Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency

John C. Middlebrooks
Kresge Hearing Research Institute, University of Michigan, 1301 East Ann Street, Ann Arbor, Michigan 48109-0506

(Received 1 September 1998; revised 11 March 1999; accepted 26 May 1999)

This study examined virtual sound localization in three conditions that differed according to the directional transfer functions (DTFs) that were used to synthesize the virtual targets. The *own-ear* and *other-ear* conditions used DTFs measured from listeners' own ears and those measured from other subjects, respectively. The *scaled-ear* condition employed other-ear DTFs that were scaled in frequency to minimize the mismatch between spectral features in the listener's and the other subject's DTFs. All measures of localization error typically were lowest in the own-ear condition. In other-ear conditions, all error measures tended to increase in proportion to the inter-subject differences in DTFs. When spectral features in an other-ear set of DTFs fell systematically lower in frequency than in a listener's own DTFs, low frontal targets typically were reported as low in the rear, and high rear targets were reported as high in front. When spectral features in a set of DTFs fell systematically higher in frequency than in a listener's own DTFs, elevation judgements showed an upward bias. In the scaled-ear condition, all measures of performance tended to improve relative to the other-ear condition. In the majority of cases, frequency scaling more than halved the penalty for use of another subject's DTFs.
[S0001-4966(99)01509-X]

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

INTRODUCTION

One can capture the transfer functions of the head and external ears by presenting a broadband sound in a free field while recording from within a subject's ear canals with miniature microphones. Transfer functions measured in that way have come to be known as ''head-related transfer function'' (HRTFs). When HRTFs are applied to an arbitrary signal and presented to the listener's two ears through headphones, he or she hears a ''virtual target'' that appears to originate from the location of the original sound source (e.g., Wightman and Kistler, 1989b; Bronkhorst, 1995; Møller *et al.*, 1996). In a similar way, one can compile a set of transfer functions for a wide range of sound-source locations and can synthesize a virtual acoustic environment. When listening through HRTFs measured from his or her own ears, a listener reports auditory events that are spatially compact and that appear ''externalized,'' i.e., that seem to arise from sources outside of the listener's head. When listening through HRTFs measured from another subject (i.e., ''nonindividualized'' HRTFs), listeners often complain that auditory events are spatially diffuse, and listeners often make incorrect judgments of the source locations (Wenzel *et al.*, 1993; Møller *et al.*, 1996). The primary goal of the current study was to explore procedures by which a generic set of HRTFs could be customized for use by any arbitrary listener.

The companion report (Middlebrooks, 1999) examined inter-subject differences in the directional components of subjects' HRTFs, which are referred to as *directional transfer functions* (DTFs). DTFs varied systematically among subjects in regard to the positions of spectral features along the frequency axis. Spectral features in the DTFs of one subject could be aligned with those of another subject by scaling his or her DTFs in frequency. To a first approximation, DTFs scaled in proportion to the sizes of subjects' heads and external ears, so that characteristic spectral features of a large subject's DTFs tended to lie at lower frequencies than those of a smaller subject. The present report compares virtual-localization performance under conditions in which virtual targets were synthesized from DTFs measured from the listeners' own ears and from the ears of other subjects. Subjects showed characteristic errors in localization that varied according to frequency differences between their own DTFs and the DTFs through which they listened. Localization performance with other subjects' DTFs improved substantially when inter-subject differences in DTFs were reduced by scaling in frequency.

I. METHODS

A. Subjects

The subject population consisted of 18 paid listeners (11 female and 7 male) drawn from the subject population of the companion study (Middlebrooks, 1999). Ages ranged from 22 to 43 yr (30.4 ± 5.7, mean$\pm$standard devation), and body heights ranged from 155 to 193 cm (169.7 ± 10.2). Fourteen of the listeners (eight female and six male), including the author, participated in the main blocks of free-field and virtual localization trials. The remaining four listeners participated only as ''naïve'' listeners in tests of virtual localization with varying frequency scale factors. All subjects were screened by conventional tests for hearing thresholds within 20 dB of audiometric zero. The DTFs were obtained in the companion study. They included measurements from 14 of the behavioral subjects plus one additional subject.

B. Apparatus, sound generation, and coordinate system

All behavioral tests were conducted in the double-walled anechoic chamber that was described in the companion paper. Free-field stimulation experiments employed a mechanism that could position two loudspeakers on the surface of an imaginary sphere that was 1.2 m in radius. The loudspeakers were Pioneer model TS-879 two-way coaxials. Virtual-environment stimuli were presented through Sennheiser model 265 circumaural headphones. Audio signals were generated with an Intel-based microcomputer, using a 16-bit analog interface and a digital-signal-processor board (Tucker-Davis-Technologies). Signals were generated with 16-bit precision at a sample rate of 50 kHz.

Free-field stimuli were presented from the moveable loudspeakers. The loudspeakers were calibrated using a $\frac{1}{2}$-in. precision microphone (ACO Pacific) that was positioned in the center of the sphere described by the speaker-movement mechanism. Probe stimuli for the calibration were pairs of 16 384-point Golay codes (Golay, 1961; Foster, 1986; Zhou *et al.*, 1992). An inverse transfer function was computed and stored for each loudspeaker. Virtual stimuli were synthesized using DTFs measured in the companion study. Complex DTFs were transformed to the time domain using an inverse fast Fourier transform (inverse FFT). The resulting impulse responses were truncated to 256 points (i.e., 5.12 ms).

Free-field and virtual stimuli consisted of independent samples of Gaussian noise bursts. The Gaussian samples had Rayleigh-distributed amplitude spectra that were flat on average but differed in detail from trial to trial. Each free-field stimulus was generated by drawing a Gaussian sample, filtering it with the inverse transfer function for the appropriate loudspeaker, and band-pass filtering between 300 and 16 500 Hz. Each virtual stimulus was generated by convolving a Gaussian sample with left- and right-ear impulse responses for the desired target. The headphones were corrected by the manufacturer for a ''diffuse field response.'' No additional headphone correction was applied. Bandwidths of virtual stimuli were limited only by the bandwidth of the headphones, which the manufacturer specified as 10–30 000 Hz. Both types of stimulus were windowed to durations of 250 ms, including 5-ms raised-cosine onset and offset ramps.

One technique that commonly is used for synthesizing virtual targets utilizes HRTFs derived from acoustical measurements at the tympanic membrane (Wightman and Kistler, 1989a). In that technique, one must also measure the transfer function from headphones to the tympanic membranes and deconvolve the headphone-to-tympanic membrane transfer functions from the free-field transfer function. Headphone-to-tympanic-membrane transfer functions vary appreciably among listeners, so individually measured headphone transfer functions should be used (Pralong and Carlile, 1996). The present study used an alternative approach. The transfer functions used in this study were DTFs, which are HRTFs that are processed to isolate the component that is specific to sound-source location (Middlebrooks and Green, 1990; Middlebrooks, 1999). DTFs are obtained from HRTFs by dividing by a listener-specific common component that is computed from an average across the set of all HRTFs. That procedure replaces the separate measurement of the headphone-to-tympanic membrane transfer function.

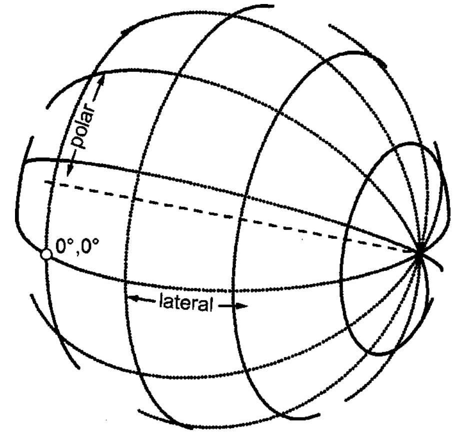

FIG. 1. Horizontal polar coordinate system.

For certain stimulus conditions, DTFs were scaled in frequency. Frequency scaling was accomplished by changing the output rate of the digital-to-analog converter (DAC). The DTFs were collected at a sample rate of 50 kHz, so the base output rate of the DAC was 50 kHz (i.e., a period of 20 μs). The maximum range of scale factors that was used was 0.757–1.320, so the output rate of the DAC was varied from 37 850 to 66 000 Hz. The total number of points in stimulus waveforms was adjusted to maintain a constant overall stimulus duration and rise/fall time.

The locations of free-field and virtual stimuli and the locations of listeners' location judgements are specified in a horizontal polar coordinate system (Fig. 1). This coordinate system is equivalent to one proposed by Morimoto and Aokata (1984). The horizontal position is given by the *lateral angle*, which is the angle formed by the stimulus, the center of the listener's head, and the vertical midline. Lateral angle ranges from left to right 90°, with 0° corresponding to the vertical midline. The vertical and front/back position is given by the *polar angle*. This is the angle around the horizontal pole that coincides with the listener's interaural axis. The polar angle ranges from −90° (beneath the interaural axis), through 0° (on the frontal horizon), through +90° (above the interaural axis), through 180° (on the rear horizon), to 270° (beneath the interaural axis). The advantage of this coordinate system is that the two coordinates correspond to the dimensions signaled primarily by two types of localization cues. The lateral angle is signaled primarily by interaural difference cues. The polar angle unifies front/back and up/down dimensions, both of which appear to be signaled by spectral shape cues. A contour of varying polar angle at a constant lateral angle corresponds to a ''cone of confusion'' (Woodworth, 1938). A disadvantage of the use of polar angle is that the spherical angle that corresponds to a given polar angle is highly compressed at large lateral angles near the poles. For instance, at a lateral angle of 85°, the angular distance between polar angles 0° and 180° is only 10° when measured at the center of the listener's head. For that reason, analysis of the polar dimension of localization judgements was restricted to stimulus locations within 30° of the vertical midline, which corresponded to half of the area of the coordinate sphere. Within that range, spherical angles were compressed by a factor of no more than 0.866 (i.e., the cosine of 30°). By excluding lateral locations from the computation of polar error, front/back confusions probably were undercounted, since localization errors of that type are more prevalent for lateral stimuli (Wightman and Kistler, 1989b; Makous and Middlebrooks, 1990; Carlile *et al.*, 1997). Nevertheless, that undercount should not have influenced the comparison of various virtual localization conditions.

C. Behavioral protocol

The behavioral protocol was similar to that used in our previous studies (Makous and Middlebrooks, 1990; Middlebrooks, 1992). The listener stood on a platform that could be adjusted in height to center his or her head in the imaginary sphere. Trials were conducted in complete darkness. The listener reported the apparent location of a free-field or virtual target by orienting to face the location. Each listener was instructed to ''point with your nose.'' The head orientation was monitored with an electromagnetic device (Polhemous FASTRAK). The sensor for the FASTRAK was attached to a cloth baseball cap from which the visor had been removed.

In free-field trials, the loudspeaker was positioned prior to the beginning of each trial. The beginning of the trial was signaled by the onset of a broadband noise from a loudspeaker that was fixed on the wall at coordinates 0°, 0°. The listener oriented toward the fixed source, then pressed a response key. The broadband noise was terminated when the response key was pressed. The test stimulus was presented after a 600- to 1200-ms delay. The listener oriented toward the apparent position of the test stimulus, then pressed the response key. The computer recorded the listener's head orientation at the time of both key presses. The 250-ms stimulus was sufficiently short that the listener did not move during the presentation of the sound.

The protocol for virtual localization trials was identical to that for the free-field trials with the exceptions that listeners wore headphones and that the initial centering stimulus consisted of a light from a light-emitting diode (LED) rather than a noise. There was concern that some of the virtual targets might seem poorly externalized or spatially diffuse such that the listener could not decide on a single apparent location. For that reason, listeners were provided with an alternate response key that they could press to identify such situations. The use of the alternate key varied widely, but across all listeners and conditions it was used on an average of 9% of trials. Trials on which the alternate key was pressed were eliminated from further analysis.

To quantify stimulus sensation levels, detection thresholds were measured for free-field targets and for each set of DTFs and each frequency scale factor with which a listener was tested. Targets were at free-field or virtual locations 0°, 0°. The sensation level produced by a source of any particular pressure level varied according to the location of the free-field or virtual source. For instance, because of shadowing of sound by the head and external ears, a source at a given level would result in a lower sensation level when presented from behind the listener than when presented from in front. This variation in sensation level is a natural localization cue but is most informative to a listener when the sound-source level is constant. To discourage listeners from relying on absolute level cues for localization, stimulus levels were varied randomly in steps of 1 dB from 40 to 60 dB above the threshold measured at 0°, 0°. Other than varying overall levels, however, no effort was made to specifically invalidate the absolute level cue by, for instance, increasing the level of stimuli presented from the rear.

Fourteen listeners participated in the main blocks of free-field and virtual localization tests. Those listeners were trained as follows: First, they were introduced to the test chamber and the speaker-movement mechanism. They were instructed to orient to targets with full head movements and to overcome the natural tendency to orient with a combination of head and eye movements. They completed two short sessions of free-field localization trials in which the room lights were left on so that the loudspeakers were visible. After each session, they were shown their data and were given verbal feedback on their orienting accuracy. Next, they completed two sessions of trials in which the room lights were turned off and in which feedback was provided immediately after each response by an LED that was illuminated at the center of the loudspeaker cone. Finally, they completed two full sessions of free-field trials in the dark with no immediate feedback. Listeners who persisted in orienting with eye movements rather than head movements received additional instruction and a few more practice sessions. After free-field training, data were collected in three free-field test sessions. Next, listeners received four virtual-localization training sessions in which virtual targets were synthesized from DTFs measured from each listener's own ears. Finally, they completed one or more blocks of seven virtual-localization test sessions. The number of blocks completed by each listener was determined by his or her willingness to continue participation.

The sets of free-field and virtual target locations were drawn from the set of 400 stimulus locations for which DTFs were measured in the companion study. The two locations more than 60° below or more than 79° above the horizontal plane were excluded because it was physically difficult for listeners to orient their heads to those locations. One set of 300 locations was drawn with replacement and was used for the free-field condition for all listeners. A second set of 300 was used for all virtual conditions for all listeners. One hundred forty-four locations fell within 30° of the vertical midline and, thus, were used for analysis of localization in the polar dimension. A block of free-field localization trials, completed in 3 sessions, consisted of 1 trial at each of 300 locations in a set. Listeners completed sessions of 100 free-field trials in 12 to 18 min. In free-field trials, the mechanism that positioned the loudspeaker made an audible noise. That noise always came from the location of the motors, straight overhead, so it did not give a cue to the test stimulus location. Nevertheless, the duration of the motor noise might have given a cue to the distance between successive stimulus locations. For that reason, the test locations were ordered so that the excursion of the speaker mechanism from one position to the next always fell in the range of 60°–120° to the left or right and 30°–70° up or down. This resulted in speaker excursions that all were around 5 s in duration. Cues to the stimulus location were further confounded by the fact that on any trial the stimulus could be presented from the front or the rear loudspeaker. Because of the special ordering of stimuli, the range of possible locations that could follow any particular stimulus was reduced by 1/3 in azimuth and by 1/2 in elevation. A listener who understood this might have realized that no two stimuli were presented from the same location on successive trials, but aside from that it is difficult to imagine that the ordering of locations had appreciable influence in the free-field localization results.

Virtual localization was tested in three conditions that differed only in the source of the DTFs from which virtual targets were synthesized. In the *own-ear* condition, the DTFs were those recorded from each listener's own ears. In the *other-ear* condition, the DTFs were those recorded from the ears of another subject. In the *scaled-ear* condition, DTFs recorded from the ears of another subject were scaled in frequency to align them with the listener's own DTFs. Trials from three conditions were interleaved within a single block of 900 trials. A block consisted of either 300 trials from each of the three conditions or 300 own-ear trials and 300 other-ear trials using DTFs from each of two other subjects. Each block of 900 trials was broken into 7 sessions of 128 or 129 trials each. Individual listeners completed from 1 to 6 blocks of virtual-localization trials, where blocks differed in the DTFs that were used. Other-ear and scaled-ear performance in each block of trials were evaluated relative to performance on own-ear trials within the same block. Listeners typically completed a session of 129 virtual-localization trials in around 12 min. Listeners usually completed two sessions in a day, separated by a 10-min rest period.

A separate series of trials was conducted to test the sensitivity of virtual-localization performance to the value of a frequency scale factor. Nine listeners participated in that series. DTFs were scaled by a factor that varied randomly between trials in powers of 1.04 (i.e., scale factors of 0.79, 0.82,...,1.22,1.27). In separate blocks, DTFs either were those measured from the listeners' ears or those measured from another subject. Virtual targets were restricted to the midline and had polar locations that ranged from -60 to $+70°$ (in front) and from $+110$ to $+240°$ (in the rear) in increments of 10°. Each combination of location and scale factor was repeated in four trials. Four of the listeners completed these trials after experience in the main blocks of free-field and virtual-localization trials. The additional five listeners received only brief instruction in the free-field condition, then conducted these variable-scale factor trials with no prior virtual-localization experience. One of these naïve listeners later completed full training and the main blocks of free-field and virtual-localization trials.

D. Data analysis

The measurement and computation of DTFs is described in the companion paper. That paper described a procedure that involved filtering with a bank of band-pass filters followed by spatial interpolation. In the present study, that procedure was used only for quantification of optimal frequency scale factors and inter-subject differences between DTFs, not for synthesis of test stimuli. The differences between the DTFs of a given pair of subjects were quantified by the *inter-subject spectral difference*. Computation of inter-subject spectral difference is discussed in the companion paper, but briefly it involved subtraction of dB amplitudes of DTF components between 3.7 and 12.9 kHz for each location, computation of the variance of the resulting difference distributions, and averaging of those variances across all locations. *Optimal frequency scale factors* were computed by systematically scaling in frequency the DTFs of one subject relative to another to find the factor that minimized the inter-subject spectral difference. The companion paper demonstrated that scale factors based on right-ear and left-ear DTFs were highly correlated. In the present study, scale factors were derived from right-ear DTFs, then were used to scale DTFs for both right and left ears.

As described above, stimulus locations and localization judgements are expressed in horizontal polar coordinates. Errors in the lateral and polar dimensions were computed independently by subtracting the target location angle from the response location angle. In some cases, 360° was added or subtracted so that the magnitude of the error was no more than 180°. The precision of localization in the lateral dimension was summarized for each subject and condition by *root-mean-squared* (*rms*) *lateral errors*.

In the polar dimension, computation of summary statistics was complicated by the occurrence of large front/back and up/down errors that appeared to form a distribution that was distinct from the scatter of judgements around the stimulus location. Conventional definitions of front/back and up/down confusions might have included localization judgments that fell within the local scatter of judgements, such as when a target at polar 80° is localized as 91°. For the purpose of distinguishing polar localization judgements that fell well outside the scatter of judgements around the target, *quadrant errors* were defined as polar errors that were larger than 90°. The criterion of 90° was selected by examining histograms of the magnitudes of polar errors across all listeners and conditions. All quadrant errors necessarily involved localization judgements that fell in a different quadrant than the target, whereas some judgments that fell in the wrong quadrant produced polar errors smaller than 90° and were not scored as quadrant errors. The *rms local polar error* was defined as the rms average of polar errors that were less than 90° in magnitude. The purpose of that measure was to characterize localization accuracy in the polar dimension on trials in which quadrant errors did not occur. The distribution of polar errors for any subject and/or condition typically was strongly bimodal, so it was inappropriate to apply a measure of central tendency. For that reason, polar errors are summarized in this report only by the percentage of trials in which quadrant errors occurred (*% quadrant errors*) and by rms local polar errors.

Two signed measures were employed to quantify systematic biases in responses. The *lateral bias* and *elevation bias* were given by the averages of signed lateral and elevation errors, respectively, where elevation is the angle above or below the horizontal plane. Biases measured the tendency of listeners to respond consistently too far lateral to the target

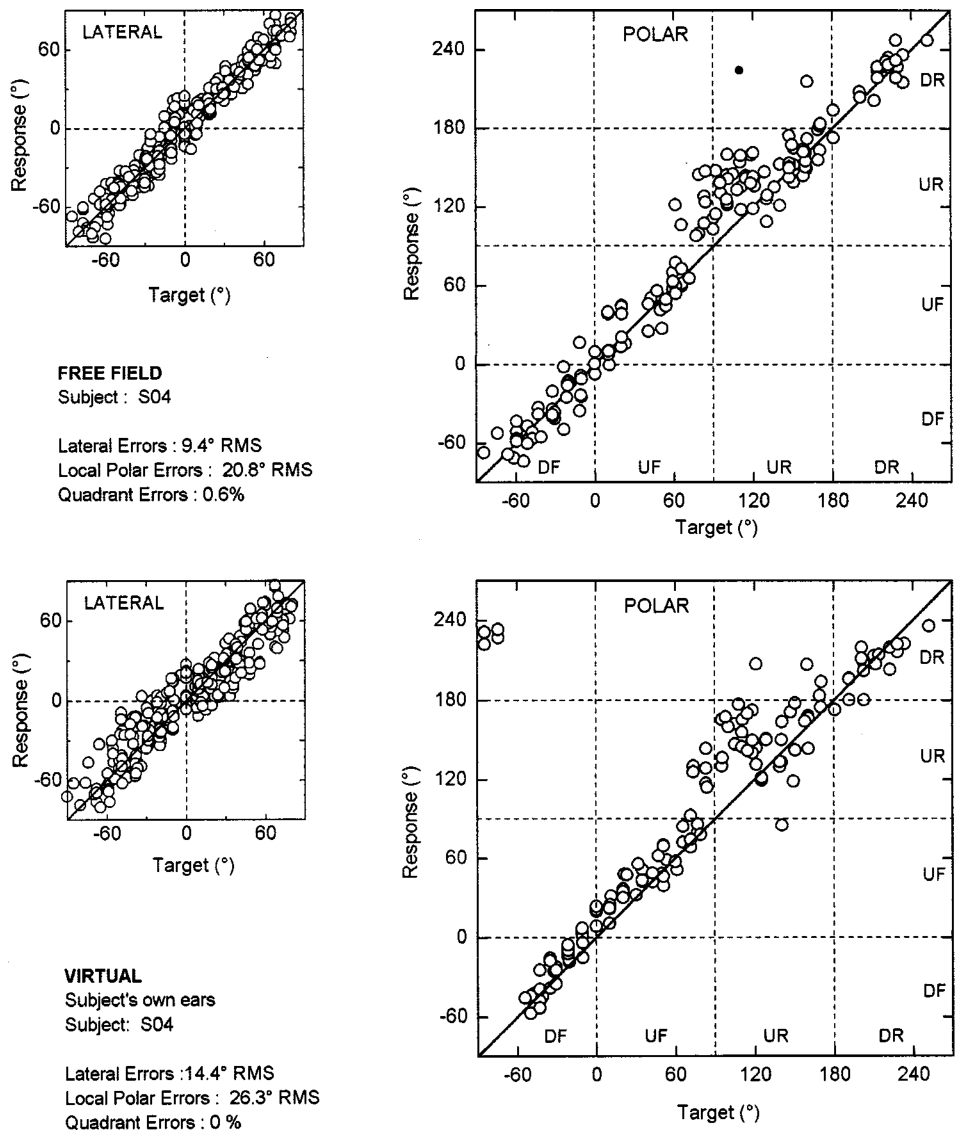

FIG. 2. Localization by listener S04 in the free-field (top panels) and own-ear virtual (bottom panels) conditions. Horizontal and vertical axes show the target locations and the listener's localization judgements, respectively. Localization in the lateral and polar dimensions is plotted in the left and right panels, respectively. Axes for lateral and polar dimensions are scaled so that one degree in either dimension occupies an equal length. Initials DF, UF, UR, and DR indicate down-front, up-front, up-rear, and down-rear quadrants. Filled circles (only one seen in this figure, in the upper right panel) represent quadrant errors. All other localization judgements are represented by open circles.

or too far above the target. Both computations of bias might have been underestimated by a ceiling effect because a response to a stimulus at, for instance, lateral 90° could be no further lateral than 90°. For that reason, biases were computed only for trials in which the stimulus locations were within 60° of the horizontal or midline planes. Two other classes of trials were excluded from the computation of elevation biases: those in which quadrant errors occurred and those in which the stimulus was within the up-rear quadrant. Up-rear quadrant trials were excluded because the responses to those stimuli tended to be so variable that they overwhelmed any systematic variation across listeners or conditions.

II. RESULTS

The results are presented in four sections. The *first* provides baseline measures of performance in the free-field localization condition and in a virtual condition in which targets were synthesized with DTFs measured from listeners' own ears. The *second* characterizes the types of virtual-localization errors that listeners made when they listened through DTFs measured from other subjects' ears. The *third* quantifies improvements in virtual localization that resulted from scaling DTFs in frequency, and it relates those improvements to a descriptive spectrum-matching model. The *final* section demonstrates the sensitivity of localization judgments to the value of the frequency scale factor that was used to scale DTFs.

A. Localization with listeners' own DTFs

Listeners localized in two conditions in which the sounds were filtered with the DTFs of their own ears. In the *free-field* condition, stimuli were presented from a movable loudspeaker and sounds were filtered naturally by the external ears as the sounds traveled from the source to the tympanic membranes. In the *own-ear* virtual localization condition, stimuli presented from headphones were filtered by DTFs that had been recorded previously from each listener's own external ears. Figure 2 shows localization results from one listener (S04) in the two conditions. This listener localized reasonably accurately throughout all lateral and polar angles, with the largest errors occurring in response to stimuli presented in the upper rear quadrant. Performance in the own-ear virtual condition was nearly as accurate as that in the free-field condition. For the listener represented in Fig. 2, rms lateral errors were 9.4° and 14.4° in the free-field and own-ear conditions, respectively, and rms local polar errors were 20.8° and 26.3°, respectively. As described in Sec. I B, polar errors were computed only for targets located within 30° of the vertical midline.

The rate of occurrence of quadrant errors (defined in Sec. I D) in both free-field and own-ear virtual conditions showed a strongly bimodal distribution across the 14 listeners. In the free-field condition, nine listeners showed quadrant errors in only 0%–2.5% of trials (median = 0.6%), whereas the other five listeners showed quadrant errors in 8.6%–16% of trials (median = 12.3%). In the own-ear-virtual condition, nine listeners made quadrant errors in only 0%–6.3% of trials (median = 2.1%), and the five other subjects made quadrant errors in 10.5%–26.2% of trials (median = 13.2%). The listeners who produced a rate of quadrant errors of 8.6% or more in the free-field condition showed a rate of 4.3%–26.2% in the own-ear virtual condition.

The listener whose localization performance is shown in Fig. 2 is representative of those who produced few quadrant errors. That listener made only one quadrant error in the free-field condition (indicated by a filled circle) and no quadrant errors in the own-ear virtual condition. Figure 3 shows localization performance by one of the five listeners who showed a high rate of quadrant errors in free-field and virtual trials. The majority of that subject's errors consisted of up-rear targets localized to down-front.

Across all listeners who showed quadrant errors, the rate of quadrant errors was highest for targets in the up-rear quadrant. The five listeners who showed a high rate of quadrant errors showed quadrant errors in 17.5%–41.3% (median = 25.4%) of free-field trials and in 32.1%–60.5% (median = 43.2%) of own-ear virtual trials in which targets were in the up-rear quadrant. Examination of DTFs (e.g., Middlebrooks, 1999) indicates that DTFs for up-rear locations tend to be relatively devoid of spectral features. That could explain why discrimination of targets within that quadrant might be imprecise, but it does not explain why targets in that quadrant would be localized to other quadrants.

Table I shows summary statistics for the free-field and own-ear conditions. With few exceptions, mean errors in the own-ear virtual condition were within about one standard deviation of mean errors in the free-field condition.

B. Localization with other subjects' DTFs

In the *other-ear* virtual localization condition, each listener listened to sounds that were filtered with DTFs measured from the ears of a different subject. The difference between each listener's DTFs and the DTFs used in a localization trial was quantified by a global measure of *inter-subject spectral difference*, which is defined in the companion paper. Listeners were matched with sets of DTFs so that cases spanned a range of the smallest to largest spectral differences. Fourteen listeners each localized using DTFs from one to seven other subjects, yielding a total of fifty-eight cases. Figure 4 shows the increase in errors in the other-ear

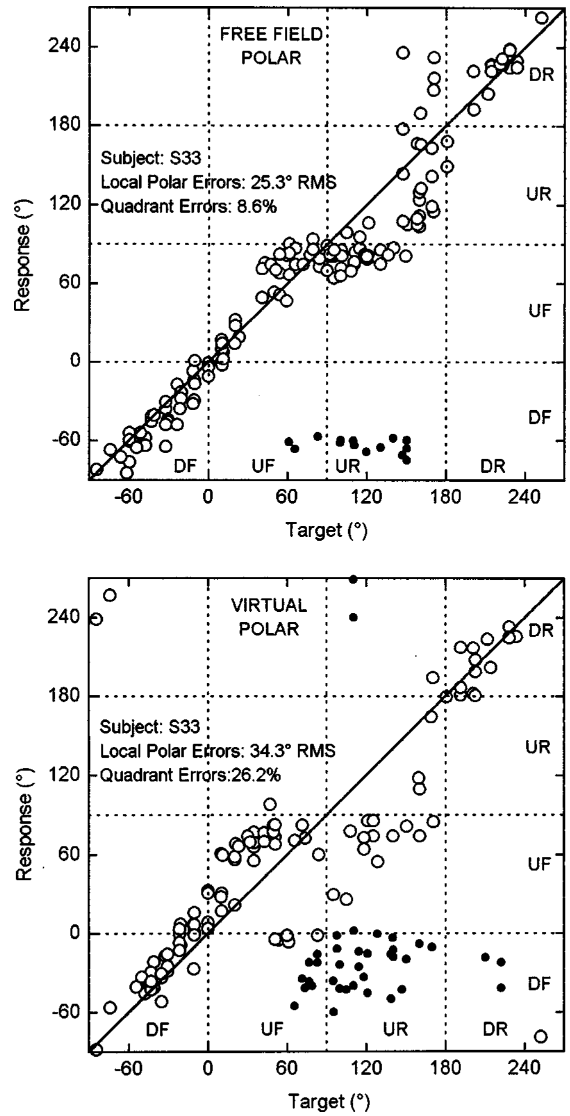

FIG. 3. Localization by listener S33, who showed a high rate of quadrant errors. All other conventions are as in Fig. 2.

condition compared to the own-ear condition as a function of inter-subject spectral difference in DTFs. All error measures of rms errors tended to increase with increases in the spectral difference between the listener's DTFs and the DTFs used in the localization trials. The rate of quadrant errors consistently was larger in the other-ear condition than in the own-ear condition, although that increase was negligible in sev-

TABLE I. Summary measures of free-field and own-ear virtual localization performance (means and standard deviations).

	Free-field	Own-ear virtual
rms lateral error (deg.)	10.6±2.0	14.5±2.2
Magnitude of lateral bias (deg.)	2.9±3.1	3.1±3.9
rms local polar error (deg.)	22.7±5.1	28.7±4.7
Magnitude of elevation bias (deg.)	5.5±4.4	10.2±6.6
Total quadrant errors (% of trials)	4.6±5.9	7.7±8.0
Quadrant error by target quadrant:		
Down-front (% of trials)	0.6±1.1	1.0±2.6
Up-front (% of trials)	1.9±3.3	5.7±6.4
Up-rear (% of trials)	10.6±13.8	21.7±21.1
Down-rear (% of trials)	0±0	1.8±5.0

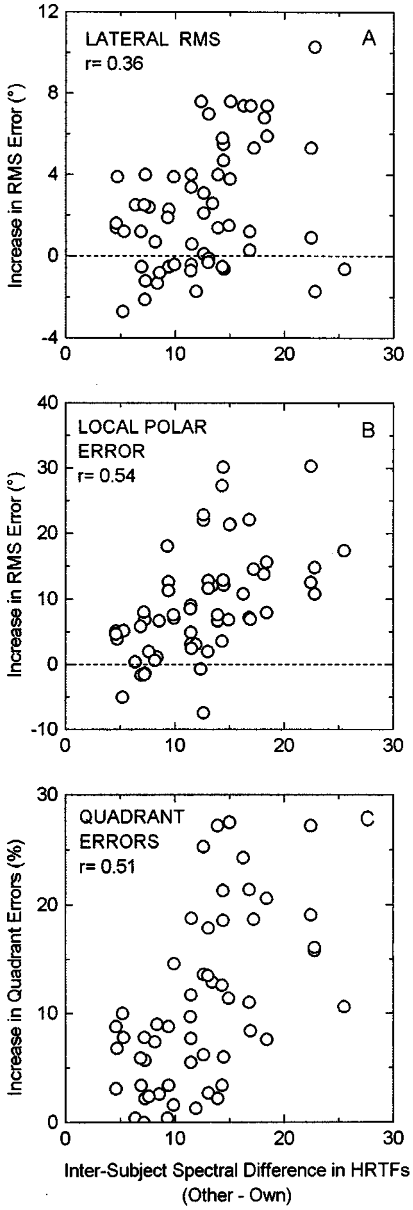

FIG. 4. Changes in localization performance resulting from use of nonindividualized DTFs. Each symbol represents one of 58 cases, each case consisting of one listener using one set of DTFs.

eral instances in which the inter-subject spectral difference in DTFs was smaller than about 10 dB^2. In a few instances, lateral errors and local polar errors were slightly smaller in the other-ear condition, although there was little evidence for a significant improvement in those measures in the other-ear compared to the own-ear condition. As a general rule, listeners' localization performance suffered when they listened through other subjects' DTFs.

Figure 5 shows two examples of virtual localization in the other-ear virtual condition. The results are from the same listener whose free-field and own-ear-virtual results are shown in Fig. 2. In Fig. 5, the upper and lower halves of the figure represent performance in the other-ear virtual condition using DTFs from two different subjects. In the lateral dimension (left two panels) one can see a moderate increase, relative to the own-ear condition, in the scatter of responses but there is still a strong correlation between target and response locations in the lateral dimension. In contrast, responses in the polar dimension (right two panels) showed a considerable increase in quadrant errors, from 0% in the own-ear condition to 27.6% or 12.9% in the other-ear conditions. The effect on local polar errors differed between sets of DTFs. In the upper right panel, local polar errors actually decreased slightly, possibly because the number of responses in the local-error computation was reduced by the large number of quadrant errors. In the lower right panel, the rms local polar error increased by nearly 50% relative to the own-ear condition. The distributions of responses in the polar dimension were different in the two cases. In the case shown in the upper right panel, nearly all down-front targets were localized to the down-rear quadrant. Most responses to down-rear targets were in the correct quadrant, but were below the target (i.e., above the diagonal in the figure). In the case shown in the lower right panel, the majority of down-rear targets were localized to the front. Responses to most targets in the front were systematically too high.

The differences in the distribution of localization judgements in the two other-ear conditions shown in Fig. 5 can be related to a property of DTFs that was described in the companion report (Middlebrooks, 1999). Specifically, that report demonstrated that DTFs tend to differ systematically among subjects in regard to the center frequencies of spectral features (e.g., spectral peaks, notches, and slopes). In addition, inter-subject frequency differences tended to correlate with the physical dimensions of subjects' heads and external ears. Therefore, for ease of presentation, this report will refer to *large* and *small* subjects, where *large* indicates a subject in whose DTFs spectral features tend to lie at lower frequencies than those of a *small* subject. In behavioral results in the other-ear virtual condition, small listeners who listened through large subjects' DTFs tended to produce different patterns of errors than did large listeners who listened through small subjects' DTFs. For instance, the listener represented in Fig. 5 was near the middle of the range of listeners. The localization judgements shown in the upper and lower panels were made when listening through the DTFs of larger and smaller subjects, respectively. Three characteristics of other-ear localization judgements correlated fairly consistently with inter-subject differences in the center frequencies of DTF spectral features or with inter-subject differences in interaural delays. Those characteristics were lateral biases, the distribution of quadrant errors, and elevation biases.

Listeners tended to show positive or negative lateral biases when listening through DTFs from larger or smaller subjects, respectively. Most models of sound localization attribute the lateral component of localization to interaural difference cues (reviewed in Middlebrooks and Green, 1991), and Wightman and Kistler (1992) have shown that interaural time differences (ITDs) dominate the lateral component of localization judgements of sound sources that contain low frequencies. The latter result predicts that lateral biases would correlate with the maximum ITDs that were produced by a given set of DTFs. To test that prediction, the ratio of the maximum ITDs was computed for each combination of behavioral listener and set of other-ear DTFs. Ratios greater

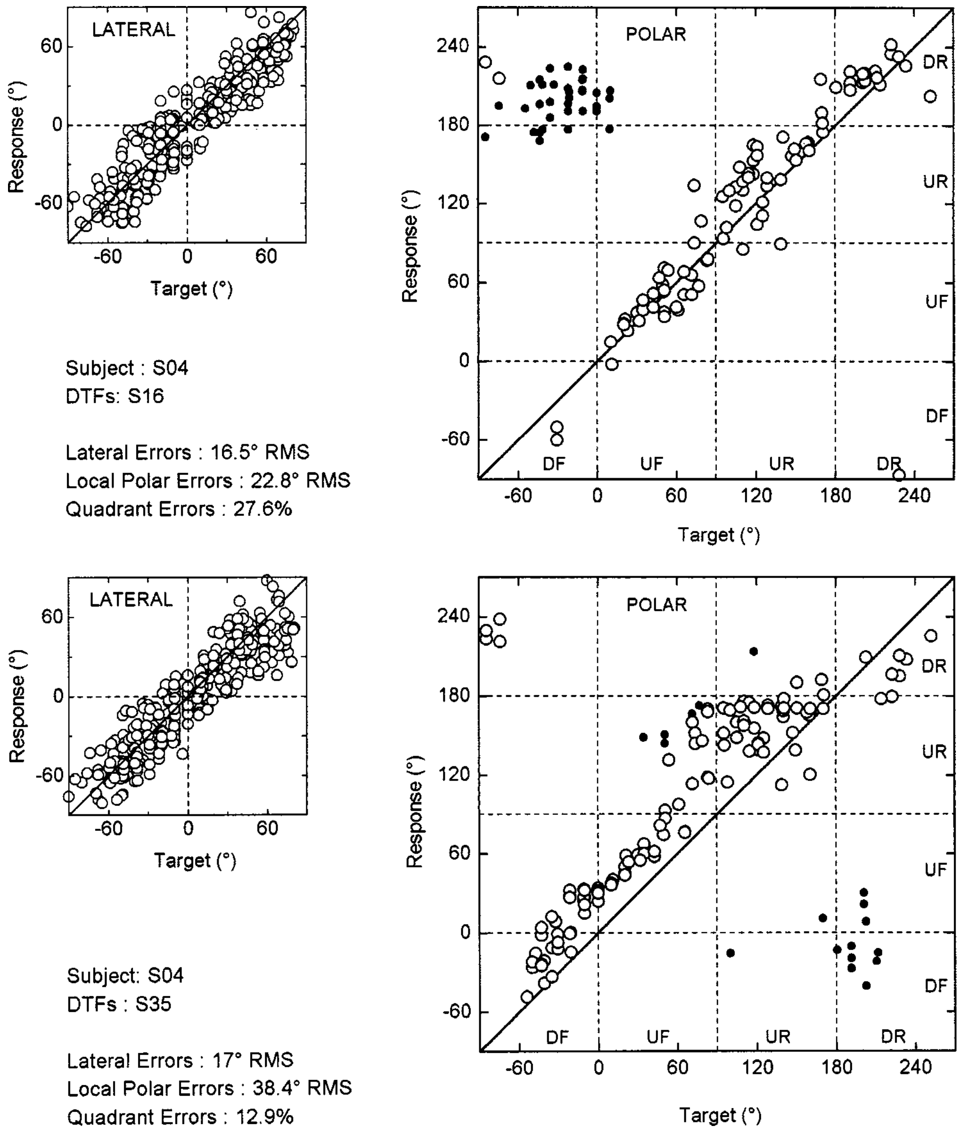

FIG. 5. Localization by listener S04 listening through DTFs from a larger subject (S16, top panels) and those of a smaller subject (S35, bottom panels). All other conventions as in Fig. 2.

than unity corresponded to small listeners localizing with larger subjects' DTFs. The increase in lateral bias was computed by subtracting the lateral bias in the own-ear condition from that in the other-ear condition. Figure 6 shows a strong positive correlation between maximum-ITD ratio and increase in lateral bias. The solid curve is a prediction based on the assumption that when a listener hears an ITD produced by a pair of other-ear DTFs for lateral angle θ_{other}, he or she reports a location θ_{own} at which his or her own DTFs would produce an identical ITD. The relationship between θ_{other} and θ_{own} is given by the following expression, which is derived from the spherical-head model of ITDs (Woodworth, 1938):

$$\text{MaxITD}_{\text{own}}(\theta_{\text{own}}+\sin(\theta_{\text{own}}))$$

$$=\text{MaxITD}_{\text{other}}(\theta_{\text{other}}+\sin(\theta_{\text{other}})),$$

where MaxITD for each set of DTFs is obtained by fitting the spherical-head model to the complete set of DTFs, then evaluating the model at $\theta=90°$. The ratio of maximum ITDs was computed for each subject and set of DTFs represented in Fig. 6. Each response θ_{own} was then estimated for every target θ_{other}, and the average of $\theta_{\text{other}}-\theta_{\text{own}}$ was averaged across all targets. The solid curve in Fig. 6 is the predicted

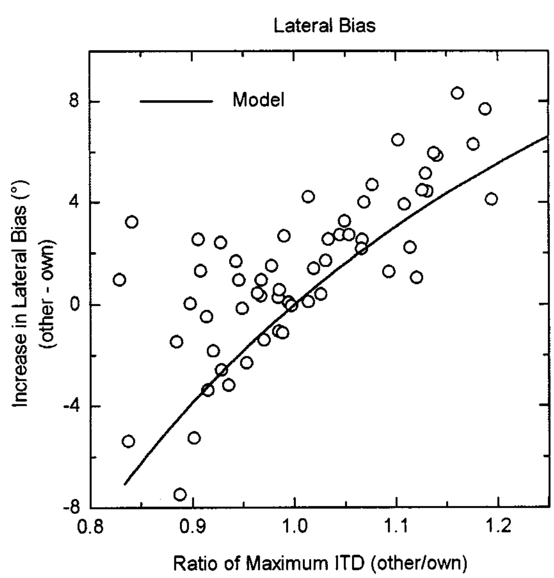

FIG. 6. Increase in lateral bias in the other-ear condition. A positive increase indicates that the listener tended to point farther to the side in the other-ear condition than in the own-ear condition. The abscissa is the ratio of maximum interaural delays, other-ear/own-ear. The solid line is a prediction described in the text ($r=0.73$, $N=58$).

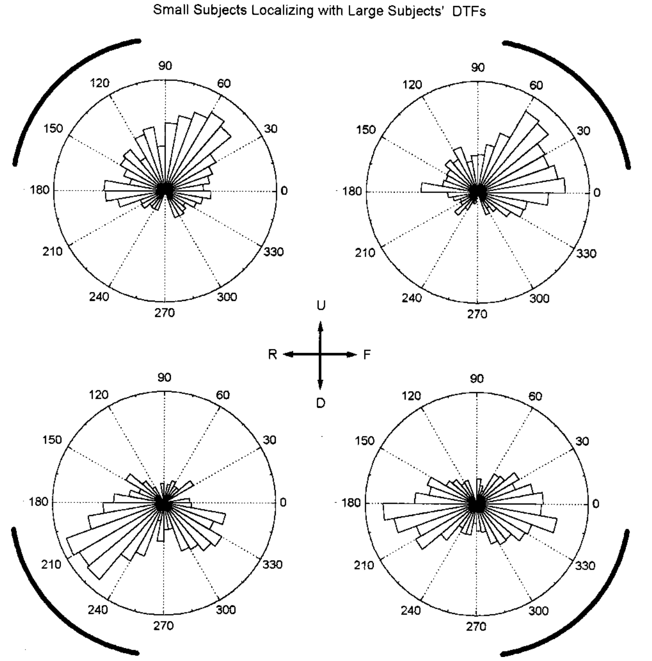

FIG. 7. Distribution of responses in the polar dimension: larger subjects' DTFs. These rose plots are drawn as if viewed from the listeners' right side. Each panel shows the response to targets presented in a particular quadrant, indicated by a thick arc. The *area* of each petal of a rose plot represents the percentage of responses that fell in a particular 10° sector (i.e., the *radius* of each petal is the square root of the percentage). This figure represents the 14 cases that showed the greatest downward frequency disparity between listeners' own DTFs and the DTFs through which they listened; i.e., spectral features in the DTFs tended to lie at lower frequencies than those in the listeners' own DTFs. The listeners' up, down, front, and rear directions are indicated by initials U, D, F, and R.

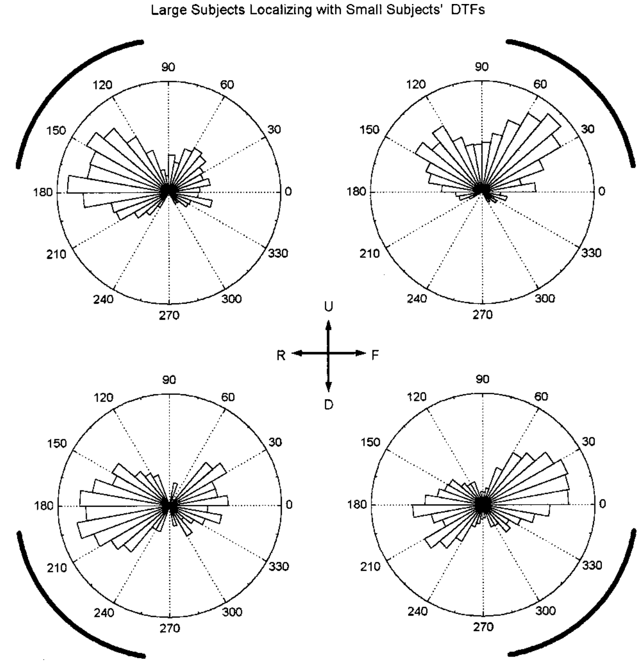

FIG. 8. Distribution of responses in the polar dimension: smaller subjects' DTFs. This figure represents the 14 cases that showed the greatest upward frequency disparity between listeners' own DTFs and the DTFs through which they listened; i.e., spectral features in the DTFs tended to lie at higher frequencies than those in the listeners' own DTFs. All other conventions are as in Fig. 7.

increase in lateral bias for each ratio of head radii. The modeled curve lies within the scatter of data and follows the general trend. The largest deviation of data points from the curve is due to some of the large listeners localizing with small subjects' DTFs (i.e., ITD ratio <1) that did not show as large an undershoot as predicted by the model.

The most conspicuous characteristic of localization errors in the other-ear condition was the quadrant errors. The results shown in Fig. 5 were fairly representative of the behavior of listeners localizing with the DTFs of larger subjects (upper panels) or of listeners localizing through the DTFs of smaller subjects (lower panels). For the purpose of presentation, all the other-ear cases were ranked according to the optimal frequency scale factors between each listener's own DTFs and the DTFs through which he or she listened. The quartile of small listeners/large-subject DTFs is represented in Fig. 7 and the quartile of large listeners/small-subject DTFs is represented in Fig. 8. The distributions of localization judgements are shown in *rose plots*, in which the *area* of each rose petal is proportional to the percentage of responses in a particular sector. Each of the four panels in each figure represents the distribution of judgments of targets presented within the quadrant that is indicated by the bold arc. In the condition of smaller listeners localizing with larger subjects' DTFs (Fig. 7), three characteristics were most common. (1) Down-front targets often were localized to down-rear ($36.3\% \pm 25.2\%$ of trials, mean $\pm$SD). (2) About half of up-rear targets were localized to up-front ($51.6\% \pm 15.0\%$). (3) Front/back confusions were more common than up/down confusions ($44.4\% \pm 7.3\%$ front/back compared to 23.4% $\pm 8.0\%$ up/down). The characteristics were different in the condition of larger listeners localizing with smaller subjects' DTFs (Fig. 8). (1) There was a considerable rate of front/back confusions ($31.2\% \pm 9.7\%$), although the rate was lower than in the small-listener/large-ear condition. (2) More conspicuous was a systematic upward bias of localization judgements from the lower quadrants to on or above the horizontal plane. More than half (58.2 ± 20.3) of targets in lower quadrants were localized to upper quadrants.

Although the quadrant errors tended to dominate the polar component of responses, trends in polar errors were evident even in trials in which quadrant errors did not occur. In particular, listeners tended to show a positive elevation bias (i.e., to point above the target) in conditions in which they listened through smaller subjects' DTFs. The tendency to point above the target is shown clearly in the lower plot of Fig. 5 in which responses above the targets appear above the identity line in the front hemifield and below the identity line in the rear hemifield. A symmetrical tendency to point below targets when listening through larger subjects' DTFs also was present. That tendency was less conspicuous, however, because down-front targets often were localized to incorrect quadrants (as in the upper panel of Fig. 5). Similarly, a clear trend is difficult to see for the up-front quadrant in the example in Fig. 5, although one can see a downward bias in targets in the down-rear quadrant (i.e., symbols fall above the identity line in the rear).

Figure 9 plots the elevation biases of 58 cases as a function of the optimal frequency scale factor that would minimize the spectral difference between the listener's own DTFs

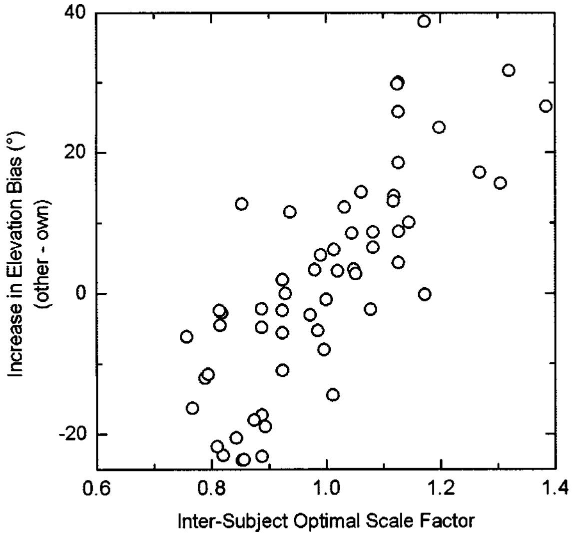

FIG. 9. Increase in elevation bias in the other-ear condition. A positive increase indicates that the listener tended to point to higher elevations in the other-ear condition than in the own-ear condition. The abscissa plots the optimal frequency scale factor that would scale the listeners' DTFs to the DTFs that were used in each case. A scale factor >1 indicates that spectral features in the DTFs tended to lie at higher frequencies than those in the listeners' own DTFs ($r=0.78$, $N=58$).

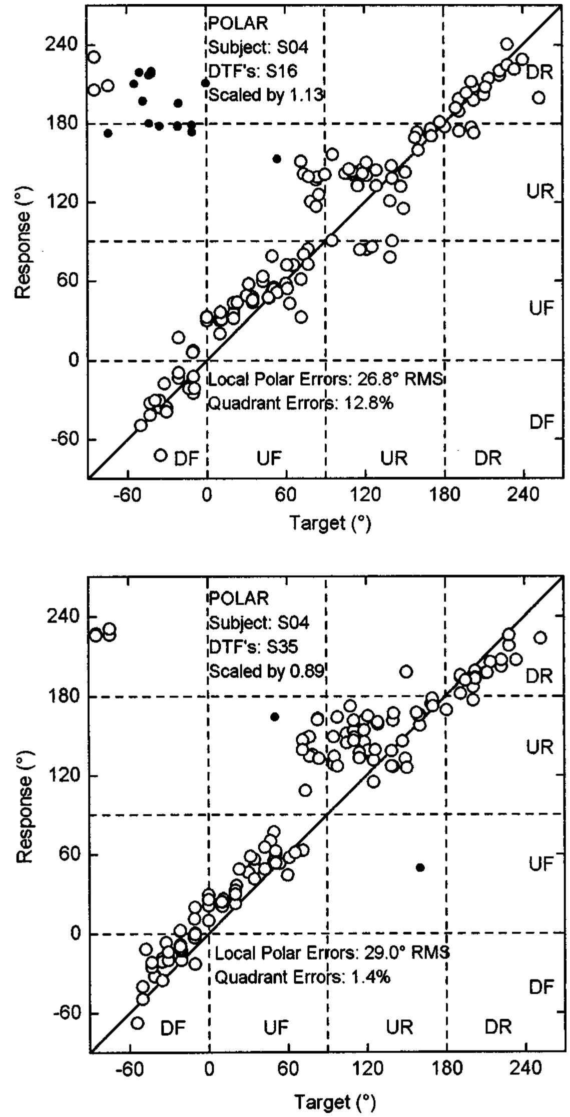

FIG. 10. Localization by listener S04 listening through scaled DTFs from subjects S16 (top panels) and S35 (bottom panels). DTFs were scaled by optimal frequency scale factors of 1.127 (top panel) and 0.888 (bottom panel). All other conventions are as in Fig. 2.

and the other-ear DTFs. A value greater than 1 on the abscissa indicates that spectral features in the DTFs that were used fell at higher frequencies than in the listener's own DTFs. Biases are expressed as increases relative to the bias measured in the own-ear condition. There is a strong positive correlation. Examination of DTFs (e.g., in the companion paper) indicates that, in general, spectral features tend to shift upward in frequency as sound sources shift upward in elevation. When a listener's localization judgments show an upward bias in response to the DTFs of a smaller subject, that might indicate that the listener hears spectral features at frequencies higher than expected and interprets those features as signaling higher elevations.

C. Localization with frequency-scaled DTFs

The companion paper demonstrated that inter-subject differences in DTFs could be reduced by appropriately scaling one set of DTFs in frequency relative to the other. That result led to the prediction that scaling DTFs in frequency could result in a reduction of errors in other-ear localization. In testing that prediction, optimal frequency scale factors were chosen that minimized the inter-subject spectral difference in DTFs (as described in the companion paper). The condition in which listeners localized with other-ear DTFs scaled in frequency is referred to here as the *scaled-ear* condition. The DTFs were scaled by changing the output rate of the digital-to-analog converter (DAC). When the DAC rate was multiplied by β, DTF spectral features were scaled in frequency by a factor β, interaural difference spectra were scaled in frequency by β, and interaural delays were scaled in time by $1/\beta$.

Figure 10 shows localization performance in the polar dimension in the scaled-ear condition for the same listener who is represented in the own-ear condition in Fig. 2 and in the other-ear virtual condition in Fig. 5. The DTFs were from the same two subjects as in Fig. 5. The top panel of Fig. 10 shows a case in which DTFs from a larger subject were scaled in frequency by a factor of 1.127. Compared to the other-ear judgments in the top panel of Fig. 5, the number of quadrant errors in response to down-front targets was halved and the accuracy of responses in the down-rear quadrant was greatly improved. Local polar errors are slightly increased, but that might be a result of the large number of trials that were excluded from the local-error computation in the other-ear case because quadrant errors occurred. The bottom panel of Fig. 10 represents a case in which DTFs from a smaller subject were in frequency by a factor of 0.888. Compared to the lower panel of Fig. 5, all of the quadrant errors in response to down-rear targets were eliminated and the upward bias in elevation responses was reduced substantially, about to the level seen in this listener's own-ear virtual localization (Fig. 2, lower panel). In this case, frequency scaling reduced

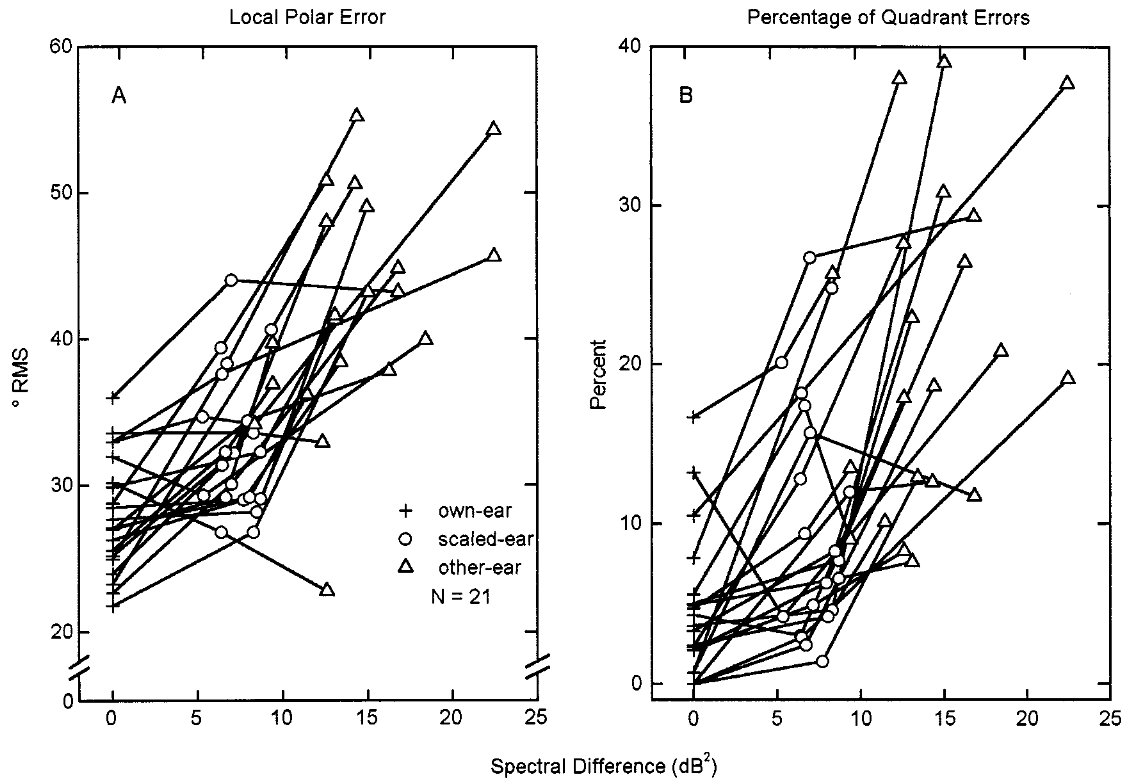

FIG. 11. rms local polar error (A) and percent quadrant errors (B) in three virtual-listening conditions. Each triad of symbols connected by lines represents one listener in the own-ear (plus signs), other-ear (triangles), and scaled-ear (circles) conditions. The abscissa shows the spectral difference between the DTFs that were used in each case and the listeners' own DTFs.

the overall rate of quadrant errors from 12.9% to 1.4%, and local polar errors from 38.4° to 29.0° rms.

The scaled-ear condition was tested in 21 cases (11 listeners tested with scaled DTFs from 1 to 4 other subjects). Figure 11 shows the rms local polar errors [Fig. 11(A)] and the percentage of quadrant errors [Fig. 11(B)] of those cases in the own-ear (plus signs), other-ear (triangles), and scaled-ear (circles) conditions. Lines connect the data points for each listener. Data are plotted against the inter-subject spectral difference between the other-ear DTFs (scaled or unscaled) and each listener's own DTFs. Frequency scaling produced substantial reductions in the rms local polar errors and in the percentage of quadrant errors relative to the other-ear condition in nearly all cases. In 17 of 21 cases, that reduction was proportionately greater than the reduction in the spectral difference between DTFs.

Figure 12 compares quadrant errors and rms local polar errors in the scaled-ear versus other-ear conditions. Errors

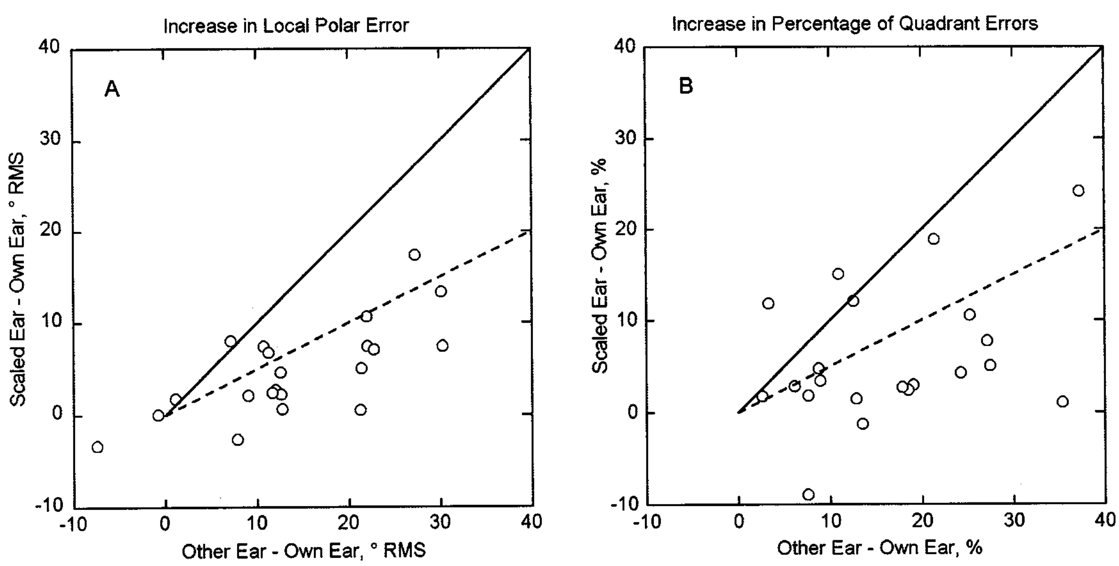

FIG. 12. rms local polar errors (A) and percent quadrant errors (B) in the scaled-ear versus the other-ear condition. Errors are expressed as the increase in rms error or the increase in percent quadrant errors relative to the own-ear condition. The solid line represents equal performance in other- and scaled-ear conditions, and the dashed line indicates 50% reduction in the increase in error.

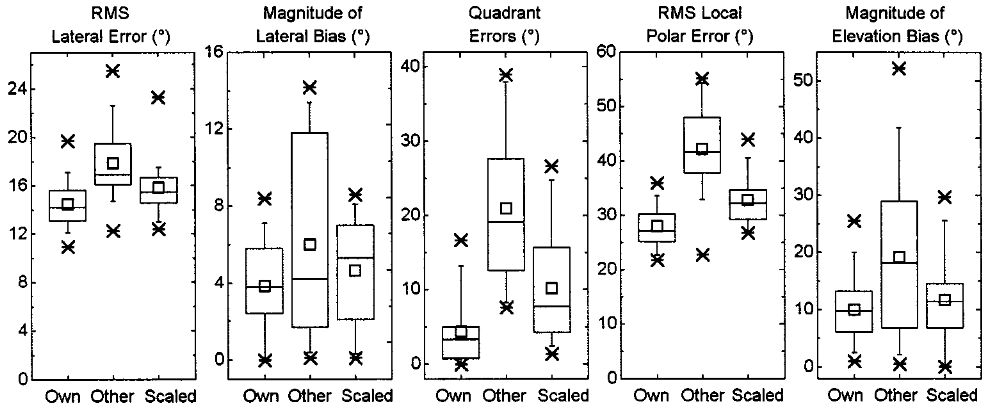

FIG. 13. Summary of performance in three virtual localization conditions. In these box plots, horizontal lines represent 25th, 50th, and 75th percentiles, the extended vertical bars represent 5th and 95th percentiles, and ×'s represent minima and maxima. The small squares indicate the means. Eleven listeners were tested in the *own* condition. The same 11 listeners were tested with 1–4 sets of DTFs from other subjects for a total of 21 cases in the *other* and *scaled* conditions.

are expressed as increases in the magnitude of rms local polar errors and the percentage of quadrant errors in scaled-ear and other-ear conditions relative to the own-ear condition. Two-thirds of points (14/21) fell below the dashed line with slope of $\frac{1}{2}$, indicating that scaling DTFs in frequency reduced by more than half the penalty for listening through another subject's DTFs. Scaling DTFs in frequency resulted in improvements relative to the other-ear condition in every summary measure of localization performance (Fig. 13).

A previous paper proposed a simple model to account for listeners' systematic errors in localization of narrow-band sounds (Middlebrooks, 1992). That model, adapted to the current results, provides some understanding of listeners' localization judgements in the other-ear and scaled-ear conditions. The model assumed that each listener's auditory system maintains in some form a library of templates of the DTFs associated with each sound-source direction. It was speculated that when a stimulus arrives at the tympanic membranes, the listener compares that proximal stimulus with templates of DTFs, and that the listener's localization judgement corresponds to the localization for which the template fits most closely. In the current study, the DTF-matching component of that model was applied to data from the other-ear and scaled-ear conditions. To conform to the procedure in this and the companion paper, comparisons were made among DTFs by computing variances of inter-subject DTF difference spectra instead of by computing correlations as in the previous study. Listeners tended to localize accurately in the lateral dimension regardless of the testing condition. For that reason and to focus on spectral shape recognition, the present analysis neglected interaural disparities and the lateral dimension of localization judgements and simply assumed that listeners evaluate DTF spectra only within the correct lateral plane (i.e., on the correct cone of confusion). The model predicted that listeners would report apparent source locations at points that minimize inter-subject variance in DTFs. Figures 14 and 15 show examples of the application of the DTF-matching model to results from one listener localizing with DTFs from two other subjects.

Figures 14 and 15 both represent localization of a virtual target at a single location: lateral $-20°$, polar $-21°$ (20° to the left, 21° below the horizon in front). In each figure, the target location is indicated by a dashed vertical line, and the listener's localization judgements are indicated by triangles. The same listener (S25) localized using DTFs from a larger subject (S27, Fig. 14) and from a smaller subject (S35, Fig. 15). The proximal stimulus in each case was synthesized with the other-ear DTF for $-20°$, $-21°$, either unscaled (upper panels) or scaled (lower panels). Each plot shows the

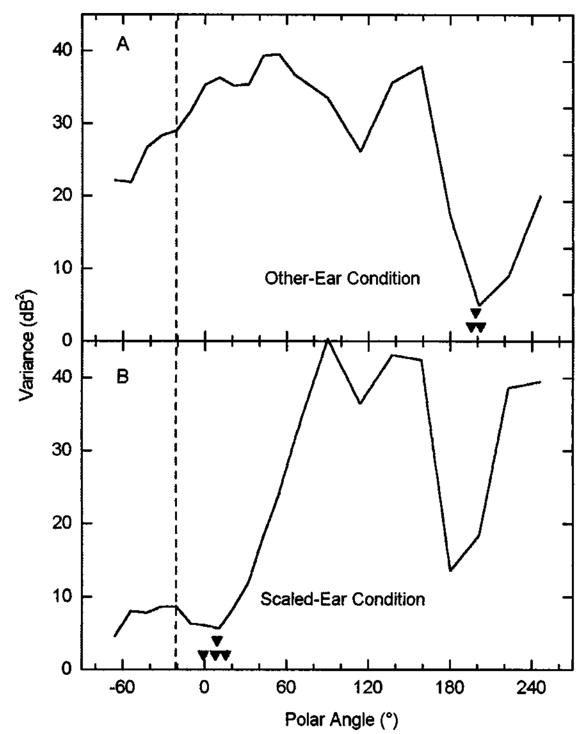

FIG. 14. Polar localization judgements related to variance in DTFs. Localization of a virtual target at left 20°, polar $-21°$ is shown for listener S25 using DTFs from a larger subject, S27. The curve in each plot represents the variance of the difference spectrum between the DTF that was used to synthesize the target (i.e., S27's DTF for -20, $-21°$) and the listener's own DTFs at each of a range of polar angles at lateral 20°. Vertical dashed lines indicate the target location. Triangles indicate the localization judgements, in some cases displaced upward to avoid overlap. In A, the target DTF was used unscaled. In B, it was scaled upward in frequency by a factor of 1.22.

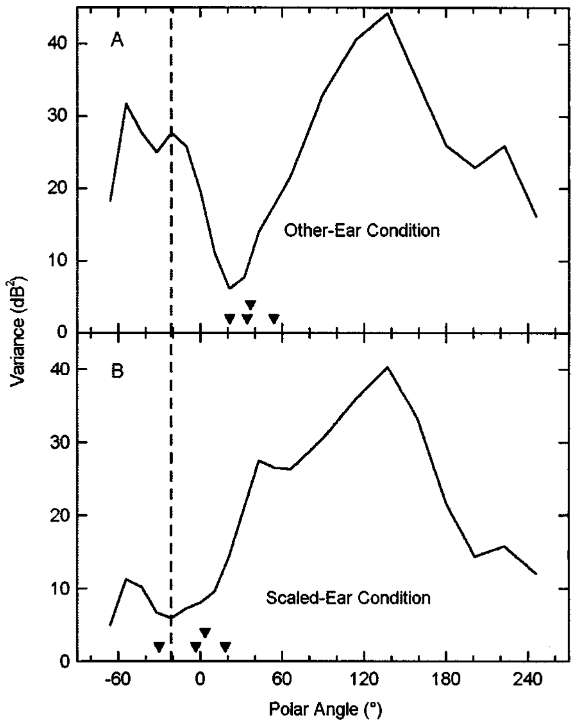

FIG. 15. Polar localization judgements related to variance in DTFs. The format is identical to that in Fig. 14, as are the listener (S25) and the target location. The stimuli in this case were synthesized with DTFs from a smaller subject, S35. In the scaled condition, S35's DTF for −21° was scaled downward by a factor of 0.889.

variance of the difference spectrum between the stimulus spectrum and the listener's own DTFs at all polar angles at lateral −20°. In the other-ear condition in which the DTF came from a larger subject [Fig. 14(A)], the variances between DTFs were fairly high near the virtual-target location but showed a sharp minimum near polar 200° (20° below the horizon in the rear). The listener's three responses to the −21° target fell near that minimum. When the other-ear DTF for polar −21° was scaled in frequency, the variance with the listener's own DTFs showed an entirely different pattern [Fig. 14(B)]. In that condition, local minima in the variance plot appeared at polar −60°, around 0°, and around 180°. The listener's four responses fell in the region of low variance near 0°.

In the other-ear condition in which the DTFs came from a smaller subject [Fig. 15(A)], the variance plot showed a sharp minimum near +20°, 41° higher than the target. The four localization judgments showed an upward elevation bias, falling near the minimum in the variance plot. In the scaled-ear condition [Fig. 15(B)], the minimum in the variance plot shifted downward to lie on top of the target location, and the listener's responses shifted downward by a corresponding amount.

Plots of variance versus polar angle, like those in Figs. 14 and 15, could predict qualitative characteristics of most listeners' localization judgements in the other-ear and scaled-ear conditions. For instance, among the quartile of small listeners localizing through larger subjects' DTFs, the variance plots computed for down-front targets consistently showed a local minimum in the up-rear or down-rear quadrant that corresponded to listeners' responses. Among the quartile of larger listeners localizing through smaller subjects' DTFs, the variance plots computed for down-front targets consistently showed a minimum at frontal locations with an upward elevation bias. Variance plots often showed two local minima, one for a frontal location and one for a rear location, with subjects' localization judgements distributed between those minima. Examples of variance plots containing both front and rear minima are considered in the next section.

D. Sensitivity of localization accuracy to frequency scale factors

A separate series of behavioral measurements tested the influence of the values of frequency scale factors on localization performance. To economize listener time, virtual targets were restricted to the vertical midline. Frequency scale factors were varied in powers of 1.04, and stimuli synthesized with various scale factors were interleaved within blocks of trials. The listener pool consisted of five listeners who had no experience in virtual localization plus four listeners who were experienced virtual localizers. The naïve listeners tended to localize less accurately than the experienced listeners, but the dependence of localization accuracy on frequency scale factor was similar between naïve and experienced listeners.

The accuracy of virtual localization tended to show sharp sensitivity for the value of the frequency scale factor. This is shown in Fig. 16, in which the solid curves and the right ordinates show the percentage of quadrant errors of listener S16 who localized with his own DTFs, scaled by various factors [Fig. 16(A)] or with scaled DTFs measured from listener S04 [Fig. 16(B)]. The sharpness of sensitivity to frequency scale factors was estimated from tuning plots like those in Fig. 16 by computing the width of the plot across which the quadrant error was within 5 percentage points of the minimum. Widths computed in that way are expressed as the maximum scale factor within the 5-percentage point criterion divided by the minimum within that criterion. Widths ranged from 1.13 to 1.22 (median = 1.15) for the four listeners who were tested with their own DTFs and from 1.04 to 1.22 (median = 1.15) for nine listeners who were tested with other subjects' DTFs.

The frequency scale factors that resulted in best behavioral performance tended to correspond to the frequency scale factors that minimized the inter-subject spectral difference. The dashed curves and left ordinates in Fig. 16 show the spectral difference between one listener's own DTFs, unscaled, and his own or another subject's DTFs scaled by various factors. Across all 13 cases in which listeners localized with their own or another subject's DTFs scaled by varying factors, the best frequency scale factors obtained from behavioral and acoustical measurements correlated with $r = 0.92$. The rms difference between the logarithms of scale factors measured behaviorally and acoustically, expressed as a linear factor, was 1.052, which is little more than the minimum increment in frequency scale factors (1.04) that was tested in the behavioral trials.

When a listener's DTFs are scaled downward or upward in frequency, the DTFs tend to approximate those of a larger

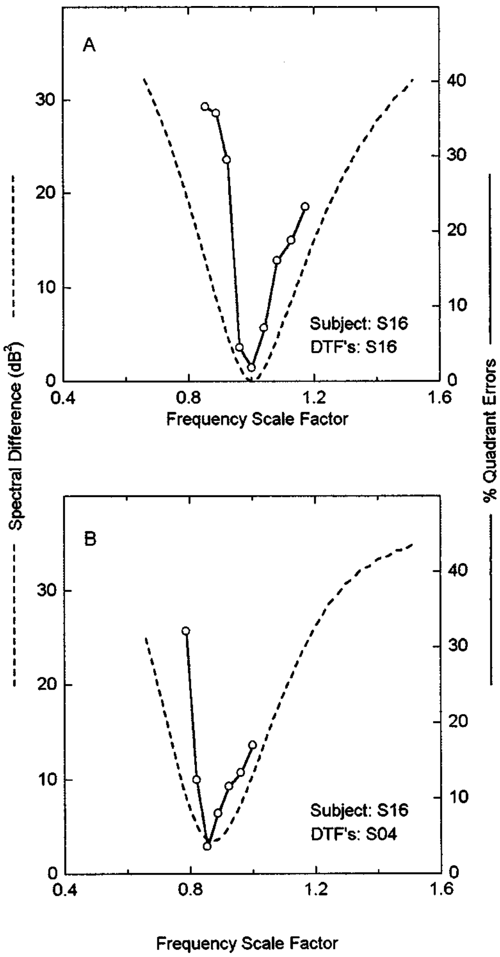

FIG. 16. Sensitivity of the percent of quadrant errors to frequency scale factor. Solid curves and right ordinate show localization performance in the vertical midline by listener S16. The DTFs were from the listener's own ears (A) or from subject S04 (B). DTFs were scaled in frequency by the scale factor that is plotted on the abscissa. Dashed curves and left ordinate show the spectral variance between the listener's own DTFs (unscaled) and either his own DTFs (A) or S16's DTFs (B), scaled by the value plotted on the abscissa.

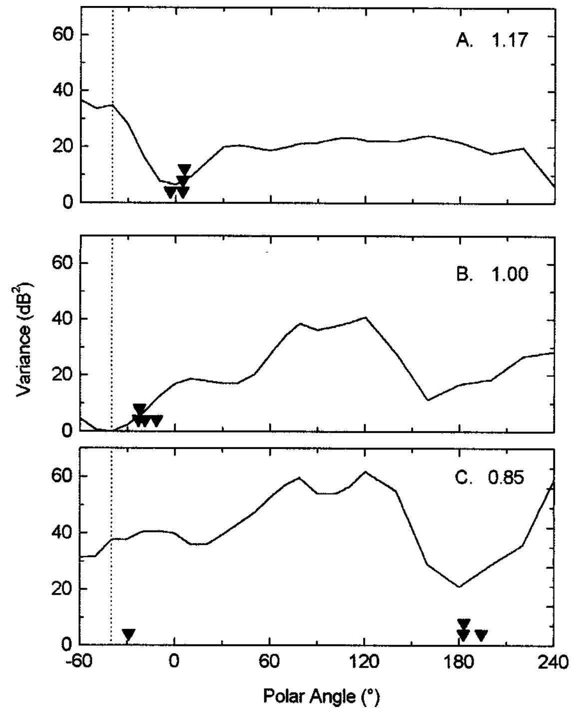

FIG. 17. Influence of frequency scale factors on cues for localization in polar dimension. Data are from trials in which listener S40 localized virtual targets at 0°, −40° using his own DTFs scaled in frequency by 1.17 (A), 1.00 (B), and 0.854 (C). Other conventions as in Fig. 14.

or smaller subject, respectively. For that reason, localization with DTFs scaled by various amounts simulates localization with DTFs from a continuum of other subjects. Figure 17 represents one listener who localized with his own DTFs scaled in frequency by three scale factors. In each case the target was on the frontal midline at −40° elevation. The curves in each panel plot the variance in the difference spectrum between the listener's DTF for −40° elevation, scaled by the stated amount, and the listener's unscaled DTFs for a range of locations in the vertical midline. The vertical dashed line indicates the target location, and the filled triangles indicate the listener's four responses in each condition. When the scale factor was one [Fig. 17(B)], the variance was zero at the target location, and the listener's responses fell slightly above that elevation. In that panel, one can see a second, less deep, minimum in the variance plot located in the rear, around 160°, but none of the listener's responses fell there. As the stimulus DTF was scaled to a lower frequency [Fig. 17(C)] the variance for DTFs at frontal locations grew larger than that in the rear, and the minimum in the variance plot at rear locations became the overall minimum. In that condition, 3/4 of the listener's judgements fell in the rear and one fell near the low-variance locus in front. The variance plot in that condition resembles the variance plot shown for a condition of a listener localizing through larger subject's DTFs (Fig. 14, top panel). When the listener's DTFs were scaled upward in frequency [Fig. 17(A)], the frontal minimum in the variance plot and the listener's responses shifted upward in elevation. The condition of an upward shift in DTFs produced a variance plot similar to that observed when a listener used a smaller subject's DTFs (Fig. 15, top).

III. DISCUSSION

The results that have been presented support two main conclusions. *First*, consistent with previous observations, sound-localization accuracy is degraded when listeners listen to targets synthesized with DTFs measured from ears other than their own. The results extend the previous observations by demonstrating particular patterns of errors that result from use of DTFs from larger or smaller subjects. Listeners showed systematic overshoots or undershoots in the lateral dimension when listening through DTFs from larger or smaller subjects, respectively. In the polar dimension, when listeners localized with DTFs from larger subjects, they tended to localize down-front targets to down-rear and they tended to localize up-rear targets to up-front. When listeners localized through DTFs from smaller subjects, the most conspicuous error was a tendency to displace low targets upward. Possible explanations for those patterns of errors are considered in Sec. III B. *Second*, the results support the working hypothesis that, when a listener localizes through other subjects' DTFs, his or her virtual localization can be improved by scaling the DTFs in frequency to minimize the spectral difference between those DTFs and the listener's own DTFs. In two-thirds of cases, the penalty for use of nonindividualized DTFs was reduced by more than half. Section III C considers the feasibility of tailoring a generic set of DTFs to individual listeners.

A. Relation to previous results

Localization performance in the free-field condition was similar to that in our previous study (Makous and Middlebrooks, 1990). The rate of quadrant errors was comparable to the rate of ''front/back confusions'' in that study. The previous study did not quantify rates of quadrant errors in specific quadrants, but Fig. 5 in that paper showed that quadrant errors were most common in response to up-rear targets. In the own-ear virtual condition, performance was similar to that in the study by Wightman and Kistler (1989b). Differences in the methods of analysis preclude quantitative comparisons, but performance errors in that study seem to be of roughly the same magnitude as in the current study, and that study showed, as in the current study, that localization errors tended to be largest for targets in the up-rear quadrant.

Several previous studies have examined virtual localization under an other-ear condition. The study most similar to the current one is by Wenzel and colleagues (1993). Those authors studied localization by 16 listeners who all heard targets synthesized from the HRTFs of a 17th subject (subject SDO, from the study by Wightman and Kistler, 1989a). Compared with own-ear data from Wightman and Kistler (1989b), subjects in the study by Wenzel and colleagues showed substantial increases in the rates of front/back and up/down confusions under the other-ear condition. Similar results were reported by Morimoto and Ando (1980), who tested three listeners that were selected for small, medium, and large external ears. Each listener localized most accurately in the own-ear condition. In the other-ear condition, each listener showed reasonable performance in the lateral dimension but an increase in the front/back confusions and in the scatter of responses in the vertical dimension.

Møller and colleagues (1996) tested virtual localization in a condition in which listeners distinguished among a limited number of virtual sources. Listeners showed a significant increase in errors in the vertical dimension in the other-ear compared to own-ear condition. Some of the most common errors resembled those observed in the present study. Specifically, down-front targets often were judged as back, characteristic of small listeners localizing through larger subjects' DTFs in the current study, and sources on or below the horizontal plane often were judged with an upward elevation bias, like the large listeners localizing through smaller subjects' DTFs. The rates of quadrant errors in the study by Møller and colleagues were much lower than in the present study, but the stimuli in that study were substantially longer than in the current study: 5 s of speech sounds, compared to 250 ms of Gaussian noise.

Butler and Belendiuk (1977) and Bronkhorst (1995) employed spatial discrimination tasks in which subjects discriminated among a limited number of virtual-target locations. Bronkhorst's ''confusion'' protocol tested the accuracy of identification of target quadrants. Surprisingly, there was no significant difference between own-ear and other-ear conditions. Bronkhorst speculated that the relatively high confusion rates among his inexperienced listeners might have overwhelmed any difference among conditions. Also, Bronkhorst's subject population might have been more uniform than in the present study, inasmuch as it consisted of seven males and only one female. Thus inter-subject differences among sets of HRTFs in that study might have been lower than in the present study. Butler and Belendiuk (1977) reported a result that is often cited as evidence that a listener can localize more accurately with the HRTFs of an accurate localizer than with his or her own. In that study, four listeners each attempted to localize five targets by listening to tapes of microphone recordings from their own ears and from the ears of the three other listeners. Three of the listeners tended to localize more accurately in the own-ear condition. In contrast, the fourth listener (''P.B.'') localized better when listening to the tapes from other listeners than when listening to his/her own tapes. Likewise, the other three listeners showed their worst performance when listening to P.B.'s tape. One interpretation of that result is that the ears of listener P.B. somehow failed to produce useful cues to elevation and that P.B. could localize better by listening through ears that produced more reliable cues. An alternate interpretation is that there might have been a technical problem with the recording of P.B.'s tape, such as some uncertainty regarding the positions of recording microphones in the ears. Also, stimulus spectra in that study were limited to frequencies below 9 kHz, which might have eliminated or distorted high-frequency spectral cues.

In a preliminary report, Wightman and Kistler (1993) described an analysis of HRTFs from 15 listeners. Listeners localized virtual targets that were synthesized with each other's HRTFs. The differences among subjects' HRTFs were characterized by their Euclidean distances. Euclidean distance is similar to the spectral difference metric used in the present study, except that the metric used here eliminated frequency-invariant offsets in DTFs for individual locations. Consistent with the present results, performance in the Wightman and Kistler study generally was more accurate when listeners used HRTFs that were closer to their own in Euclidean distance than when they used HRTFs that were more distant.

B. Implications for models of sound localization

Most models of sound localization acknowledge that listeners' localization judgements are derived from some combination of interaural difference cues and spectral shape cues (see Middlebrooks and Green, 1991, for review). Our procedure for scaling DTFs in frequency involved changing the output rate of the digital-to-analog converter, which resulted in a concomitant upward scaling in frequency of DTFs and of interaural difference spectra and a shortening of interaural delays (or downward scaling in frequency and lengthening of delays). For that reason, the design of the study did not lend itself to independent manipulation of specific localization cues. In another sense, however, this equivalent scaling of multiple localization cues was ideal for the purpose of aligning DTFs functions among subjects. Results of the companion study indicated that optimal frequency scale factors for DTFs were very highly correlated with the optimal scale factors for interaural difference spectra ($r=0.96$) and showed substantial correlation with the inter-subject ratios of maximum interaural delays ($r=0.71$). That is, adjustment of a single parameter, the output rate, could simultaneously minimize inter-subject differences in spectral shape cues, interaural level differences, and interaural delays.

In the lateral dimension, listeners tended to overshoot the target (i.e., point too far to the side) when listening through the DTFs of a larger subject and tended to undershoot when listening through the DTFs of a smaller subject. Scaling DTFs in frequency resulted in a significant reduction in lateral biases. Nevertheless, the typical increase in magnitude of the lateral bias that resulted from use of another subject's DTFs, scaled or not, was fairly small compared to the scatter in the lateral component of localization judgements. That is, the mean change in unsigned magnitudes of biases was 2.7°, whereas the rms average of lateral errors in the own-ear condition was 14.7°.

In the polar dimension, interaural differences tend to be ambiguous, so listeners are thought to rely on spectral shape cues for localization in that dimension. Models that attempt to explain the use of spectral shape cues generally assume, either explicitly or implicitly, that a listener has some knowledge of the characteristics of his or her own DTFs that are associated with particular sound-source locations. Some models have emphasized the importance of specific spectral features, such as ''boosted bands'' (Blauert, 1969/70), ''covert peaks'' (Humanski and Butler, 1988), or other spectral peaks, notches, or slopes (e.g., Hebrank and Wright, 1974; Bloom, 1977; Watkins, 1978). When listeners attempted to localize narrow-band noise bursts in a previous study (Middlebrooks, 1992), their localization judgements in the polar dimension were determined more by the center frequency of the noise burst than by the actual location of the source. The location that a listener reported in response to a particular center frequency tended to vary with the physical height of the listener. That paper reported that ''in general, the spatial distributions of responses were most similar among subjects of similar physical size, and a given distribution of responses at a particular frequency observed for a tall subject tended to be found at a higher frequency for a shorter subject (Middlebrooks, 1992, p. 2611).'' Based on the companion study (Middlebrooks, 1999), one would account for that result by noting that the spectral features found in DTFs of smaller subjects tend to be found at higher frequencies relative to those of larger subjects.

As an alternative to examining specific spectral peaks and notches, several investigators have compared HRTFs or DTFs with proximal stimulus spectra across relatively broadbands (Watkins, 1978; Middlebrooks, 1992; Zakarauskas and Cynader, 1993; Hofman and Van Opstal, 1998). Studies that used narrow-band free-field stimuli (Middlebrooks, 1992) and broadband virtual stimuli (the present study) both show that localization judgements tend to fall near locations at which the associated DTF most closely resembles the proximal stimulus. Previous localization studies have reported large ''front/back'' or ''up/down'' confusions, which are referred to here as ''quadrant errors.'' The variance plots in Figs. 14, 15, and 17 demonstrate a likely explanation for the quadrant errors in that the plots often showed two widely separated local minima. A discontinuous distribution of localization judgements, with some judgements near the target and others far away, might reflect the dual minima in variance plots.

In the companion study, a residual inter-subject spectral difference in DTFs of around 6.2 dB^2 remained after scaling DTFs in frequency by an optimal scale factor. That residual spectral difference was largely independent of the initial frequency disparity among the DTFs. Inspection of DTFs indicates that frequency scaling is successful in aligning major spectral features along the frequency axis, but scaling can have no effect on idiosyncratic features such as the detailed shapes or magnitudes of peaks or notches. In the current behavioral results, the percentage improvement in localization performance that resulted from scaling DTFs in frequency often was greater than the percentage reduction in spectral difference. That suggests that the general distribution of spectral features in frequency is more important for virtual localization than is the detailed shapes of those features. That conclusion is supported by the work of Kistler and Wightman (1992), who attempted to reconstruct HRTFs from small numbers of principal components. They found that reconstructions based on as few as three components supported some degree of virtual localization in front/back and up/down dimensions and that reconstructions based on only five components supported virtual localization that was nearly as accurate as that obtained with intact HRTFs. Similarly, Kulkarni and Colburn (1998) found that a veridical virtual synthesis could be obtained with highly smoothed HRTFs.

C. Tailoring generic DTFs to specific users

The present results confirm observations by others that performance in virtual localization is most accurate when listeners use DTFs measured from their own ears. Nevertheless, it is not feasible to make detailed DTF measurements from every listener who might want to utilize a virtual acoustic environment. Situations are considered here in which it might or might not be acceptable to use generic ''nonindividualized'' DTFs, and practical procedures are evaluated for customizing generic sets of DTFs to individual listeners.

In evaluating the potential impact of the use of nonindividualized DTFs for a particular application, one must consider the requirements of the application. In the horizontal dimension, the liability for use of nonindividualized DTFs for horizontal localization is relatively small. In the data, the rms errors in the horizontal dimension averaged 17.1° in the other-ear condition compared to 14.7° in the own ear condition, an increase in error of only 16%. The largest increases in lateral bias in the other-ear condition were around 8°, which is less than 10% of the entire range of lateral angle from 0° to 90°. Moreover, the largest lateral biases were observed only in cases in which the largest listeners localized through DTFs from the smallest subjects, and *vice versa*.

Lateral bias could be minimized by maintaining sets of DTFs from one large subject and from one small subject and choosing the appropriate set for use by large or small listeners. In addition, Shinn-Cunningham and colleagues (1998) have shown that, given feedback, listeners can adapt to modified horizontal localization cues. In applications that require only localization of virtual targets that are known to be restricted to the frontal horizontal plane, nonindividualized DTFs might be entirely adequate.

In contrast to the results in the horizontal dimension, localization in the vertical and front/back dimensions showed, on average, a quadrupling of the rate of quadrant errors and a sizeable increase in rms local polar errors when listeners localized through other subjects' DTFs. It is difficult to imagine an application in which such degradation in performance would be acceptable. Our results from scaling of DTFs suggest two ways in which one might improve virtual localization with nonindividualized DTFs. One possibility would be to maintain a single set of DTFs from one listener and to adapt those DTFs to individual listeners by scaling in frequency. Alternatively, one might maintain sets of DTFs from multiple subjects, indexed by a mean frequency scale factor. An individual listener would select the optimal set of DTFs on the basis of his or her own mean frequency scale factor. The difficulty with either of these schemes is that both require knowledge of a mean frequency scale factor for each listener. That is, if one has DTF measurements from subjects A and B, one can scale A to match B, but one does not know the scale factor needed to scale A to another subject, subject C, from whom there are no acoustical measurements.

One way to obtain a frequency scale factor might be to estimate it from physical measurements of the width of a subject's head and the height of his or her external ear. The companion paper compared scale factors estimated from those two parameters with scale factors measured from full sets of DTFs from pairs of subjects. The rms error in estimation was only 5.8%. Alternatively, one might attempt to identify an optimal scale factor behaviorally by conducting virtual localization trials with a range of scale factors. That approach, when tested with virtual targets restricted to the vertical midline, yielded scale factors that differed from acoustically measured scale factors by an rms value of 5.2%. That accuracy is only slightly better than the estimates obtained from physical measurements of the head and ear. Moreover, the behavioral approach required about 2 h of testing per listener.

It is not yet clear how precisely a frequency scale factor would need to be measured in order to obtain a significant improvement in behavioral performance. When virtual localization was tested for targets restricted to the vertical midline, the rate of quadrant errors increased rapidly as the frequency scale factor was varied away from the optimal value (e.g., Fig. 16), with widths of the tuning curves ranging from 4% to 22% across listeners. This suggests that an error of, say, 6% in estimating a frequency scale factor might produce a small but noticeable degradation in virtual localization performance. Nevertheless, in acoustical measurements from a population of 45 subjects (Middlebrooks, 1998), the scale factors obtained from every pairwise combination of subjects ranged from 0.72 to 1.38. Therefore, an estimate of a scale factor with an rms error of 6% (i.e., a factor of 1.06) would substantially reduce the uncertainty in the scale factor.

One final possibility that we have not evaluated is that, with experience, subjects might adapt to another subject's DTFs. In the present study, own-ear, other-ear, and/or scaled-ear trials normally were interleaved within a test session, so a listener seldom heard more than two or three DTFs in a row from the same set. That issue was addressed to some extent by Møller and colleagues (1996), who compared two other-ear conditions: one in which all the HRTFs were from one subject and the other in which HRTFs from several subjects were interleaved. They observed no significant difference in performance in the two conditions. Another issue is one of feedback. All of the trials in the present study were conducted without trial-by-trial feedback, so listeners had no means of associating particular other-ear HRTFs with correct locations. Shinn-Cunningham and colleagues (1998) reported that, when provided with trial-by-trial feedback, subjects could adjust to modified horizontal localization cues. A recent report by Hofman and colleagues (1998) tested the ability of listeners to adapt to long-term modifications of the acoustics of their own external ears. Subjects wore plastic inserts in their external ears. Initially after placement of the inserts, the listeners showed dramatic errors in free-field localization. After wearing the inserts for a period of 6 weeks, however, the listeners' free-field localization performance improved substantially. The changes in HRTFs introduced by the plastic inserts almost certainly were larger than the abnormalities in HRTFs that would be encountered by a listener who used nonindividualized HRTFs that were scaled imperfectly in frequency.

D. Concluding remarks

The companion study (Middlebrooks, 1999) demonstrated that, for any pair of subjects, scaling in frequency with a single scale factor can reduce substantially the intersubject differences in DTFs for locations throughout the coordinate sphere. The present study exploited that result to attempt to improve virtual localization in conditions in which a listener uses DTFs measured from another subject's ears. The results of that exercise actually were better than expected, in that the percentage improvements in localization performance often were greater than the percentage reduction in spectral difference between sets of DTFs. The sign of the frequency difference between the DTFs through which a listener localized and the listener's own DTFs turned out to be predictive of particular types of localization errors. That is, localization errors differed depending on whether spectral features in DTFs generally lay at higher or lower frequencies than spectral features in a listener's own DTFs. Those localization errors can be accounted for qualitatively by assuming that listeners' localization judgements tend to correspond to locations for which their own DTFs best resembles the DTF that was used to synthesize the virtual target.

The companion study showed that optimal frequency scale factors between pairs of subjects can be predicted with some accuracy from certain physical dimensions of subjects, and the present behavioral results suggest that those predictions likely are accurate enough to provide appreciable improvements in virtual localization. That result offers encouragement for the use of nonindividualized DTFs. Specifically, in a setting in which one had no facility for measuring an individual listener's DTFs, one could estimate a frequency scale factor from measurements of the width of the listener's head and the height of the external ear. That scale factor then would be used to tailor a generic set of DTFs so that they approximate the listener's own DTFs fairly closely. Initial errors in virtual localization would be corrected with feedback. One would anticipate that this procedure would result in a realistic virtual synthesis of auditory space for the large number of listeners for whom it would not be practical to make individual acoustical measurements of DTFs.

ACKNOWLEDGMENTS

I am pleased to acknowledge the contributions of Zekiye Onsan, who collected the behavioral data and made the illustrations. Shigeto Furukawa, Ewan Macpherson, Brian Mickey, and Li Xu provided stimulating discussion and commented on an early draft of the manuscript. I thank Professor Frederic Wightman and a second anonymous reviewer for constructive comments on an earlier draft. This research was supported by ONR Grant No. N0014-94-1-0481 and NIH Grant No. DC000420.

Blauert, J. (**1969–1970**). ''Sound localization in the median plane,'' Acustica **22**, 205–213.

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

Bronkhorst, A. W. (**1995**). ''Localization of real and virtual sound sources,'' J. Acoust. Soc. Am. **98**, 2542–2553.

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.

Carlile, S., Leong, P., and Hyams, S. (**1997**). ''The nature and distribution of errors in sound localization by human listeners,'' Hearing Res. **114**, 179–196.

Foster, S. (**1986**). ''Impulse response measurement using Golay codes,'' in *IEEE 1986 Conference on Acoustics, Speech, and Signal Processing, Vol. 2* (IEEE, New York), pp. 929–932.

Golay, M. J. E. (**1961**). ''Complementary series,'' IRE Trans. Inf. Theory **7**, 82–87.

Hebrank, J., and Wright, D. (**1974**). ''Spectral cues used in the localization of sound sources on the median plane,'' J. Acoust. Soc. Am. **56**, 1829–1834.

Hofman, P. M., and Van Opstal, J. A. (**1998**). ''Spectro-temporal factors in two-dimensional human sound localization,'' J. Acoust. Soc. Am. **103**, 2634–2648.

Hofman, P. M., Van Riswick, J. G. A., and Van Opstal, J. A. (**1998**). ''Relearning sound localization with new ears,'' Nature Neurosci. **1**, 417–421.

Humanski, R. A., and Butler, R. A. (**1988**). ''The contribution of the near and far ear toward localization of sound in the sagittal plane,'' J. Acoust. Soc. Am. **83**, 2300–2310.

Kistler, D. J., and Wightman, F. L. (**1992**). ''A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction,'' J. Acoust. Soc. Am. **91**, 1637–1647.

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

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

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

Middlebrooks, J. C. (**1998**). ''Individual diffrences in external-ear transfer functions reduced by scaling in frequency,'' J. Acoust. Soc. Am.

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

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

Middlebrooks, J. C., and Green, D. M. (**1990**). ''Directional dependence of interaural envelope delays,'' J. Acoust. Soc. Am. **87**, 2149–2162.

Møller, H., Sorensen, M. F., Jensen, C. B., and Hammershoi, D. (**1996**). ''Binaural technique: Do we need individual recordings?,'' J. Aud. Eng. Soc. **44**, 451–469.

Morimoto, M., and Ando, Y. (**1980**). ''On the simulation of sound localization,'' J. Acoust. Soc. Jpn. (E) **1**, 167–174.

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

Pralong, D. and Carlile, S. (**1996**). ''The role of individualized headphone calibration for the generation of high fidelity virtual auditory space,'' J. Acoust. Soc. Am. **100**, 3785–3793.

Shinn-Cunningham, B. G., Durlach, N. I., and Held, R. M. (**1998**). ''Adapting to supernormal auditory localization cues. I. Bias and resolution,'' J. Acoust. Soc. Am. **103**, 3656–3666.

Watkins, A. J. (**1978**). ''Psychoacoustical aspects of synthesized vertical locale cues,'' J. Acoust. Soc. Am. **63**, 1152–1165.

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

Wightman, F. L., and Kistler, D. J. (**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.

Wightman, F. L., and Kistler, D. J. (**1992**). ''The dominant role of low-frequency interaural time differences in sound localization,'' J. Acoust. Soc. Am. **91**, 1648–1661.

Wightman, F. L., and Kistler, D. J. (**1993**). ''Multidimensional scaling analysis of head-related transfer functions,'' Proc. ASSP (IEEE) Workshop on Applications of Signal Processing to Audio and Acoustics.

Woodworth, R. S. (**1938**). *Experimental Psychology* (Holt, Rinehart, and Winston, New York), pp. 349–361.

Zakarauskas, P., and Cynader, M. S. (**1993**). ''A computational theory of spectral cue localization,'' J. Acoust. Soc. Am. **94**, 1323–1331.

Zhou, B., Green, D. M., and Middlebrooks, J. C. (**1992**). ''Characterization of external ear impulse responses using Golay codes,'' J. Acoust. Soc. Am. **92**, 1169–1171.