis, each desired experimental stimulus is digitally filtered. The transfer functions of the filters (one for the left ear stimulus and one for the right) are defined in Eq. (4). Ideally, when the filtered stimuli are presented to the subject over the headphones, the waveforms reaching the eardrums should be identical to those produced by a free-field stimulus. The error in the procedure is quantified by recording the stimuli at the eardrums in the free-field and headphone conditions and computing the difference.

### A. Transfer function measurement

Both free-field and headphone transfer function measurements were made using a technique loosely based on the procedure described by Mehrgardt and Mellert (1977). A wideband, noiselike signal was presented (either by loudspeaker or headphone) repetitively, and the response at the listener's eardrum was obtained by averaging the output of a probe microphone. The Fourier transform of this response was divided by the Fourier transform of the signal to produce an estimate of the transfer function in question. The signal was 20.48 ms in duration and was computed via an inverse DFT so that both the amplitude and phase components of its spectrum could be tailored to maximize the signal-to-noise ratio in the response recordings.<sup>1</sup> Specifically, the amplitude spectrum of the signal was flat from 200 to 4000 Hz, where it increased abruptly by 20 dB. Thereafter, it was flat to 14 kHz. The signal contained no energy below 200 Hz or above 14 kHz. The phase spectrum was computed to minimize the peak factor of the signal (Schroeder, 1970). The signal was output continuously (hence, with a repetition frequency of about 50 Hz), via a 16-bit D/A converter (controlled by an IBM-PC) at a rate of 50 kHz. No antialiasing filters were used. For the free-field measurements, the signal was transduced by a miniature loudspeaker (Realistic Minimus-7). For the headphone measurements, the signal was transduced by a pair of Sennheiser HD 340 headphones, driven in phase. Signals were presented at approximately 70 dB SPL, a level chosen to reduce the contaminating effects of the acoustic reflex.

The acoustical response at the eardrum was measured with a miniature electret microphone (Etymotic) coupled to

a silicone rubber probe tube with an outer diameter of less than 1 mm. This probe microphone system, with its matching preamplifier and compensation network, had a sensitivity of about 50 mV/Pa, and a frequency response which was relatively flat ( $\pm 2.5$  dB) from 200 Hz to 14 kHz. Two matched microphones were used, one for each ear, and the responses from both were measured simultaneously. The amplified microphone outputs were digitized (simultaneously) using 16-bit A/D converters (controlled by the IBM-PC) at a 50-kHz sampling rate. The responses to 1000 periods of the signal were averaged with floating-point precision, a spectral resolution of 48.8 Hz, and a worst-case signal-to-noise ratio of well over 20 dB in the range 200 Hz–14 kHz.

The acoustical measurements were made with the tips of the probe tubes positioned roughly in the middle of the subject's ear canal, about 1–2 mm from the eardrum. This position was chosen in order to be certain the measurements would capture all direction-dependent effects (which may not be the case for measurements at the ear-canal entrance) and to avoid standing-wave nulls at high frequencies. At 14 kHz, the highest frequency of interest in our work, the first standing-wave null would occur at about 6 mm from the eardrum (assuming the ear canal is a uniform tube closed at one end). To avoid occluding the ear canals, the probe tubes were held in place with custom (i.e., different for each subject) Lucite earmold shells, trimmed so that they did not extend into the concha when inserted, and bored out to a thickness of less than 0.5 mm. With the earmold shell in place, the probe tube was inserted into a thin, semirigid guide tube that was cemented to the wall of the earmold shell. The length of each guide tube was calibrated, at the time the earmold assembly was made, so that with the probe inserted as far as its collar-stop would allow, the probe tip was about 1 mm from the eardrum. This calibration was accomplished by inserting a human hair into the guide tube until the subject indicated that the hair had touched the eardrum. The hair was then marked and withdrawn so that the appropriate length for the guide tube could then be determined. The body of the microphone was left hanging at the side of the subject's ear. Figure 1(a) shows the earmold shell with microphone attached, and Fig. 1(b) shows the whole assembly in place in a subject's ear.

For free-field measurements, the periodic wideband signal was transduced by one of eight loudspeakers, each positioned 1.38 m from the subject in an anechoic chamber. The loudspeakers were mounted on a semicircular arc (2.76-m diameter), the ends of which were attached directly above and directly below the subject. The loudspeakers were aimed at the position of the subject's head in order to minimize the influence of loudspeaker directionality (which we found to be virtually nonexistent within 10 deg of the speaker axis). The entire arc assembly could be rotated (by hand crank) around the vertical axis, and positioned with a precision of about 0.5 deg. The subject was seated on an adjustable stool (with back) so that his/her head was at the center of the arc. The speakers were mounted at  $-36^\circ$ ,  $-18^\circ$ ,  $0^\circ$ ,  $+18^\circ$ ,  $+36^\circ$ ,  $+54^\circ$ ,  $+72^\circ$ , and  $+90^\circ$ -deg elevation relative to the horizontal plane passing through the subject's ears. The

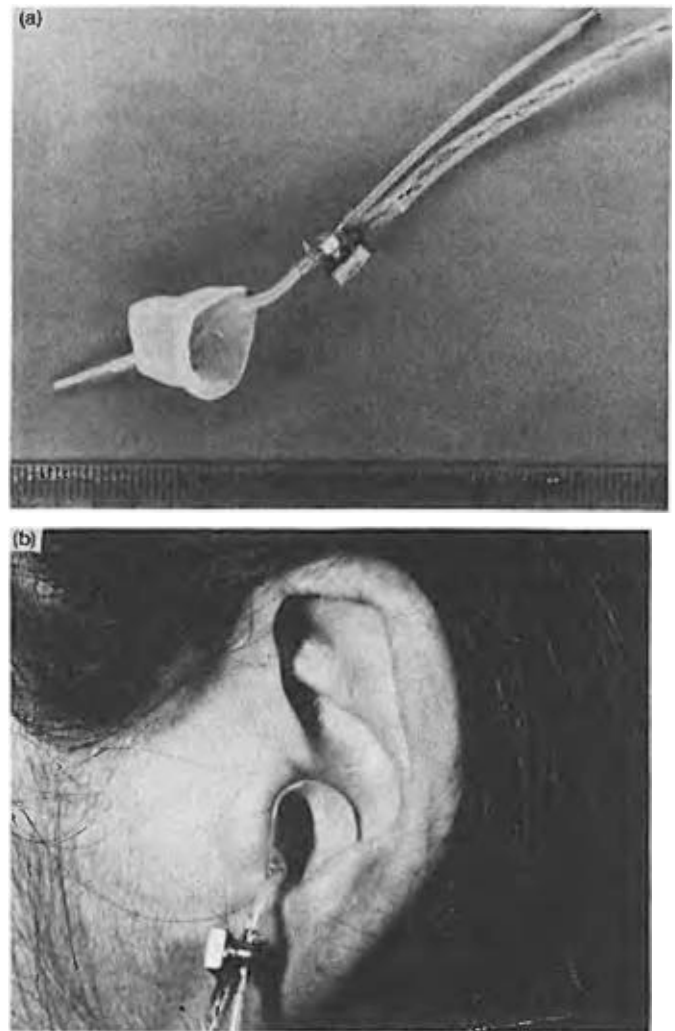


FIG. 1. Photographs of the custom earmold shell microphone holder used for acoustical transfer function measurements. (a) The shell with microphone probe tube inserted. The major divisions on the scale are 1 cm and the minor divisions are 1 mm. (b) The assembly in place in a subject's ear.

measurements were made at all elevations except  $+72^\circ$  and  $+90^\circ$  deg, and at all azimuths around the circle in 15-deg steps. Thus transfer functions were measured from both ears at 144 source positions.

A typical measurement session lasted about an hour. After the microphones were fitted in the subject's ear canals, the subject was seated in the anechoic chamber, and instructed on how to set the azimuth of the loudspeaker speaker arc using the hand crank to turn the arc. Then, with the subject alone in the chamber, the arc was moved to the first azimuth setting (usually directly behind the subject). Depending on the condition under study, the subject either looked directly forward and held his/her head still, or bit down on a bitebar, which could be attached rigidly to the subject's seat. After the subject signaled the experimenter that all was ready, measurements were made in rapid succession at all six elevations, in both ears simultaneously. About 2 min were required to make the six pairs of measurements at each azimuth. The subject then moved the arc to the next location and the sequence was repeated. Finally, after measurements

had been made at all 24 azimuths, the subject put on the headphones, taking care not to disturb the position of the microphones, and a pair of transfer function measurements was taken with the headphones being used to transduce the wideband test signal. Ten young adults (six females, four males), with no history of hearing problems, participated as subjects.

## B. Digital filter construction

Each raw data record consisted of the time-domain representation of a signal recorded from a probe microphone in a subject's ear canal. This signal included not only the direction-specific characteristics of the subject's outer ear (and head, shoulders, etc.), but also the characteristics of the original test signal, the loudspeaker (or headphones), and the measuring microphone. To obtain an uncontaminated free-field-to-eardrum transfer function characteristic (HRTF) or an uncontaminated headphone-to-eardrum transfer function, the effects of the signal, loudspeaker (or headphone), and microphone must be removed. This could be done by transforming the raw data record into the frequency domain (via an FFT) and dividing by the frequency domain representation of the characteristics of the signal, the microphone, and the loudspeaker or headphone. In our case, to produce the digital filters required for stimulus synthesis, we divided the frequency domain representations of the signals recorded in free field by the frequency domain representations of the same signals recorded under headphones.<sup>2</sup> Since the stimulus and microphone characteristics appear in both the numerator and denominator terms, they cancel. The loudspeaker characteristics were not removed from the digital filters used to synthesize stimuli. All digital signal processing, including test stimulus generation, FFT computations, digital filter design and implementation, and waveform analysis, was accomplished on a DEC VAX-11/750 computer using the ILS (Signal Technology, Inc.) software package.

## II. RESULTS AND DISCUSSION

We have made transfer function measurements from both ears of ten subjects with sources at 144 positions. For most of the subjects, we have several sets of measurements from a subset of the source positions, allowing quantitative assessments of the reliability of the measurements. While a complete analysis of these data would be inconsistent with the purpose of this article, we feel it is important to assess the stability and the validity of the measurements and the degree to which our synthesis procedures can actually duplicate a free-field stimulus with headphone presentation.

The measurements will be evaluated in three ways. First, we will present data on the stability of the measurements from a single subject and on the variability of the measurements from subject to subject. The data will show that for frequencies under 14 kHz, the measurements are stable, provided the signal-to-noise ratio is high. These stability data are difficult to evaluate, however, because it is not known how much instability can be tolerated for a successful simulation. The data will also show that the between-sub-

jects variability in the measurements is quite large, and that the pattern of between-subject differences across frequency is independent of source position. Second, we will present the results of an acoustical verification of our simulation technique, in which ear-canal recordings of real free-field stimuli and synthesized headphone-presented stimuli are directly compared. Third, we will show that our measurements are roughly consistent with comparable measurements reported elsewhere in the literature.

## A. Error and variability in the HRTF measurements

For the purposes of the measurements reported here, we make a distinction between "error" and "variability." Error refers to the influence of those factors that affect the repeatability of the measurements obtained on a single subject. Variability refers to the differences in the measurements from subject to subject.

Those components of error that are probably most relevant to the success of our simulations are the variations from measurement to measurement caused by head movements, the variations caused by the uncertainty of microphone placement, and the variations caused by the uncertainty of headphone placement. Figure 2 illustrates the magnitude of these errors for a typical subject. For purposes of illustration, we have represented the HRTF magnitudes in this figure both as raw, unprocessed DFT output and as  $\frac{1}{3}$ -oct band levels. The  $\frac{1}{3}$ -oct analysis, which smooths out high-frequency detail, was done not only for computational convenience but also because we feel the  $\frac{1}{3}$ -oct representation reflects cochlear output more accurately than the raw DFT representation. Most of the analyses of HRTFs presented later in this article will concern only the  $\frac{1}{3}$ -oct results. Figure 2(a) and (b) shows the raw HRTFs and the mean (and standard deviation) of the  $\frac{1}{3}$ -oct band levels from ten separate measurements of the HRTF for both ears from a source on the horizontal plane directly facing the left ear. The subject's head was fixed by means of a bitebar during each measurement and the earmold assembly was removed after each measurement. Thus these panels illustrate the influence of variability in microphone placement. Figure 2(c) and (d) shows the results from two comparable sets of ten measurements of the left and right HRTFs from the same source position and subject. For these measurements, the subject's head was not restrained during the measurements, and the earmold assembly was left in place until all ten measurements had been completed. Thus these panels illustrate the variability resulting from head movement during measurement. Note that at least for the ear facing the source (where signal-to-noise ratio is best), the  $\pm 2$  s.d. interval is less than 5 dB wide, even at high frequencies. Figure 2(e) and (f) provides an indication of the variability in the headphone transfer function measurements that results from uncertainty of headphone placement. The results from ten different headphone placements are shown. For these measurements, the earmold assembly was left in place and the headphones were repositioned after each of the ten measurements.

Substantial intersubject variability in the HRTF for a single source position is to be expected, given differences in head size and pinna shape. This HRTF variability has been

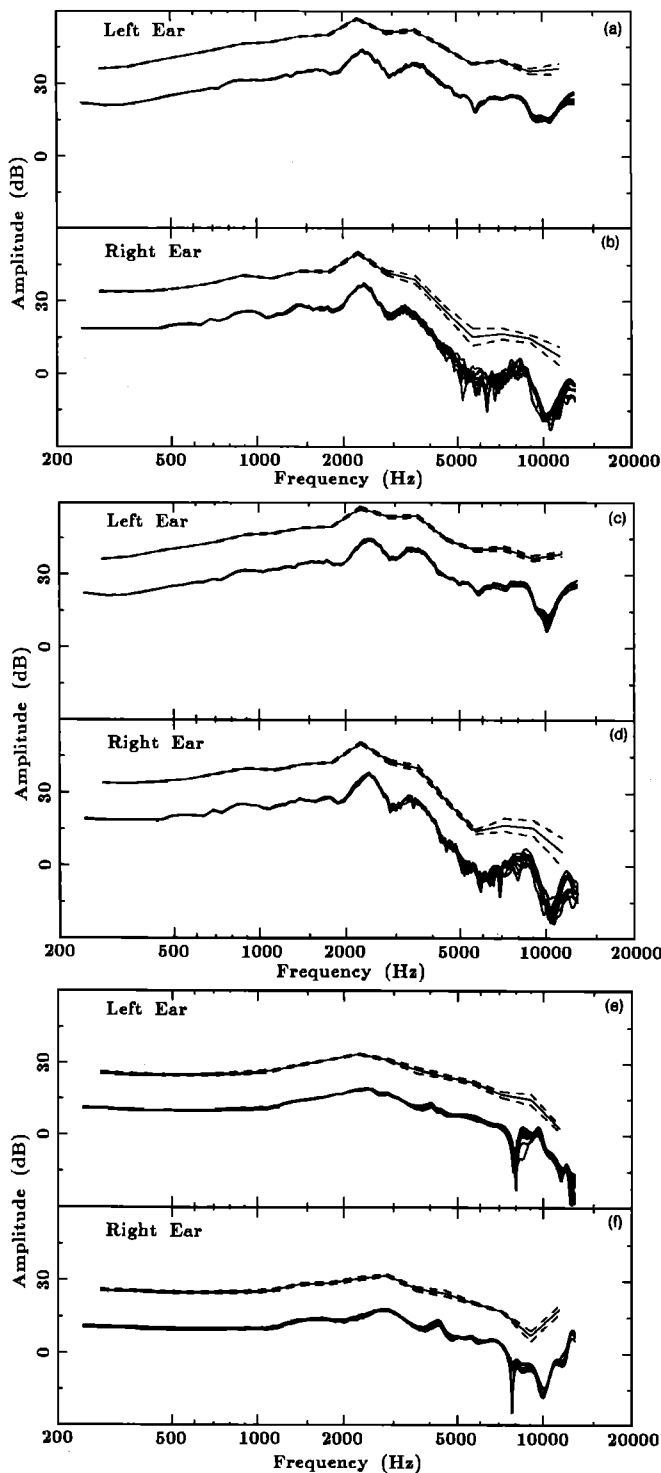


FIG. 2. Illustrations of typical variability in the acoustical transfer function measurements. The dB scale is based on an arbitrary reference; the spectrum level of the measurement system noise is at approximately  $-40$  dB on this scale. Each panel shows raw DFT magnitude functions from ten measurements. In addition each panel shows the means (solid line) and the  $\pm 2$  s.d. interval (dashed lines) of  $\frac{1}{3}$ -oct band magnitude levels from the ten measurements (displaced vertically by 10 dB for visibility). (a) and (b) Measurements from a subject's left and right ear, respectively, of the free-field-to-eardrum transfer function for a source directly opposite the left ear, at ear level. The subject's head was held in place with a bitebar during the measurements and the microphone assembly was removed after each measurement. (c) and (d) Comparable data except that, in this case, the microphone assembly was left in place for all ten measurements, and the subject's head was not held in place. (e) and (f) The variability resulting from headphone placement. For these measurements, the microphone assembly was left in place, and the headphones were removed and replaced after each measurement.

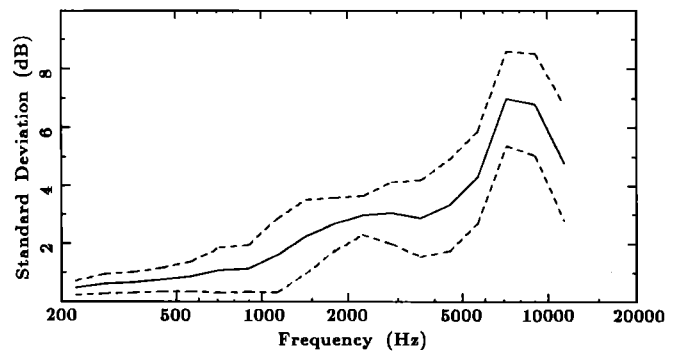


FIG. 3. Intersubject variability in the HRTF measurements ( $\frac{1}{3}$ -oct band levels). The solid line represents the mean standard deviation of  $\frac{1}{3}$ -oct band levels from ten subjects. Standard deviation in each  $\frac{1}{3}$ -oct band is computed from the  $\frac{1}{3}$ -oct band levels of the HRTF magnitude measurements from the ten subjects at a single source position; the mean standard deviation is computed by averaging the standard deviations in each  $\frac{1}{3}$ -oct band across the 48 different source positions. The dashed lines represent the  $\pm 2$  s.d. interval of the standard deviation estimates and reflect the extent to which the pattern of intersubject variability changes across the 48 positions.

reported before (e.g., Shaw, 1965) and is prominent in our data. In order to evaluate the intersubject variability, we transformed each subject's left-ear HRTF measurements from 48 source positions (all six elevations at each of eight azimuths, from 0 to 315 deg at 45-deg intervals) into  $\frac{1}{3}$ -oct band levels. We then analyzed these band levels (using a principal components analysis) to determine if the pattern of intersubject variability across frequency changed with source position. This analysis revealed no effect of source position (a single component accounted for 94% of the total variance). Figure 3 shows the intersubject standard deviation of the HRTF magnitude (in dB) as a function of band center frequency, averaged across the 48 source positions. The dashed lines represent  $\pm 2$  s.d. of the intersubject standard deviation estimates across the 48 source positions. As Fig. 3 shows, the variability in HRTF from subject to subject grows with frequency until it reaches a peak of almost 8 dB between 7 and 10 kHz. The intersubject standard deviation is significantly lower on each side of the peak. Note that an intersubject standard deviation of 7 dB implies that individual HRTFs would be expected to differ by as much as 28 dB ( $\pm 2$  s.d.).

## B. Acoustical verification of the simulation procedure

The working assumption of our simulation efforts was that if, using headphones, we could produce ear-canal waveforms identical to those produced by a free-field source, we would duplicate the free-field experience. The only conclusive test of this assumption would come from psychophysical experiments in which free-field and simulated free-field listening were directly compared (see Wightman and Kistler, 1989). To complement the psychophysical experiments, we evaluated the extent to which our simulation techniques could produce "acoustically correct" waveforms in subjects' ear canals. In other words, we compared ear-canal

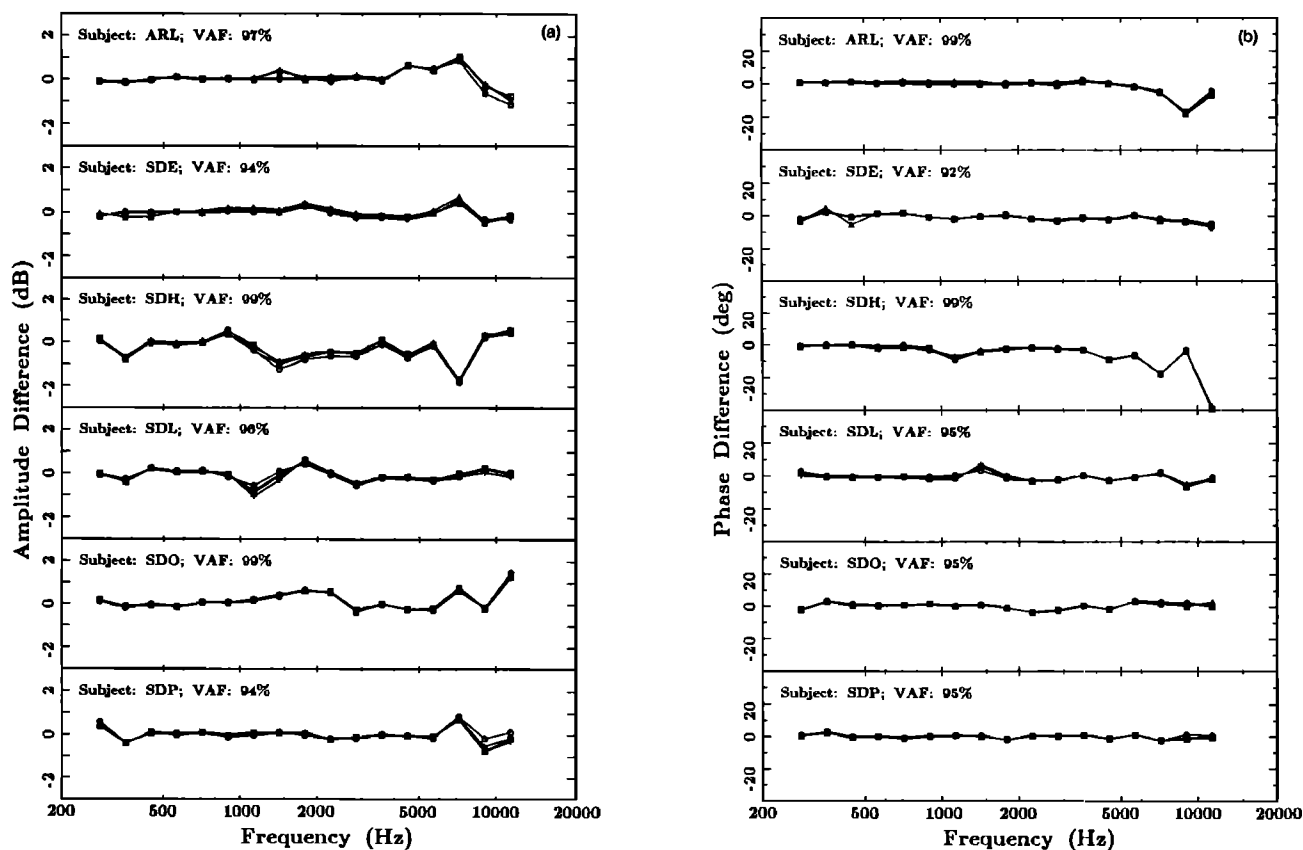


FIG. 4. Results from the acoustical verification experiments. Data are shown from six subjects and four source positions each. The different symbols represent different source positions. (a) The 1/3-oct amplitude difference in dB, measured at the listener's eardrum, between the spectrum of a free-field stimulus and the spectrum of a synthesized stimulus presented over headphones. (b) The phase difference in degrees. The percent variance accounted for (VAF) in each panel reflects the extent to which the pattern of differences as a function of frequency is constant across the four positions.

waveforms produced by free-field sources with ear-canal waveforms produced by simulated free-field sources.

The first step of the acoustical verification procedure was to measure the HRTF of a subject for a given source position. Then, the headphone transfer function was measured. Finally, the frequency domain representation of an FIR digital filter was computed by dividing the HRTF by the headphone transfer function. Normally, this digital filter would be used to process a stimulus for simulated free-field listening. In theory, if such a processed stimulus is presented over the headphones, the headphone response cancels, leaving the free-field HRTF characteristics imposed on the stimulus. To evaluate the extent to which the theory would hold in our situation, we used the impulse-response of the digital filter as a stimulus, presented that stimulus to listeners over headphones, and recorded the result.<sup>3</sup> With no error, the Fourier transform of the recorded stimulus should equal the originally measured HRTF. Figure 4 shows the differences in 1/3-oct bands (magnitude in dB and phase in degrees) between the original HRTFs and the transforms of the test stimulus recordings for six of the ten subjects at four source positions.<sup>4</sup> In general, it can be seen that the error is less than 1–2 dB in magnitude and 10 deg in phase. Most important is the fact that the error is independent of source position. A principal components analysis of the patterns of errors across frequency reveals that the variance accounted for by the first extracted component (a measure of the overall simi-

larity of the patterns) is greater than 92% of the total in all cases. In other words, the errors in our procedure are comparable at all source positions, and thus are similar to those that would be introduced by a slight change in headphone characteristics. We feel it is highly unlikely that this error is perceptually significant.

One possible source of the error revealed by our acoustical verification experiment is the acoustic reflex. The stimuli in the experiment, which produced the data shown in Fig. 4, were delivered at about 70 dB SPL in the headphone condition, and about 70 dB SPL in free field (exact sound-pressure levels in the two conditions were not recorded.) When the headphone stimuli were delivered at 90 dB SPL, the difference between free-field and headphone recordings were much greater, as shown for two subjects in Fig. 5. Since 90 dB SPL is well within the linear range for all our equipment, we feel the differences between the 90- and 70-dB SPL conditions are probably caused by changes in ear-canal acoustics brought about by the acoustic reflex. If so, it is possible that the usual practice of making measurements of HRTFs at high stimulus levels in order to maximize signal-to-noise ratio leads to contamination of the data with acoustic reflex effects.

### C. Comparison with other HRTF data in the literature

More than 40 years ago, the first acoustical measurements of the transfer function of the external ear were re-

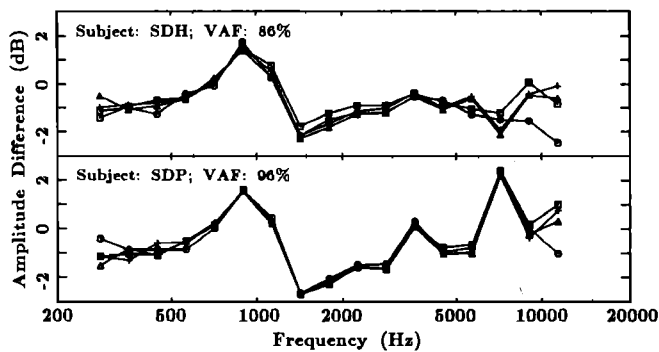


FIG. 5. Same as Fig. 4, except only magnitude data from two of the subjects are shown. For these cases, the headphone stimulus was 20 dB more intense than for the conditions represented in Fig. 4.

ported by Wiener and Ross (1946). Since that time, more than a dozen additional studies of external ear acoustics have been published (Blauert, 1983, provides a comprehensive summary of this work). Nearly all of the published studies include measurements only of the magnitude of the external ear transfer function, and most of the measurements apply to source positions on the horizontal plane.

### 1. Magnitude of the HRTF

The most comprehensive and readily accessible data on the magnitude of the HRTF are those reported by Shaw (1974) and by Mehrgardt and Mellert (1977). Shaw's data, which apply to source locations on the horizontal plane only, represent a compilation of measurements obtained from several laboratories using a variety of techniques. Mehrgardt and Mellert (after whose techniques our own are modeled) report measurements from sources both on the horizontal plane and on the vertical median plane. Figure 6 shows average HRTF data from Shaw (1974), Mehrgardt and Mellert (1977), and our own study for a source directly in front of the subject on the horizontal plane. The figure shows that for frequencies below 2.5 kHz, our data are in good agreement with those from the earlier studies. Above 2.5 kHz, however, there are sizable (5–10 dB) discrepancies, which are greatest near 10 kHz. A similar pattern emerges when we compare our data to those from the earlier studies from other source positions on the horizontal and vertical median plane. Our HRTF estimates are about 5 dB lower than others between 2.5 and 7 kHz, and show a deep spectral notch at around 7–10 kHz, which is not shown in the other data.

The discrepancies between our HRTF data and others in the literature are neither surprising nor worrisome. The differences could have arisen from any one or more of several sources, among which are the following.

(1) The data from Shaw (1974) and Mehrgardt and Mellert (1977) were not obtained by arithmetically averaging the HRTF measurements across subjects. Rather, a scheme of averaging was used which would tend to preserve better the details of the HRTF functions. In the case of Mehrgardt and Mellert, for example, the individual HRTF functions were shifted along the log-frequency axis until the best subject-to-subject match was found, and then the func-

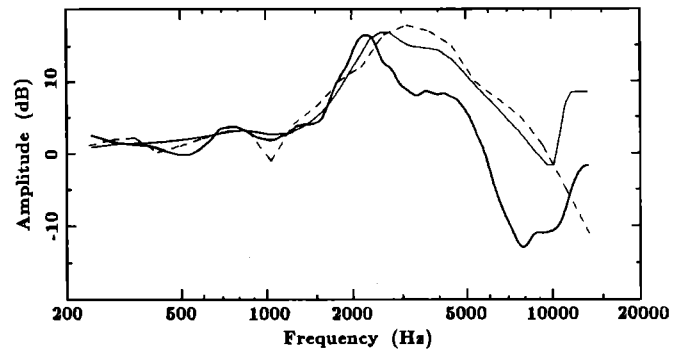


FIG. 6. Average HRTF for a source at 0-deg azimuth, 0-deg elevation, from the ten subjects in the current study (dark solid line) compared to data from the studies of Shaw (1974, light solid line) and Mehrgardt and Mellert (1977, dashed line.)

tions were averaged. Our data were obtained by arithmetic averaging.

(2) Neither Shaw's nor Mehrgardt and Mellert's HRTF data were obtained by measuring at the eardrum. Most of the data compiled by Shaw (1974) represent measurements taken at the entrance to the ear canal and later corrected by an independent measurement of the canal entrance to eardrum transfer function. The data reported by Mehrgardt and Mellert (1977) were similarly corrected. Our measurements were made at the eardrum, and are thus not subject to the errors that must be assumed to accompany any kind of *post hoc* correction.

(3) The acoustics of our measuring conditions were different from those in previous studies. For example, our microphone probe tube was probably much smaller than those used in previous studies (e.g., our probe tube had an outer diameter of about 0.8 mm, while the probe used by Mehrgardt and Mellert had a 1.7-mm outer diameter), and our measurements were made with a bored-out earmold in place, while in other studies the ear canal was perhaps less occluded (except by the probe tube itself). These acoustical differences, which would affect primarily the high-frequency HRTF estimates, may be significant.

We feel that since our primary interest is in the dependence of the HRTF on the location of the sound source, a more meaningful way to compare our data with those of Shaw (1974) and Mehrgardt and Mellert (1977) is to examine the changes in the HRTFs brought about by changes in source position. For this purpose we use azimuthal dependency and elevational dependency functions. The azimuthal dependency functions show, as a function of frequency, the dB increase or decrease in the magnitude of the HRTF caused by a change in azimuth from directly ahead to the azimuth in question (for sources on the horizontal plane.) The elevational dependency functions are computed for sources on the vertical median plane and have meaning comparable to the azimuthal dependency functions.

Figure 7 shows a subset of the azimuthal dependency functions presented by Shaw (1974). Also in the figure are our data, averaged over ten subjects, transformed into azimuthal dependency functions. The similarity between our data and Shaw's is gratifying, especially given the dramatic

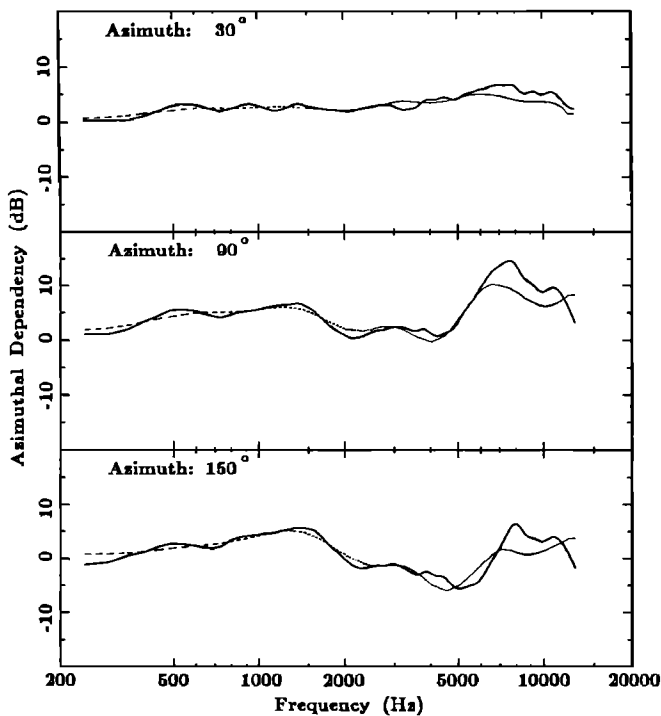


FIG. 7. The solid line in each panel represents an azimuthal dependency function (change in HRTF from 0° azimuth) averaged over ten subjects for a single source azimuth (0° elevation). The dotted line represents comparable data from Shaw (1974). Each panel shows the data from a different source azimuth.

differences of procedure across the various studies. Figure 8 shows some of the vertical median plane HRTF data from the Mehrgardt and Mellert (1977) study, transformed into elevational dependency functions. The figure also shows comparable data from our study, averaged across ten subjects. While direct comparison is not possible, since the source positions on the median plane were different in the two studies, the general agreement in form between our data and that from the Mehrgardt and Mellert (1977) study is obvious.

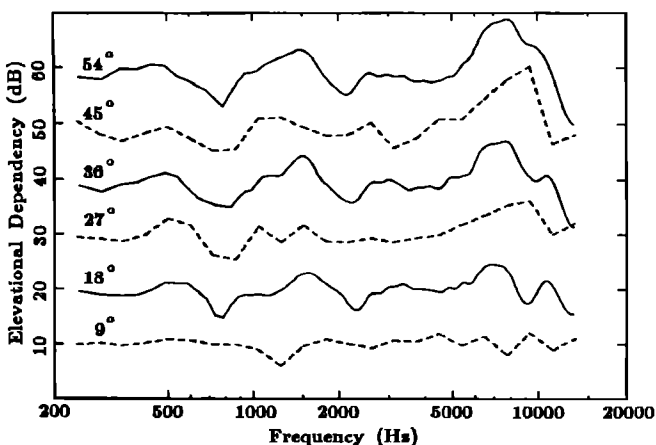


FIG. 8. Elevational dependency functions (change in HRTF from 0° elevation to the elevation noted at the left of each function) for a source at 0° azimuth. The solid lines represent average 1/3-oct band levels from the 10 subjects in the current study; the dashed lines represent data from the study of Mehrgardt and Mellert (1977).

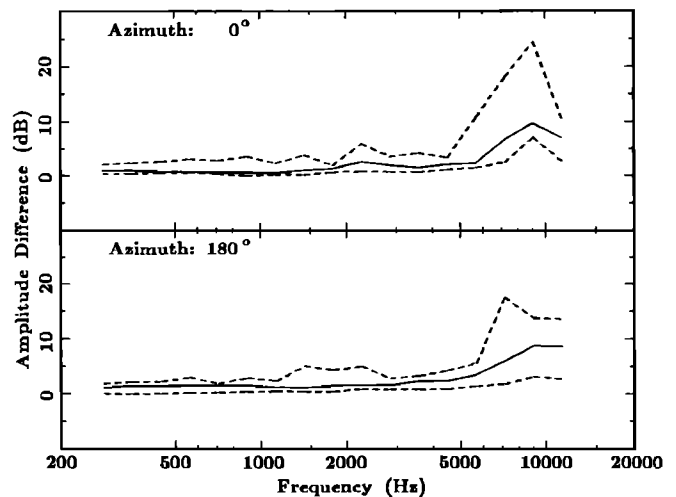


FIG. 9. Mean 1/3-oct band interaural asymmetry in the HRTF measurements. The two panels show the absolute value of the dB difference between the left and right HRTFs for a source at 0° azimuth and 0° elevation (upper panel), and for a source at 180° azimuth and 0° elevation (lower panel). The solid lines indicate the mean interaural asymmetry and the dashed lines indicate the upper and lower limits of the range of asymmetries.

## 2. Interaural asymmetry in the magnitude of the HRTF

It is typical to assume (e.g., Shaw, 1974) that HRTF measurements are symmetric about the vertical median plane. Under this assumption, the left HRTF and right HRTF would be identical for sources on the median plane, and symmetric elsewhere. For example, the left HRTF for a source on the horizontal plane directly opposite the left ear should equal the right HRTF for a horizontal plane source directly opposite the right ear. However, actual measurements show that the HRTFs are not symmetric (e.g., Searle *et al.*, 1975), and there have been suggestions that the asymmetries are perceptually important, especially for localizing sources on the vertical median plane. Figure 9 shows the interaural asymmetry we obtained in our transfer function measurements. The figure shows the mean and the range of the unsigned dB differences (in 1/3-oct bands) between the left and right HRTFs for sources at 180- and 0° azimuth and 0° elevation. These interaural asymmetries are comparable to those that have been reported previously (Searle *et al.*, 1975).

## 3. Phase of the HRTF

There are very few data in the literature on the phase of the HRTF. There are several possible reasons for this. First, many acoustical transfer function measurement techniques (e.g., those using 1/3-oct band noise) provide only magnitude information. Second, the relevance of the monaural phase characteristic to sound localization, primarily a binaural process, is questionable. Third, interpretation of the phase function is complicated by the influence of acoustical delays, which are difficult to quantify and which add an uninteresting linear component to the phase functions. Fourth, phase measurement techniques typically resolve phase at each frequency only within a  $\pm 2\pi$  range, so a complete appreciation of a phase function requires that the raw phase data be



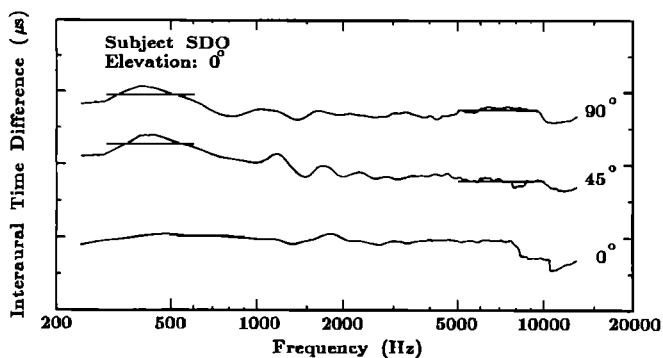


FIG. 10. Interaural time differences obtained from the HRTF measurements from one subject for sources at 0-deg elevation and three different azimuths (0, 45, and 90 deg). The horizontal lines indicate estimates of the low- and high-frequency asymptotic values of interaural time difference at each of the three azimuths. At 45-deg azimuth the low- and high-frequency asymptotes are 511 and 310  $\mu\text{s}$ , respectively; at 90-deg azimuth, they are 780 and 701  $\mu\text{s}$ , respectively.

“unwrapped,” a complex process that is susceptible to many sources of error (Tribolet, 1977). Mehrgardt and Mellert (1977) present data on the phase of the HRTF, which they obtained by addressing both the acoustical delay and the unwrapping issues. However, it is difficult to compare our phase data with theirs since the shape of the function as plotted depends critically on how the acoustical delay is quantified, and Mehrgardt and Mellert present no detail on this point. Nevertheless, we should mention that a decomposition of our HRTF data into all pass and minimum-phase components supports Mehrgardt and Mellert’s conclusion that the HRTF can be modeled by a minimum-phase system up to 10 kHz.

The phase *difference* between the transfer functions measured from the two ears is of considerable interest, since it is the basis of the important interaural time-difference cue. Kuhn (1977) has presented a rigid-sphere model for predicting interaural time differences in the azimuthal plane. He also presented empirical measurements of interaural time difference from a manikin’s ears. The important feature of Kuhn’s model is the prediction that the ratio of the low-frequency interaural time difference to the high-frequency interaural time difference is 3:2. His data suggest that for angles of incidence around 45 deg, the prediction is confirmed, but for angles near 90 deg, there is less variation of interaural time difference with frequency than the model predicts. Figure 10 shows interaural time differences extracted from our HRTF measurements from one subject and three angles of incidence on the azimuthal plane. As can be seen, for a 45-deg incidence, the low-frequency asymptote is about 511  $\mu\text{s}$  and the high-frequency asymptote is about 310  $\mu\text{s}$ , close to the prediction of the Kuhn model. For a 90-deg incidence, the low-frequency asymptote is about 780  $\mu\text{s}$  and the high-frequency asymptote is about 700  $\mu\text{s}$ . The latter is greater than Kuhn’s model would predict ( $\frac{3}{2} * 780 = 520$ ), but is generally consistent with Kuhn’s empirical data.

### III. CONCLUSIONS

This article has described procedures we developed for synthesizing acoustical stimuli that, when presented over

headphones, simulate free-field stimuli. The procedures require measurements of each subject’s free-field-to-eardrum transfer functions, or HRTFs, deep in the subject’s unoccluded ear canals. The method we have developed is fast and produces precise and repeatable measurements. The measurements obtained from ten subjects were shown to be generally consistent with previous data in the literature, in spite of substantial differences in procedure. An objective verification procedure, in which ear-canal waveforms from free-field sources and from headphone-presented simulations are directly compared, shows that the simulation procedure can duplicate free-field waveforms within a few dB of magnitude and a few degrees of phase at frequencies up to 14 kHz.

### ACKNOWLEDGMENTS

This work was supported by grants from the National Institutes of Health (NINCDS) and the Air Force Office of Scientific Research, and by a Joint Research Interchange with NASA. The authors gratefully acknowledge the assistance of Marianne Arruda, Douglas Swiggum, and Linda Pettersson for their contributions to the technical phases of the work, and the helpful comments of Dr. Robert Lutfi on several earlier drafts of this article.

<sup>1</sup>Dividing one spectrum by another in order to “deconvolve” the denominator signal from the numerator signal is a potentially unstable operation. If at any frequency the magnitude of the spectrum in the denominator is very small compared to the magnitude of the spectrum in the numerator, numerical overflow problems can occur. In our application, this could happen if the headphone transfer function had a very deep spectral notch at some frequency. Fortunately, while 20- to 30-dB spectral notches appeared in many of our headphone transfer function (the denominator in our case) measurements, they did not cause computational problems.

<sup>2</sup>In theory, any signal could be used to measure the HRTF as long as its spectrum was nonzero in the frequency range of interest. In pilot work with two of our ten subjects, we evaluated two signals in addition to the signal described in the article: a 20- $\mu\text{s}$  impulse and a flat-spectrum random-phase noise burst. We found that while the three signals produced indistinguishable HRTF estimates in spectral regions where the signal-to-noise ratio was high, the stepped-spectrum, minimum peak factor signal produced more stable estimates in regions of low signal-to-noise ratio.

<sup>3</sup>The filter impulse-response was used as a test stimulus for reasons of convenience. In fact, because this stimulus often had prominent peaks and valleys in its spectrum, it proved to be a severe test of the simulation. Pilot work with two of the ten subjects showed that stimuli with flatter spectra usually produced less error than what was obtained with the impulse-response stimulus.

<sup>4</sup>The acoustical verification experiments were logistically complicated and time consuming. For this reason, a pilot study was used to select the four positions at which the greatest differences were likely. Given the pattern of data obtained from the first six subjects tested, we elected not to test the remaining four subjects and not to test at additional source positions.

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

Blauert, J. (1983). *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT, Cambridge, MA).

Bloom, P. J. (1977). “Creating source elevation illusions by spectral manipulation,” *J. Audio Eng. Soc.* **25**, 560–565.

Butler, R. A. (1975). “The influence of the external and middle ear on auditory discriminations,” in *Handbook of Sensory Physiology*, edited by W. D. Keidel and W. D. Neff (Springer, Berlin).

Butler, R. A., and Belendiuk, K. (1977). “Spectral cues utilized in the localization of sound in the median sagittal plane,” *J. Acoust. Soc. Am.* **61**, 1264–1269.

Butler, R. A., and Planert, N. (1976). “The influence of stimulus bandwidth on localization of sound in space,” *Percept. Psychophys.* **19**, 103–108.

- Fisher, H., and Freedman, S. (1968). "The role of the pinna in auditory localization," *J. Aud. Res.* **8**, 15–26.
- Gardner, M. B., and Gardner, R. S. (1973). "Problem of localization in the median plane: Effect of pinnae cavity occlusion," *J. Acoust. Soc. Am.* **53**, 400–408.
- Hebrank, J., and Wright, D. (1974). "Spectral cues used in localization of sound sources on the median plane," *J. Acoust. Soc. Am.* **56**, 1829–1834.
- Jongkees, L., and Groen, J. (1946). "On directional hearing," *J. Laryngol. Otol.* **61**, 494–504.
- Kuhn, G. (1977). "Model for the interaural time differences in the azimuthal plane," *J. Acoust. Soc. Am.* **62**, 157–167.
- Mehrgardt, S., and Mellert, V. (1977). "Transformation characteristics of the external human ear," *J. Acoust. Soc. Am.* **61**, 1567–1576.
- Oldfield, S., and Parker, S. (1984). "Acuity of sound localisation: a topography of auditory space. II. Pinna cues absent," *Perception* **13**, 6601–617.
- Plenge, G. (1974). "On the difference between localization and lateralization," *J. Acoust. Soc. Am.* **56**, 944–951.
- Searle, C. L., Braida, L. D., Cuddy, D. R., and Davis, M. F. (1975). "Binaural pinna disparity: Another auditory localization cue," *J. Acoust. Soc. Am.* **57**, 448–455.
- Searle, C. L., Braida, L. D., Davis, M. F., and Colburn, S. (1976). "A model for auditory localization," *J. Acoust. Soc. Am.* **60**, 1164–1175.
- Schroeder, M. R. (1970). "Synthesis of low-peak-factor signals and binary sequences with low autocorrelation," *IEEE Trans. Inform. Theory* **IT-16**, 85–89.
- Shaw, E. A. G. (1965). "Ear canal pressure generated by a free sound field," *J. Acoust. Soc. Am.* **39**, 465–470.
- Shaw, E. A. G. (1974). "Transformation of sound pressure level from the free field to the eardrum in the horizontal plane," *J. Acoust. Soc. Am.* **56**, 1848–1861.
- Strutt, J. W. (Lord Rayleigh). (1907). "On our perception of sound direction," *Philos. Mag.* **13**, 214–232.
- Tribolet, J. M. (1977). "A new phase unwrapping algorithm," *IEEE Trans. Acoust. Speech Signal Process* **ASSP-25**, 170–177.
- Watkins, A. J. (1978). "Psychoacoustical aspects of synthesized vertical locale cues," *J. Acoust. Soc. Am.* **63**, 1152–1165.
- Wiener, F. M., and Ross, D. A. (1946). "The pressure distribution in the auditory canal in a progressive sound field," *J. Acoust. Soc. Am.* **18**, 401–408.
- Wightman, F. L., and Kistler, D. J. (1989). "Headphone simulation of free-field listening. II: Psychophysical validation," *J. Acoust. Soc. Am.* **85**, 868–878.