A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction[a)]

Doris J. Kistler
Waisman Center, University of Wisconsin—Madison, Madison, Wisconsin 53705

Frederic L. Wightman
Department of Psychology and Waisman Center, University of Wisconsin—Madison, Madison, Wisconsin 53705

(Received 6 May 1991; revised 29 June 1991; accepted 28 October 1991)

Free-field to eardrum transfer functions (HRTFs) were measured from both ears of 10 subjects with sound sources at 265 different positions. A principal components analysis of the resulting 5300 HRTF magnitude functions revealed that the HRTFs can be modeled as a linear combination of five basic spectral shapes (basis functions), and that this representation accounts for approximately 90% of the variance in the original HRTF magnitude functions. HRTF phase was modeled by assuming that HRTFs are minimum-phase functions and that interaural phase differences can be approximated by a simple time delay. Subjects' judgments of the apparent directions of headphone-presented sounds that had been synthesized from the modeled HRTFs were nearly identical to their judgments of sounds synthesized from measured HRTFs. With fewer than five basis functions used in the model, a less faithful reconstruction of the HRTF was produced, and the frequency of large localization errors increased dramatically.

PACS numbers: 43.66.Qp, 43.66.Yw, 43.66.Pn (WAY)

INTRODUCTION

The major determinants of the apparent direction of a sound are encoded in the acoustical transfer functions from the sound source to a listener's eardrums. These transfer functions are often called "head-related" transfer functions (HRTFs) to acknowledge the primary acoustical importance of the head and pinnae. The phase components of the HRTFs incorporate the interaural time cues thought to be important primarily for determining apparent source azimuth. The gain or magnitude components code the interaural intensity, and spectral shape cues (pinna effects) from which apparent source elevation is thought to be derived.

Most empirical research on human HRTFs has focused on the magnitude components (Mehrgardt and Mellert, 1977; Middlebrooks *et al.*, 1989; Shaw, 1974; Wightman and Kistler, 1989a). These studies suggest three sources of variability in HRTF magnitudes: (1) Frequency: Individual HRTFs contain a tremendous amount of spectral detail (peaks and valleys), especially in the high frequencies, (2) Source position: The center frequencies of the peaks and valleys in the HRTFs change in complicated ways as a function of the azimuth and elevation of the source, (3) Subject: At any given source position, HRTFs vary considerably from subject to subject, especially in the high-frequency regions (Wightman and Kistler, 1989a).

The perceptual importance of the across-frequency, across-position, and across-subjects variability is not fully understood. In the case of across-frequency variability, it is generally agreed that spectral peaks (e.g., Blauert, 1983) and valleys (e.g., Musicant and Butler, 1985) carry important information about sound source position. However, it is unlikely that all the spectral detail in HRTFs is important. One reason is that, at high frequencies, poorer peripheral resolution (wider critical bandwidth, or auditory filter width) would effectively smooth out much of the detail. To the extent that any of the spectral detail in HRTFs is relevant to determining apparent position, the across-position variability in HRTFs is obviously important. Greater changes in the HRTF with changes in source position might be expected to produce more effective cures. The perceptual relevance of the observed differences in HRTFs across subjects is less clear. However, there is some evidence that a given subject's judgments of apparent position are determined to some extent by idiosyncratic features of that subject's own HRTFs (Butler and Belendiuk, 1977; Wenzel *et al.*, 1988).

This article describes a mathematical model of HRTFs that quantifies the observed variations of HRTF magnitude with frequency, source position, and subject. The parameters of the model correspond to spectral features or patterns in the HRTFs. By relating these model parameters to psychophysical data on sound localization, we hope to identify those features of HRTF magnitude functions that serve as localization cues.

Modeling the spectral variation in HRTFs is similar to the problem of modeling the spectral variation in speech sounds. Much of the work with speech sounds suggests that principal components analysis (PCA) can be an efficient and useful way to represent the underlying structure in a highly variable set of spectral data (Li *et al.*, 1969; Plomp *et*

[a)] Portions of this paper were presented at the 120th meeting of the Acoustical Society of America held in San Diego, CA, 26–30 November 1990.

 0001-4966/92/031637-11$00.80

al., 1967; Pols *et al.*, 1969; Pols *et al.*, 1973; Zahorian and Rothenberg, 1981).

PCA decomposes a set of magnitude spectra into basis functions in a manner analogous to the way Fourier analysis decomposes a waveform into sine and cosine components. The basis functions can be considered the basic spectral shapes from which each spectrum in the set is built. Each spectrum can be approximated by a weighted sum of these basis functions. Thus the weights define the relative contribution of each basis function (basic spectral shape) to the spectrum. The entire collection of weights that represents the contribution of a given basis function to each of the magnitude spectra in the set is called a "principal component" or PC.

A recent application of PCA to the problem of modeling HRTFs was reported by Martens (1987). Martens computed basis functions and principal components from critical-band-filtered HRTFs measured from 35 source positions on the horizontal plane. Prior to the PCA, the HRTF magnitudes in each critical band were normalized to eliminate direction-dependent level differences in HRTFs. For the four basis functions and PCs described, Martens noted a systematic variation of the weights as source azimuth varied from one side of the head to the other and from front to rear.

The two experiments we describe here are intended to evaluate more extensively the utility of the PCA approach to modeling HRTFs. The first involves PCA analysis of HRTF measurements from 10 listeners at 265 source positions. The second is a psychophysical experiment in which we assess the perceptual validity of the PCA model. HRTFs are reconstructed from the principal components, and these model HRTFs are used to synthesize stimuli that, when presented over headphones, simulate real free-field sources. Listeners' judgments of the apparent directions of synthesized and real free-field sources are compared as a way of determining the perceptual validity of the PCA model.

I. EXPERIMENT I: PRINCIPAL COMPONENTS ANALYSIS OF HRTFS

A. Subjects

Ten young adults (3 male, 7 female) served as paid participants. All had normal hearing and no history of hearing problems. Five of the subjects also served in experiment II.

B. HRTF measurement procedure

Our procedure for measuring HRTFs is described in detail in Wightman and Kistler (1989a), so will only be summarized here. The only substantive differences between the previous and current procedures are the number of sources used and the details of the measuring signal. The subject is seated (head unrestrained) on a stool in the center of an anechoic chamber. A wideband (0.15–18 kHz), periodic (10.24-ms period) noiselike signal, is presented continuously, transduced by one of 23 miniature loudspeakers (Realistic Minimus 3.5) mounted on a vertical semicircular arc 1.4 m from the subject. The signal is output at a 50-kHz sample rate via a 16-bit D/A converter controlled by an IBM-PC. The response at the listener's eardrums is obtained by averaging (500 periods of the signal) the output of a probe microphone (Etymotic), the tip of which is held in place, a few mm from the eardrum, by a customized Lucite earmold shell.

The signal used to make HRTF measurements was computed via an inverse FFT procedure. The magnitude and phase of the signal spectrum were computed to maximize the signal-to-noise ratio in the response recordings. The magnitude increased at a rate of 3 dB per octave from 150 Hz to approximately 18 kHz and decreased sharply from 18 to 20 kHz. The signal contained no energy below 150 Hz or above 20 kHz. The phase of the signal spectrum was constructed to minimize the peak factor of the signal (Schroeder, 1970).

HRTFs were measured at 265 source positions: directly over the subject's head and at 264 other positions given by the combination of 11 elevations, ranging from $-48°$ to $+72°$ in 12° increments, and 24 azimuths, ranging from $-165°$ to $+180°$ in 15° increments. A typical measurement session lasted approximately 1.5 h.

C. Principal components analysis

Principal components analysis is a statistical procedure that attempts to provide an efficient representation of a set of correlated measures. The central idea of PCA is to reduce the dimensionality of a data set in which there are a large number of interrelated measures, while retaining as much as possible of the variation present in the data. A small set of basis functions is derived, and these basis functions are used to compute the principal components. Recall that the principal components are the sets of weights that reflect the relative contributions of each basis function to the original data. The basis functions are derived in such a way that the first function and its weights (which we designate PC-1) capture the majority of common variation present in the data and that the remaining functions and weights (PC-2, PC-3, etc.) reflect decreasing common variation and increasing unique variation. The number of basis functions required to provide an adequate representation of the data is largely a function of the amount of redundancy or correlation present in the data. The greater the redundancy, the smaller the number of basis functions needed.

We derived the basis functions and PCs from the HRTF magnitudes in a single analysis that included left- and right-ear log-magnitude functions of all 10 subjects (a total of 5300 HRTFs). Only the 150 log magnitudes in the 0.2- to 15-kHz frequency region were included in the analysis. Prior to the PCA, mean (across the 265 positions) HRTF log-magnitude functions were computed for the measurements from each ear of each subject. These mean functions include the subject-dependent and direction-independent spectral features shared by all 265 HRTFs recorded from an individual ear (e.g., the ear canal resonance at about 2.5 kHz). They also include measurement artifacts such as the spectral notches caused by standing waves (Wightman and Kistler, 1989a). To remove these features, the appropriate mean function was subtracted from each HRTF. With means removed, the resulting 5300 log-magnitude functions represent primarily direction-dependent spectral effects. We call

these functions "directional transfer functions" or DTFs, to distinguish them from HRTFs which include both directional and nondirectional (e.g., ear canal resonance) spectral effects. It is the set of 5300 DTFs that was subjected to a PCA.

The first step in the principal components analysis is computation of a frequency covariance matrix. These covariances provide a measure of similarity across the 5300 DTFs for each pair of frequencies. The covariance S for a given pair (i,j) of frequencies is given by

$$S_{ij} = (1/n)[\Sigma D_{ki} D_{kj}], \quad \text{for } i,j = 1,2,...,p, \tag{1}$$

where n is the total number of transfer functions (5300 in this case), p is the total number of frequencies (150 in this case), and D_{ki} is the log magnitude at the ith frequency of the k th directional transfer function.

The basis functions (more correctly basis vectors, since we are dealing with discrete representations), $\mathbf{c}_q$, are the q eigenvectors of the covariance matrix that correspond to the q largest eigenvalues. For a given DTF, the weights representing the contribution of each basis function to that DRF, are given by

$$\mathbf{w}_k = \mathbf{C}'\mathbf{d}_k, \tag{2}$$

where $\mathbf{C}$ is a matrix, the columns of which are the basis vectors, and $\mathbf{d}_k$ is the k th DTF magnitude vector. Note that if the terms in Eq. (2) are rearranged, the DTF magnitude vector is equal to a weighted sum of the basis vectors:

$$\mathbf{d}_k = \mathbf{C}\mathbf{w}_k. \tag{3}$$

However, this equality holds only if $q = p$, or the maximum possible number of eigenvectors and basis vectors are retained. In practice, $q \ll p$, and thus the right side of Eq. (3) provides only an estimate of $\mathbf{d}_k$.

There is considerable inconsistency in the terminology used to describe the basis vectors (columns of C above) and the weights w that result from principal components analysis. Technically, the term "principal component" refers to the entire set of weights (one for each DTF being modeled) associated with a particular basis vector. Thus PC-1 refers to the weights associated with the first basis vector, PC-2 to the weights associated with the second basis vector, etc. However, it is often convenient to use the term principal component to denote the set of weighted basis vectors (or a single weighted basis vector). Thus, depending on context, we may use the term "PC-1" to refer either to the weights or to the first-weighted basis vector. Where it is important to be precise, we will use the terms "PC-1 weights" to refer to the weights alone, and "PC-1 basis vector" to denote the basis vector alone.

D. Results

PCA provides a convenient means for quantifying the amount of spectral detail present in DTFs. As Eq. (3) indicates, each DTF can be reconstructed from a linear combination of principal components. (Reconstruction of an entire HRTF requires addition of the grand mean and the appropriate subject-ear mean to the reconstructed DTF.) The amount of detail in the reconstructed DTF is determined by the number of PCs used in the reconstruction. As stated above, the original DTFs would be reproduced exactly if 150 PCs were derived and retained in the reconstruction (the maximum number of PCs is equal to the number of spectral points in each DTF). However, our objective is to reduce the number of PCs required to represent a DTF. Figure 1 shows the results of the reconstruction process for a DTF measured at the left ear of a representative subject for a source at $-30°$ azimuth and 12° elevation (on the left side, above the horizontal plane). The solid line in each panel of this figure is the log-magnitude spectrum of the measured DTF and the dotted line corresponds to the reconstructed DTF. In the top panel it is apparent that the DTF reconstructed using only PC-1 bears little resemblance to the measured DTF. As additional PCs are included, the amount of spectral detail increases, as does the similarity of the reconstructed and measured DTFs.

The decision of how many PCs to retain is traditionally based on some estimate of how much of the variance in the original data must be recovered. We chose, arbitrarily, to extract the number of PCs necessary to account for approximately 90% of the variance in the DTF magnitude functions. (The total variance in the collection of 5300 DTFs may be computed by summing the 150 eigenvalues.) Applying this criterion, we retained five PCs. The percentage of

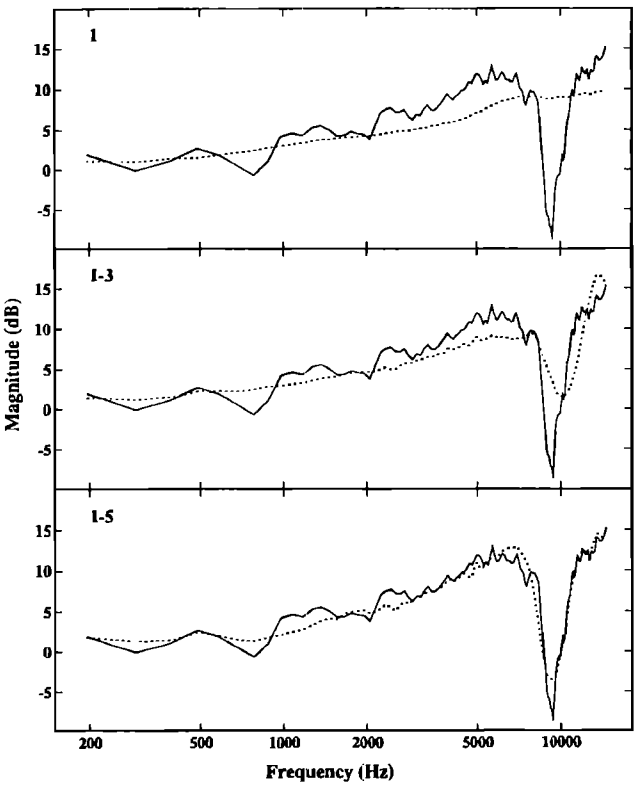

FIG. 1. Measured (solid lines) DTFs (HRTFs with means removed) from the left ear of a single subject with the sound source at $-30°$ azimuth and 12° elevation, and DTFs reconstructed from one principal component (top panel), three principal components (middle panel), and five principal components (bottom panel).

total variance accounted for by PCs 1–5 is 74.3, 6.6, 4.5, 2.7, and 2.2, respectively. To assess how well the reconstruction captured the variation in DTF magnitude functions for individual subjects, we computed percentage of variance accounted for in the left- and right-ear DTF data from each subject. These percentages are given in Table I. The data show similar trends across subjects, suggesting that the first five PCs capture features of the DTFs that are common to all subjects.

The five principal component basis functions are plotted in Fig. 2. Note that all five functions are roughly constant and close to zero at frequencies below 2–3 kHz. This reflects the fact that there is little direction-dependent variation in the DTFs in this frequency region. Regardless of the weights applied to the basis functions, the resulting weighted sum will be close to zero in this region. Above about 3 kHz, all five basis functions have nonzero values. It is clear that with the exception of the first function, the high-frequency variation in these basis functions represent the direction-dependent high-frequency peaks and notches in the DTFs.

Given that DTFs can be represented by a relatively small number of basic spectral shapes, it seems reasonable to expect that the amount each basic shape contributes to the DTF at a given source position would relate,in some simple way, to source azimuth and elevation. Figure 3 shows the PC-1 weights for a representative subject's DTFs, plotted as a function of source azimuth. The left-ear weights for the 265 source positions are plotted in the top panel, and the right-ear weights are plotted in the bottom panel. Note that there is a tendency for the weights to increase in magnitude as the source moves from the median plane. Ipsilateral sources

TABLE I. Percentage of variance in the DTFs from 10 subjects explained by 5 principal components.

ID	Ear	PC-1	PC-2	PC-3	PC-4	PC-5	Total
SDL	L	74	7	6	3	2	92
	R	74	6	5	3	2	90
SGE	L	70	11	2	5	2	90
	R	72	7	4	4	3	90
SGG	L	73	8	3	4	2	90
	R	75	7	4	3	2	91
SHG	L	79	5	5	2	2	93
	R	78	5	5	2	2	92
SIK	L	76	5	5	3	3	92
	R	77	6	4	3	2	92
SKR	L	72	6	6	5	2	91
	R	74	7	4	4	1	90
SKS	L	72	8	4	3	3	90
	R	74	6	5	2	3	90
SKT	L	74	8	5	2	2	91
	R	71	7	6	3	2	89
SLG	L	80	5	4	2	2	93
	R	80	5	4	2	2	93
SLN	L	70	5	6	3	4	88
	R	71	6	5	3	3	88

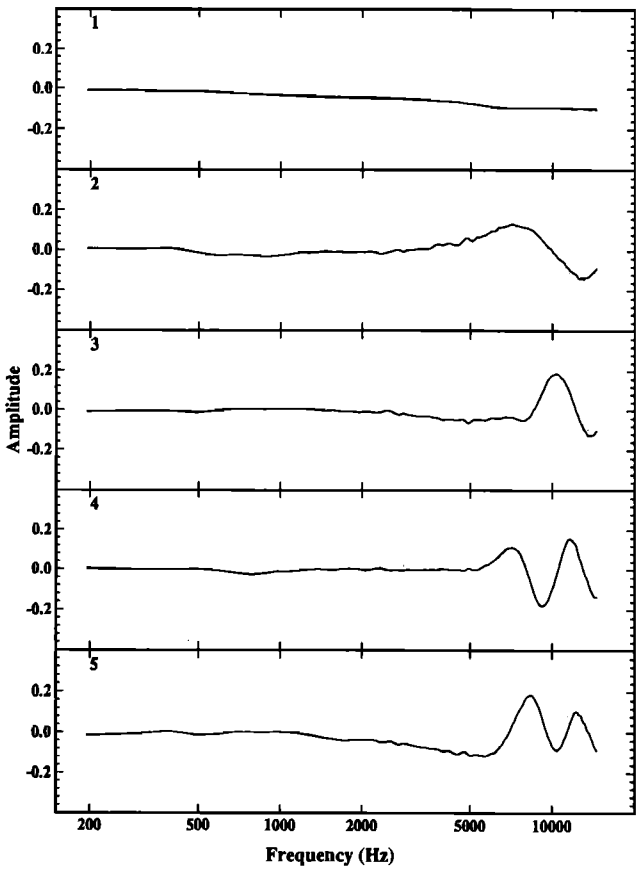

FIG. 2. The first five basis functions extracted from a principal components analysis of DTFs from both ears of 10 subjects at 265 source positions.

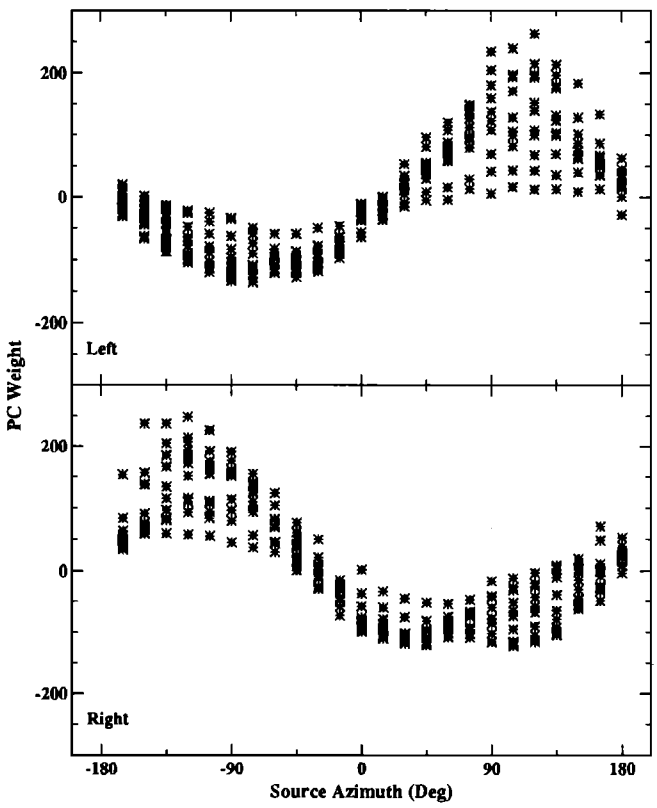

FIG. 3. PC-1 weights for a representative subject's DTFs plotted as a function of source azimuth. Weights for left-ear DTFs are in the top panel and weights for right-ear DTFs are in the bottom panel.

(sources with negative azimuths in the top panel and positive azimuths in the bottom panel) have negative weights, while contralateral sources have positive weights. Though not apparent from the figure, sources at the highest elevations also have weights near 0. This distribution of PC-1 weights is very similar across the 10 subjects. Figure 4 illustrates the low intersubject variability. In this figure, we plot the mean and standard deviation of PC-1 weights for all sources on the horizontal plane.

In order to interpret the regular variation of PC-1 weights with source azimuth, consider the form of the first basis function (Fig. 2, top panel). Note that since the trend of this function is toward more negative values as frequency increases, a positive weight on this function would lead to a contribution of negative log magnitudes at high frequencies, thus deemphasizing the high frequencies (i.e., with a gradual slope) in the DTF. Conversely, a negative weight would emphasize the high frequencies. Given that the first basis function accounts for more than 70% of the total variance, we can see that the major difference between ipsilateral and contralateral DTFs is the slope of the log-magnitude function in the high frequencies. Ipsilateral DTFs will have a more positive slope than contralateral DTFs. This suggests that the major source of variance among the DTFs is the amount of energy in the high frequencies. Since the mean HRTF is removed, the DTFs will have roughly equal energy in the low frequencies. The correlation between total DTF energy and PC-1 weights is -0.96.

The remaining four basis functions are less amenable to interpretation, but taken together, they seem to capture the high-frequency spectral variation (see Fig. 2) resulting from changes in source elevation. They also reflect spectral differences between sources in front and sources behind the subject. However, neither elevation changes nor front/back distinctions can be ascribed to a single function. Figure 5 presents PC-2 through PC-5 weights derived from the right-ear DTFs of a representative subject plotted as a function of source elevation. Weights associated with DTFs measured in front of the subject and behind are plotted with different symbols. Consider first the PC-2 weights (top panel of Fig. 5). While not as convincing as with the PC-1 weights, there are some notable trends. DTFs from rear positions (diamonds) are given positive weights at high elevations and zero weight at low elevations. Thus in DTFs from rear positions the broad peak in the PC-2 basis function at around 7 kHz (Fig. 2) would be more prominent at high elevations than at low elevations. For DTFs from positions in the front, the situation is reversed in two ways: Low elevations are given negative weights, and high elevations zero weight. The pattern of PC-3 weights shown in Fig. 5 suggests that the PC-3 basis function can also distinguish front- from rear-source positions, but only at low elevations. In this case, rear positions are given positive weights (thus emphasizing the peak at 10 kHz in the PC-3 basis function) at low elevations, while front positions are given negative weights (thus adding a notch at 10 kHz to the resulting DTF). The pattern of PC-4 and PC-5 weights shown in Fig. 5 is considerably more complicated, so no interpretation will be offered here.

Although there was greater intersubject variability in the distributions of the higher-order PC weights, the patterns were roughly similar across subjects and ears. As an

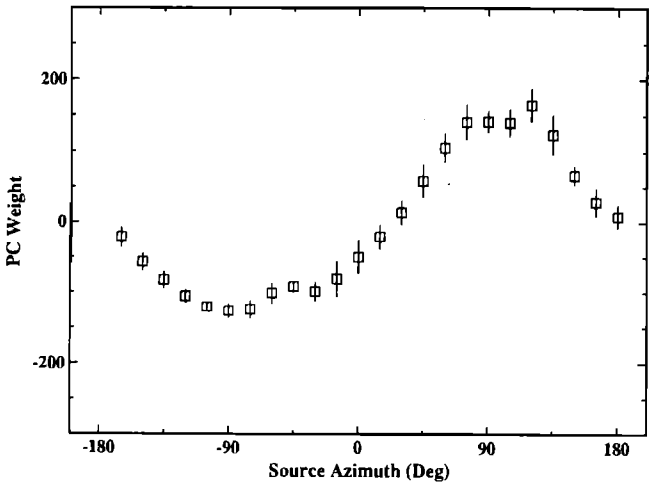

FIG. 4. Means (data points) and plus/minus 1 standard deviation (error bars) of PC-1 weights for all left ear DTFs (10 subjects) from sources on the horizontal plane.

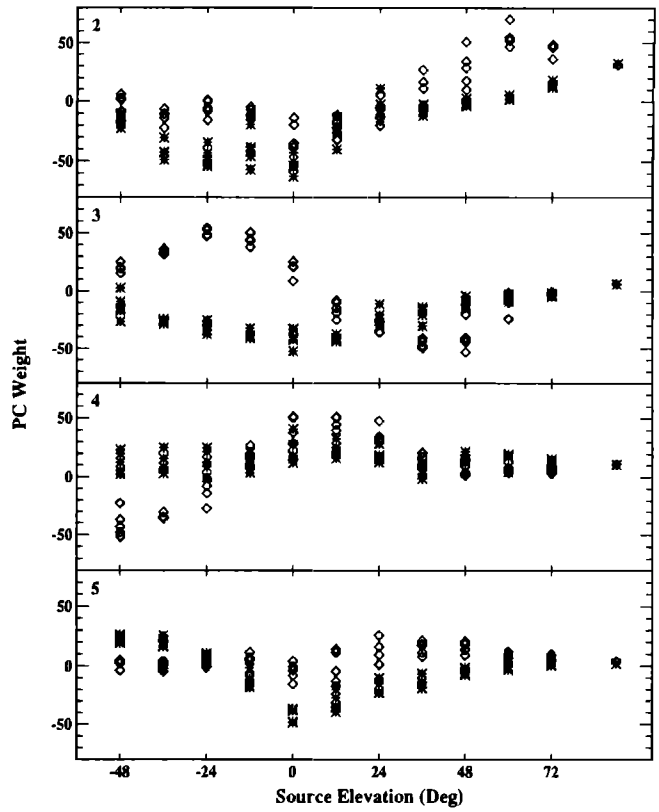

FIG. 5. PC weights for PCs 2–5 extracted from a representative subject's DTFs plotted as a function of source elevation. Weights for rear DTFs (azimuths from 120° to 180° and from $-120°$ to $-180°$) are plotted with diamonds, and weights for front DTFs (azimuths from 60° to $-60°$) are plotted with asterisks.

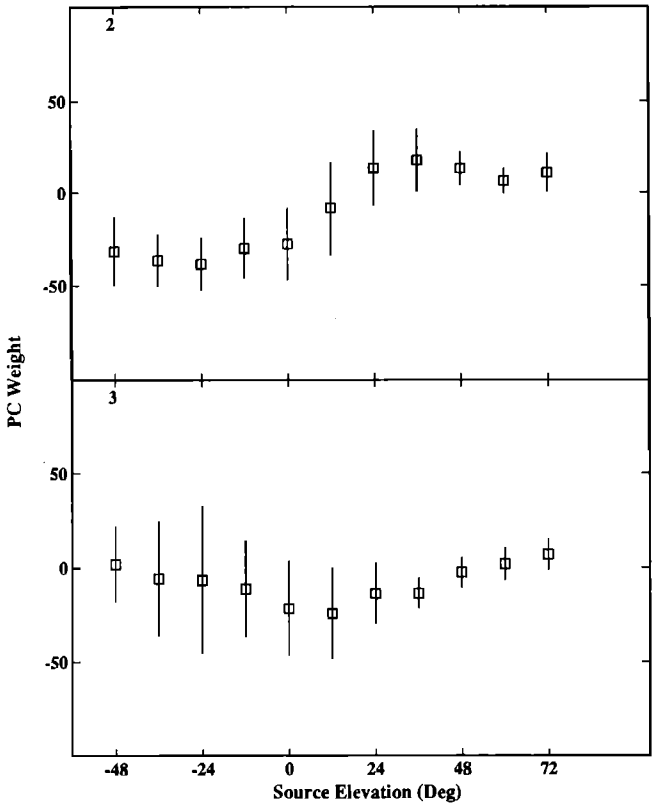

FIG. 6. Means (data points) and plus/minus 1 standard deviation (error bars) of PC-2 and PC-3 weights for all left ear DTFs (10 subjects) from sources at an azimuth of 45°.

example, means and standard deviations of the PC-2 and PC-3 weights for all sources at an azimuth of 45° are plotted in Fig. 6.

II. EXPERIMENT II: PERCEPTUAL VALIDATION OF THE PCA MODEL

Figure 1 shows that with only five of the 150 possible basis functions used in the reconstruction, the differences between original and reconstructed DTFs are small. This implies that HRTFs (DTFs with means added back) can be effectively modeled as linear combinations of five basic spectral shapes. If this model is perceptually valid, then an HRTF reconstructed from the five functions should be equivalent to the original HRTF, as far as sound localization is concerned. The most convenient way of testing this prediction is by comparing listeners' judgments of the apparent directions of sounds synthesized from original and modeled HRTFs.

Our previous work with localization of headphone-presented sounds that simulate free-field sounds (Wightman and Kistler, 1989b) suggests that when a subject's own measured HRTFs are used in the synthesis, the apparent directions of the synthesized sounds are nearly the same as the apparent directions of sounds actually presented in free field. The question here relates to the degree of correspondence between direction judgments of sounds synthesized from modeled HRTFs and sounds synthesized from measures HRTFs.

Our PCA approach models only the magnitude components of the HRTFs, but synthesis of stimuli requires both magnitude and phase components. Therefore, we constructed a simple model of the phase components of HRTFs, based on two assumptions. First, we assume that HRTFs are "minimum-phase" functions (Oppenheim and Schafer, 1975). Indirect evidence supporting this assumption comes from the HRTF measurements of Mehrgardt and Mellert (1977). Second, we assume that the frequency dependence of interaural time delay (ITD), reported by Kuhn (1977) and by Wightman and Kistler (1989a), is of no perceptual relevance. Thus we assume that the ITD at each source position can be represented by a constant. The validity of these two assumptions is tested in the control condition of this experiment.

A. Subjects

Five of the subjects that participated in experiment I also served in experiment II. One of these had also had previous experience in similar psychophysical experiments.

B. Stimuli

The general features of our stimulus-generation procedure have been described in detail elsewhere (Wightman and Kistler, 1989b), so only the essential elements are summarized here. The basic stimulus is a 250-ms burst of white Gaussian noise (20-ms cosine-squared onset–offset ramps), repeated 8 times, with 300 ms of silence between repetitions. In contrast with our previous work, the spectrum of the noise stimulus was not scrambled, so that potentially subtle spectral features would not be obscured by trial-to-trial stimulus variability. The characteristics of a given spatial location are superimposed on the noise bursts by filtering them with the HRTFs (either measured or PC derived) appropriate to that location. Next the HRTF-filtered noises are filtered by the inverse of the headphone transfer functions, to compensate for the characteristics of the headphone, and are bandpassed with an 512-tap zero-phase FIR filter (0.2–14 kHz) to eliminate processing artifact.

The experiment includes five conditions in which stimuli are synthesized and presented over headphones. In one, which we call the "baseline" condition, stimuli are synthesized from measured HRTFs. Thus this condition is identical to the headphone condition that we previously reported to have produced localization judgments similar to those obtained with free-field stimuli (Wightman and Kistler, 1989b). The second condition, a "control," was included as a test of the phase model described above. For this condition, stimuli were synthesized using HRTFs with magnitude components from measured HRTFs and with phase components derived according to our model of the phase of HRTFs. The phase functions for a given left-right pair of HRTFs are modeled by computing the left and right minimum phase functions from the log-magnitude functions (Oppenheim and Schafer, 1975), and adding a pure delay to the left minimum phase function. The right minimum phase

function is used without modification. To estimate the appropriate interaural delay, we compute the cross-correlation function (via an inverse Fourier transform of the cross spectrum of the measured left- and right-ear HRTFs) and determine the delay corresponding to the maximum in the cross-correlation function. This time delay is converted to phase and added to the minimum phase function derived from the left HRTF. In the experimental conditions, stimuli are synthesized using HRTFs with PC-derived magnitudes and with phase functions estimated from the PC-derived magnitudes using the method described above. These conditions differ in terms of the number of basis functions used in the HRTF reconstruction (five, three, or one).

The stimuli for each subject and each run were precomputed on a MicroVax 3500 and stored on disk. Stimuli were presented over headphones (Sennheiser HD-340) at an overall level of about 70 dB SPL. Stimulus presentation was controlled by an IBM-PC, with 16-bit D/A converters operated at a 50-kHz sample rate. No antialiasing filters were used.

C. Procedure

The primary purpose of the psychophysical experiments is to compare the apparent directions of sounds filtered by real HRTFs to the apparent direction of sounds filtered by PC-derived HRTFs. In addition, comparison of the localization performance in the "baseline" and "control" conditions allows us to evaluate the validity of the simplifying assumptions we made about phase components of HRTFs. The psychophysical procedure is the same as was used previously to compare performance with free-field stimuli and headphone-presented stimuli (Wightman and Kistler, 1989b). This procedure requires subjects to indicate the apparent direction of a simulated free-field sound source, presented over headphones, by calling out numerical estimates of azimuth and elevation, using standard spherical coordinates. The subject is instructed to give a positive azimuth for sounds appearing on the right side and a negative azimuth for those on the left side. A positive elevation judgment indicates sounds appearing to be above ear level and a negative judgment indicates sounds appearing to be below ear level.

Subjects are tested either in an anechoic chamber or a sound booth (pilot experiments indicate no performance difference between testing sites). On each trial the subject listens to the stimulus and reports estimates of azimuth and elevation. The experimenter, who is in an adjacent room listening over an intercom, enters the responses on the IBM-PC keyboard.

During a single run, subjects judged the directions of simulated sources from 36 pseudorandomly selected positions with a 360° range of azimuths, and a 108° range of elevations ($-48°$ to $+60°$). Each subject completed 10 runs per condition, with a different randomized order of stimuli on each run. Six were completed in a single session, which lasted approximately 90 min. Each subject was tested first in the baseline condition, followed by the control condition, and the 5-, 3-, and 1-basis function conditions.

All subjects were also tested with real free-field stimuli before participating in the main experiment (as in Wightman and Kistler, 1989b). A comparison of the free-field judgments to the judgments in the baseline condition allows us to assess the validity of our free-field simulation procedures, for these subjects. The free-field results, which are not presented here, are wholly consistent with the results of our earlier study (Wightman and Kistler, 1989b) and suggest that for these subjects, the basic free-field simulation technique is successful.

D. Results

Recall that each subject provided 10 judgments of azimuth and elevation for each of the 36 source locations in each stimulus condition. Since apparent distance is not measured, we assume that it is constant (an assumption consistent with informal reports from the subjects), and thus that the judgments can be represented as points on the surface of a sphere. The use of common spherical statistics (Fisher *et al.*, 1987) to compute measures of central tendency and variance assumes that the data are unimodally distributed. This assumption is frequently violated with localization data because of the presence of front–back confusions. A front–back confusion is thought to occur when, for example, a subject reports an apparent azimuth of 170° for a stimulus actually presented at an azimuth of 10°, or reports an azimuth of 30° with the stimulus presented at 120°. Front–back (and even up–down) confusions are not unexpected, given the importance of the interaural time difference cue and the roughly spherical symmetry of the head (recall the infamous "cone-of-confusion" described by Mills, 1972). Researchers usually adopt one of two strategies for handling confusions. Responses that appear to represent confusions are either eliminated from the data set and analyzed separately (Makous and Middlebrooks, 1990), or confusions are "resolved," by reflecting the azimuth component of the response around 90° (Oldfield and Parker, 1984; Wightman and Kistler, 1989b). In most cases, since confusions represent a small fraction of the data set, the choice of strategy is unimportant. In some conditions of this experiment, however, subjects produced substantial numbers of front–back and up–down confusions. As a result, we chose not to resolve or eliminate confusions, but rather to present and analyze the raw data.

To facilitate interpretation of the results, we represent the azimuth component of each judgment in terms of two angles: right–left, the angle subtended by the judgment vector and the median plane, and front–back, the angle subtended by the judgment vector and the transverse plane (the vertical plane that passes through the ears). The right–left angle, which is equivalent to the azimuth angle in the "double-pole" coordinate system favored by some investigators (Middlebrooks *et al.*, 1989), expresses only the extent of laterality of a judgment, since directions in the front and rear hemispheres are not distinguished. Consequently, a judgment of 30,0 (30° azimuth, 0° elevation) has the same right–left angle (30°) as a judgment of 150,0. Right–left angles of 90° (right) or $-90°$ (left) represent the extremes on this dimension. The front–back angle distinguishes between judgments in the front and rear hemispheres, but does not reflect differences between directions on the left and direc-

tions on the right. Thus judgments of 15,0 and − 15,0 are assigned the same front–back angle (75°), as are judgments of 165,0 and − 165,0 (− 75°). On the front–back dimension, the extremes are represented by angles of 90° (front) and − 90° (rear). To maintain consistency in the terminology, we will refer to the elevation component of the judgments as an up–down angle, the angle subtended by the judgment vector and the horizontal plane.

Figures 7 and 8 show the results from a representative subject in the baseline and control conditions. In these figures, the right–left, front–back, and up–down components of the judgments are plotted as a function of the coordinates of the targets. There are 360 judgments represented in each panel, 10 judgments of the apparent directions of 36 sources.

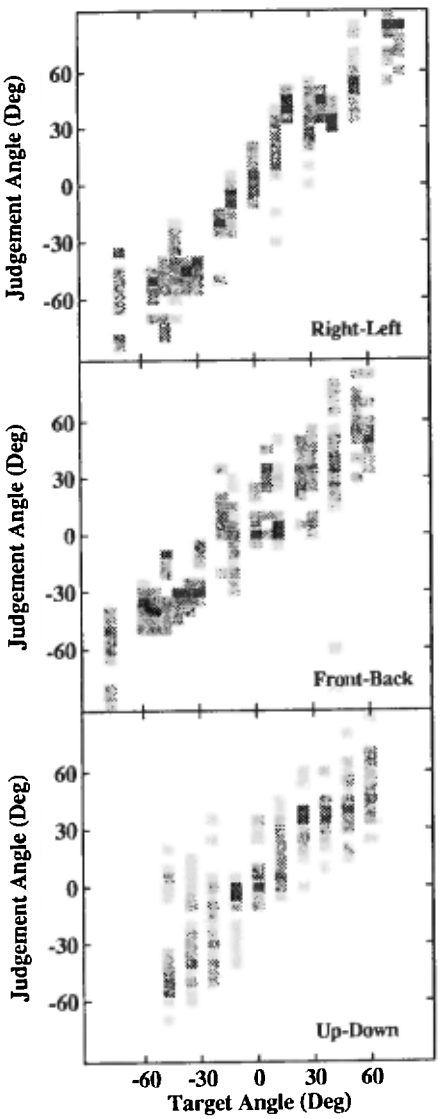

FIG. 7. Scatterplots of judged direction versus target direction from one subject in the baseline headphone condition of experiment II. As described in the text, the judged and target directions are represented in terms of three angles, right–left, front–back, and up–down. Each data cell includes all the judgments within a 5° wide interval. The darker the cell, the more judgments represented in that cell. The lightest cells represent a single judgment. There are 360 judgments shown in each panel, corresponding to the 10 judgments made to each of 36 stimuli.

In these displays the judgments are grouped in bins 5° wide. The darkness of each cell reflects the number of judgments falling in each bin. The lightest cells indicate a single judgment. If the judged directions were identical to the target directions, then the cells would lie on a diagonal line drawn between the lower left corner and the upper right corner of each panel. The data in these figures show that the pattern of judgments in the control condition, in which stimuli were synthesized from HRTFs with modeled phases, is very similar to the patterns of judgments in the baseline condition, in which measured HRTFs were used for stimulus synthesis. This result suggests that HRTF phase can be adequately represented by a combination of minimum-phase functions and a pure time delay.

Figures 9, 10, and 11 show the results obtained from the same subject in the experimental conditions, in which the stimuli were synthesized from PC-derived HRTFs. Note first that the right–left components of the judgments are the

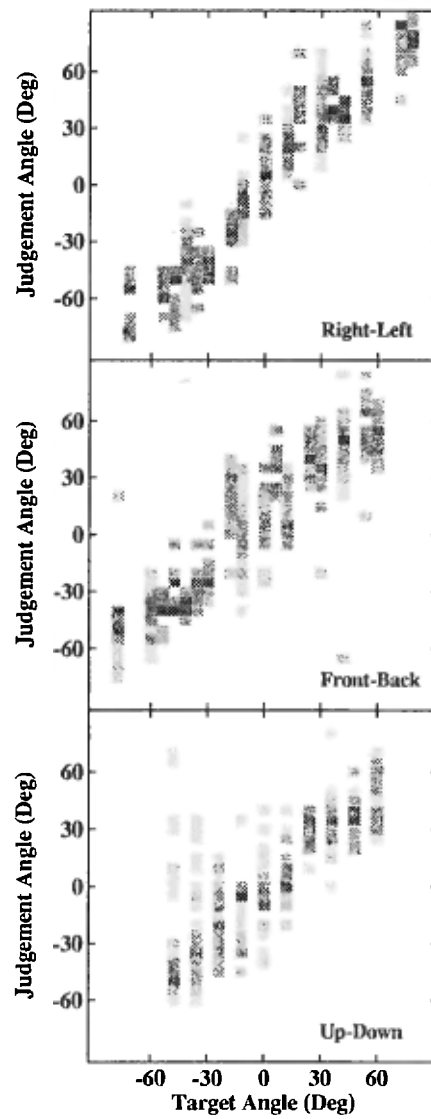

FIG. 8. Same as Fig. 7, but data are from the control (minimum-phase) condition.

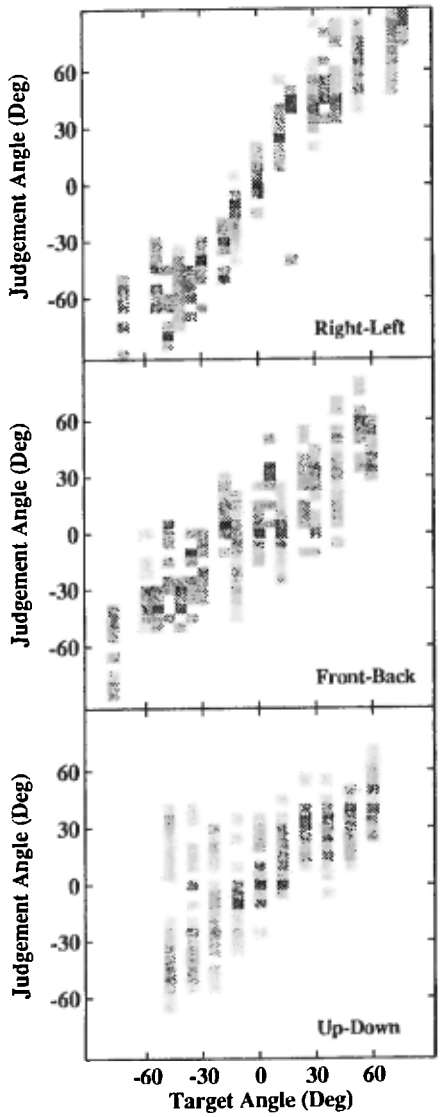

FIG. 9. Same as Fig. 7, but data are from the PC:1–5 condition, in which HRTFs were reconstructed from the first five basis functions.

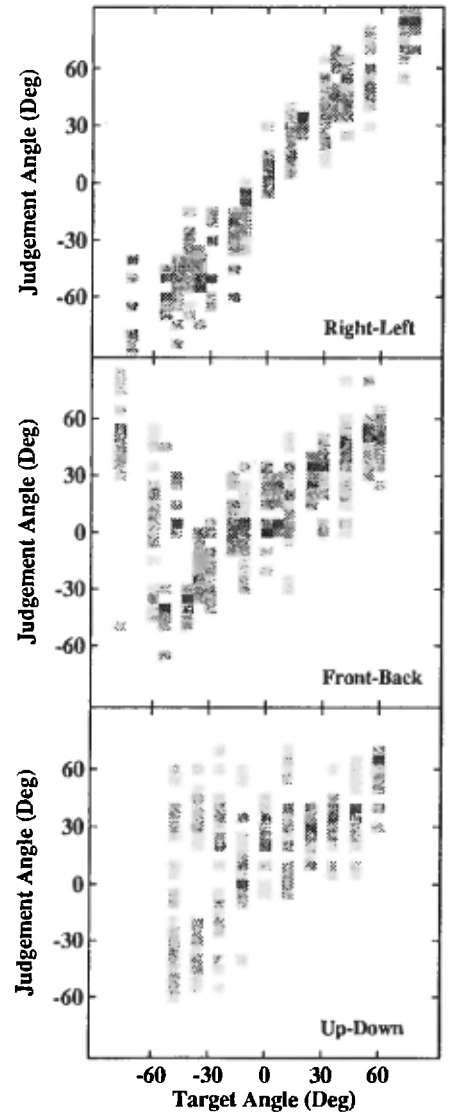

FIG. 10. Same as Fig. 7, but data are from the PC:1–3 condition, in which HRTFs were reconstructed from the first three basis functions.

same in all three experimental conditions, and indistinguishable from the right–left components of the judgments in the baseline (Fig. 7) and control (Fig. 8) conditions. This suggests that judgments of laterality (right–left) are accurate even if HRTFs are reconstructed from PC-1 alone. PC-1 provides information only about gross interaural intensity differences (interaural time differences are provided in the phase spectra which are modeled as in the control condition). Note also that as successively more primitive representations of the HRTFs are used (from PC:1–5 to PC:1) both front–back and up–down performance deteriorates. Along the front–back dimension, the primary indicator of degradation is an increase in front–back confusions, manifest here by "front" (positive angles) responses to "rear" (negative angles) targets. The front–back confusion rate increased for all subjects as fewer PCs were used to reconstruct HRTFs, but subjects differed in terms of the "preferred hemisphere." Three subjects (including the one whose data are shown in Figs. 7–11) tended to respond with primarily front directions, and two responded with primarily rear directions. Along the up–down dimension, we also note an increase in confusions as fewer PCs are used in the HRTF reconstruction process. Up–down confusions can be seen in Fig. 9 (PC:1–5) and 10 (PC:1–3), as high (positive angles) responses to low (negative angles) targets. With PC-1 alone represented in the HRTF (Fig. 11), discrimination along the up–down dimension is quite poor.

To quantify the effects on localization performance of reduced fidelity in the modeled HRTFs, we computed correlations between target and judgment angles on each of the three dimensions, right–left, front–back, and up–down. The correlations for all subjects are shown in Fig. 12. Unfortunately, the correlation statistic by itself does not allow us to distinguish between error variance and confusions, since a

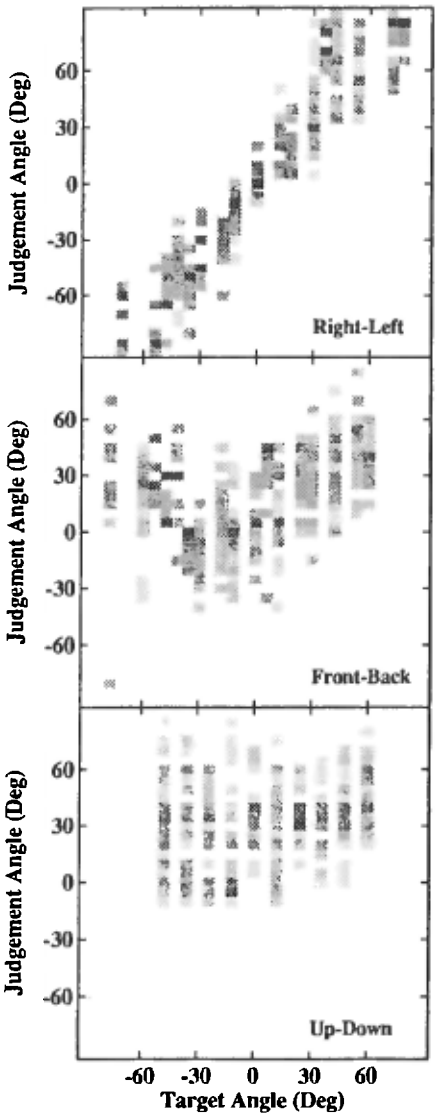

FIG. 11. Same as Fig. 7, but data are from the PC:1 condition, in which HRTFs were reconstructed from only the first basis function.

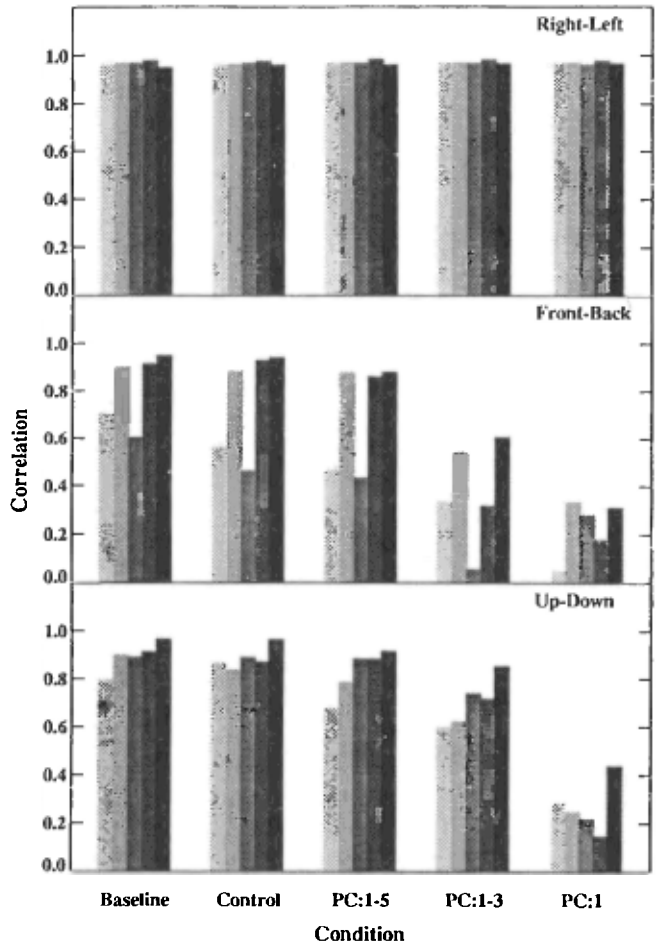

FIG. 12. Correlations between target and judged direction from all five subjects and all five conditions of experiment II. As described in the text, direction is represented by three angles, right–left, front–back, and up–down. Data from an individual subject are coded by the shading of the bars.

decrease in correlation could have resulted from either an increase in variance or an increase in confusions. However, a visual inspection of the data reveals that in the case of the front–back dimension, the primary determinant of a low correlation is an increase in front–back confusions. Along the up–down dimension, both error variance and up–down confusions seem to contribute to the low correlations.

The increase in confusion rates in the PC:1–3 and PC:1 conditions suggests that subjects normally resolve front–back and up–down ambiguities by analyzing the fine spectral detail provided by their own HRTFs. Such spectral detail is not faithfully represented by the modeled HRTFs in these conditions. This hypothesis about the use of fine spectral detail is consistent with data obtained from experiments in which subjects localized stimuli synthesized from another subject's HRTFs (Wenzel *et al.*, 1988). Confusion rates are much higher in these conditions than in normal free-field listening conditions. It also agrees with our informal observations that when some features of the HRTF magnitude spectrum are not well represented (e.g., as produced by certain kinds of errors in our synthesis procedures), confusion rates increase dramatically.

III. CONCLUSION

The results reported here confirm the utility of the PCA model for representing those features of human HRTFs that are important for sound localization. HRTFs can be adequately approximated by a linear combination of five basic spectral shapes or basis functions. The first basis function represents the gradual high-frequency emphasis or deemphasis in the HRTFs that accompanies changes in the azimuth and elevation of the sound source. Of greater interest are basis functions 2–5, which represent the high-frequency peaks and valleys in the HRTFs. Our results support the hypothesis that these spectral features are important primarily for resolution of front–back and up–down confusions. The weights applied to these basis functions are systematically related to source direction along the front–rear and up–

down dimensions, although no single function can account for the complex spectral changes that result from changes in source direction along these dimensions.

The psychophysical results argue strongly that the only cues needed for accurate determination of source laterality are interaural time differences and the gross interaural intensity differences that are provided by the first basis function. The results also suggest that basis function 2–5 probably mediate the distinction of sources in the front from sources in the rear, discrimination of directions along the vertical dimension, and the distinction of sources above the horizontal plane from sources below the horizontal plane.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the assistance of Ms. Marianne Arruda and Ms. Ramona Agrawal in the technical phases of the work, and the very helpful comments of Dr. John Middlebrooks on an earlier version of this paper. Financial support was provided by NASA (Cooperative Agreement NCC2-542) and the NIH-NIDCD (DC00116).

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

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.

Fisher, N. I., Lewis, T., and Embleton, B. J. J. (1987). *Statistical Analysis of Spherical Data* (Cambridge U.P., England).

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

Li, K. P., Hughes, G. W., and House, A. S. (1969). "Correlation characteristics and dimensionality of speech spectra," J. Acoust. Soc. Am. **46**, 1019–1025.

Makous, J. C., and Middlebrooks, J. C. (1990). "Two-dimensional sound localization by human listeners," J Acoust. Soc. Am. **87**, 2188–2200.

Martens, W. L. (1987). "Principal components analysis and resynthesis of spectral cues to perceived direction," in *Proceedings of the International Computer Music Conference*, edited by J. Beauchamp (International Computer Music Association, San Francisco, CA), pp. 274–281.

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

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.

Mills, W. (1972). "Auditory localization," in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York).

Musicant, A. D., and Butler, R. A. (1985). "Influence of monaural spectral cues on binaural localization," J. Acoust. Soc. Am. **77**, 202–208.

Oldfield, S. R., and Parker, S. P. A. (1984). "Acuity of sound localization: A topography of auditory space. I. Normal hearing conditions," Perception **13**, 581–600.

Oppenheim, A. V., and Schafer, R. W. (1975). *Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ).

Plomp, R., Pols, L. C. W., and van de Geer, J. P. (1967). "Dimensional analysis of vowel spectra," J. Acoust. Soc. Am. **41**, 707–712.

Pols, L. C. W., van der Kamp, L. J. Th., and Plomp, R. (1969). "Perceptual and physical space of vowel sounds, J. Acoust. Soc. Am. **46**, 458–467.

Pols, L. C. W., Tromp, H. R. C., and Plomp, R. (1973). "Frequency analysis of Dutch vowels from 50 male speakers," J. Acoust. Soc. Am. **53**, 1093–1101.

Schroeder, M. R. (1970). "Synthesis of low-peak-factor signals and binary sequences with low autocorrelation," IEEE Trans. Inform. Theory **16**, 85–89.

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.

Wenzel, E. M., Wightman, F. L., Kistler, D. J., and Foster, S. H. (1988). "Acoustic origins of individual differences in sound localization behavior," J. Acoust. Soc. Am. Suppl. 1 **84**, S79.

Wightman, F. L., and Kistler, D. J. (1989a). "Headphone simulation of free-field listening I: Stimulus synthesis," J. Acoust. Soc. Am. **85**, 858–867.

Wightman, F. L., and Kistler, D. J. (1989b). "Headphone simulation of free-field listening II: Psychophysical validation," J. Acoust. Soc. Am. **85**, 868–878.

Zahorian, S. A., and Rothenberg, M. (1981). "Principal-components analysis for low-redundancy encoding of speech spectra," J. Acoust. Soc. Am. **69**, 832–845.