

Individual differences in external-ear transfer functions reduced by scaling 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 10 March 1999; accepted 26 May 1999)

This study examined inter-subject differences in the transfer functions from the free field to the human ear canal, which are commonly known as head-related transfer functions. The directional components of such transfer functions are referred here to as *directional transfer functions* (DTFs). The DTFs of 45 subjects varied systematically among subjects in regard to the frequencies of spectral features such as peaks and notches. Inter-subject *spectral differences* in DTFs were quantified between 3.7 and 12.9 kHz for sound-source directions throughout the coordinate sphere. For each pair of subjects, an optimal frequency scale factor aligned spectral features between subjects and, thus, minimized inter-subject spectral differences. Frequency scaling of DTFs reduced spectral differences by a median value of 15.5% across all pairs of subjects and by more than half in 9.5% of subject pairs. Optimal scale factors showed a median value of 1.061 and a maximum of 1.38. The optimal scale factor between any pair of subjects correlated highly with the ratios of subjects' maximum interaural delays, sizes of their external ears, and widths of their heads.

© 1999 Acoustical Society of America. [S0001-4966(99)01409-5]

PACS numbers: 43.66.Pn, 43.64.Ha, 43.66.Qp [DWG]

INTRODUCTION

The physical structures of the head and external ears transform the spectrum of a sound as it travels from the source to the ear canal. The transfer function from the source to any of several points in the ear canal has come to be known as the head-related transfer function (HRTF; e.g., Morimoto and Ando, 1980; Butler, 1987; Wightman and Kistler, 1989a). The HRTF contains prominent spectral features, such as peaks and notches, that vary according to source direction (e.g., Batteau, 1967; Blauert, 1968/69; Shaw, 1974). Direction-specific spectral features provide cues to the directions of sound sources (reviewed by Middlebrooks and Green, 1991). When a listener hears through headphones sounds that have been filtered with HRTFs measured from his or her own ears, a "virtual acoustic environment" results. The listener reports sensations of sounds that appear to originate from well-defined directions in the space surrounding the subject (Wightman and Kistler, 1989a, b; Bronkhorst, 1995; Hartmann and Wittenberg, 1996; Møller *et al.*, 1996). Subjects can localize such virtual targets with accuracy that approaches the accuracy of localization of actual free-field sound sources. Conversely, when targets are synthesized using HRTFs that have been recorded from the ears of a different subject, the virtual environment is less vivid and localization is less accurate. In particular, virtual localization with nonindividualized HRTFs is characterized by a high incidence of large errors in vertical and front/back localization (Morimoto and Ando, 1980; Wenzel *et al.*, 1993; Møller *et al.*, 1996). The goal of the present study was to identify systematic inter-subject differences in HRTFs that might account for this degradation of the virtual environment.

Previous reports from several research groups have men-

tioned inter-subject differences in HRTFs (e.g., Shaw, 1966; Mehrgardt and Mellert, 1977). For instance, Wightman and Kistler (1989a) reported that inter-subject standard deviations within 1/3-octave bands of HRTFs could reach nearly 8 dB for bands centered between 7 and 10 kHz. In our previous work, we have noted that the directional sensitivity (i.e., direction-specific amplitude gain) of a particular subject's ear at one frequency can closely resemble the directional sensitivity of another subject's ear at a different frequency (Middlebrooks *et al.*, 1989). Moreover, we found that the frequencies at which subjects showed similar directional sensitivity tended to rank inversely with the overall sizes of subjects. Intuitively, one can think of lower-frequency sounds, with longer wavelengths, as interacting with larger heads and ears in a similar way that higher-frequency sounds, with shorter wavelengths, interact with smaller heads and ears. These observations suggest that it might be possible to reduce inter-subject differences in HRTFs by scaling them in frequency. Reductions in inter-subject differences in HRTFs, in turn, might lead to improvements in virtual localization using nonindividualized HRTFs.

This report considers the directional components of HRTFs, which are referred to as *directional transfer functions* (DTFs). Two hypotheses are tested. The first is that DTFs differ systematically among subjects in regard to the frequencies of spectral features. The second is that the optimal scale factor between the DTFs of two subjects can be predicted by the relative physical sizes of the subjects. The results largely confirm both hypotheses. The companion study (Middlebrooks, 1999) tests the hypothesis that reductions in inter-subject differences produced by frequency scaling can result in improvements in virtual localization.

I. METHODS

A. Subjects

Subjects were recruited from the student population of the University of Michigan and from members and associates of the Kresge Hearing Research Institute. Subjects ranged from 21 to 50 yr in age (30.2 ± 6.3 , mean $\pm$ standard deviation). Twenty-three were male, and 22 were female. On the basis of predominant racial heritage, 34 of subjects were European, 9 were Asian, and 2 were African. No subjects were recruited or excluded because of properties of their ears, but none of the subjects exhibited any conspicuous abnormalities of external ears, and none reported any hearing disorder. Some effort was made to sample subjects representing a broad range of physical sizes. Subjects ranged from 152 to 193 cm in height (171 ± 10.4). During measurements, subjects removed earrings and eyeglasses and fastened long hair so that the paths to the ears were clear.

B. Apparatus and sound generation

Acoustical measurements were made inside a double-walled anechoic chamber. The usable space inside the sound-absorbent cones was $2.6 \times 3.7 \times 3.2$ m. The chamber contained a mechanism that could position two loudspeakers on the surface of an imaginary sphere that was 1.2 m in radius. The mechanism was driven by computer-controlled stepping motors. The loudspeakers were separated by 160° on the sphere, and one loudspeaker or the other could be positioned at any azimuth throughout 360° and at any elevation from 70° below to 90° above the horizontal plane. The subject stood on a platform that was adjusted in height to place the center of his or her interaural axis at the center of the coordinate sphere.

The longest acoustical wavelength of interest in this study was 113.3 mm (at 3 kHz) and the largest head width was 153 mm (Sec. IID). Thus the 1200-mm distance between the loudspeaker and the center of the subject's head corresponded to roughly 10.6 wavelengths and 7.8 head widths. For that reason, the stimulus conditions are regarded as a reasonable approximation of the acoustical far field.

The loudspeakers were Pioneer model TS-879 two-way coaxials. Audio signals were generated and recorded using an Intel-based personal computer that was equipped with a 16-bit analog interface and floating-point array processor from Tucker-Davis Technologies (TDT). Output and input sample rates were 50 000 samples per second.

C. Measurements of transfer functions

Transfer functions were measured by recording from subjects' ear canals with miniature microphones while presenting broadband sounds from various locations. Recordings were made simultaneously from Knowles model 1934 microphones placed in both ears approximately 5 mm deep to the entrances of the ear canals. The transfer function from the free field to a microphone in an ear canal varies substantially with changes in the position of the microphone in the standing-wave pattern in the canal. Nevertheless, we have shown previously that, at any particular frequency, measure-

ments of the directional gain of the ear are largely insensitive to the position of a microphone at any point deep to the ear canal entrance (Middlebrooks *et al.*, 1989). Moreover, we have described a procedure, reviewed below, by which one can isolate a directional component of the transfer function that is essentially independent of microphone position in the ear canal (Middlebrooks and Green, 1990).

Probe stimuli consisted of "Golay codes," which are complementary pairs of binary sequences that have correlation structure equivalent to many repetitions of a single impulse (Golay, 1961; Foster, 1986; Zhou *et al.*, 1992). At each sound-source location, a single pair of 512-point codes was presented at an output rate of 50 000 samples per second. That resulted in a 256-point spectrum that ranged from 0 to 25 kHz with resolution of 97.7 Hz. Normally, for the purpose of rapid data acquisition, complementary Golay codes were presented immediately one after the other. There was initial concern that reverberation from the response to the first code might contaminate the response to the second code, but preliminary control experiments proved that contamination to be negligible. Probe stimuli were presented from 400 loudspeaker locations that ranged through 360° of azimuth and in elevation from -70° to 90° . The locations were distributed in azimuth and elevation so that each location represented a roughly equal area on the coordinate sphere and so that neighboring locations were separated by angles of no more than 10° measured at the center of the sphere. At each location, the probe sound level was adjusted so that the amplified microphone signal into the analog-to-digital converter was 50%–99% of the clipping level of the converter.

On the day of each set of measurements, the transfer functions of the two loudspeakers were measured using the Golay probe stimuli and a $\frac{1}{2}$ -in. precision microphone (ACO Pacific) positioned in the center of the testing chamber in the absence of the subject. The transfer functions from the loudspeakers to the ear canal were computed by dividing the spectra measured with the ear canal microphones by the transfer functions of the loudspeakers. The influence of the transfer functions of the ear canal microphones was eliminated along with the lumped "common component" described below.

Measurements from each subject were completed in approximately 8.5 min. The subject was instructed to stand stationary during the measurements, to hold his or her mouth closed with the jaw slack so that the teeth did not touch, and to refrain from swallowing. A handrail was provided to help the subjects remain stationary, but otherwise there was no head rest or other restraint. Subjects were observed with a video monitor. In a few instances in which a subject coughed or moved or reported swallowing, measurements were halted and begun again. With those exceptions, however, movement of subjects appeared not to contribute appreciably to variance in measurements. This was confirmed by the test-retest reliability of results (Sec. IIC).

Each external-ear transfer function was treated as the product of two transfer functions. The first, which contained components that were specific to each sound-source direction, was referred to as the *directional transfer function* (DTF; Middlebrooks and Green, 1990). The second con-

tained components that were common to all directions. Contributors to the common component included the ear canal resonance and the microphone transfer function. The magnitude spectrum of the common component was estimated for each set of measurements in each ear by computing the root-mean-squared (rms) of the sound pressure at each frequency averaged across all 400 sound-source locations. The minimum-phase spectrum of the common component was given by the Hilbert transform of the logarithm of the magnitude spectrum (Oppenheim and Schaffer, 1975; Middlebrooks and Green, 1990). The complex common component was formed by the common magnitude spectrum and the minimum phase spectrum. The DTF for each sound-source location was calculated by dividing the complex transfer function measured at each location by the complex common component.

D. Data analysis

The linear spacing of frequency components in measured DTFs sampled higher octaves more densely than lower octaves. To avoid over-representing higher frequencies and for ease in scaling DTFs in frequency, DTFs were processed with a bank of band-pass filters to sample frequency components at equal intervals on a base-2 logarithmic (i.e., octave) scale. The filter bank consisted of 85 triangular filters. The 3-dB bandwidth of the filters was 0.057 octave, filter slopes were 105 dB per octave, and center frequencies were spaced in equal intervals of 0.0286 octave from 3 to 16 kHz. The interval of 0.0286 octave was chosen to give intervals of 2% in frequency, which provided suitable resolution in preliminary tests of frequency scaling.

We scaled DTFs in frequency by applying constant offsets in logarithmic frequency. An upward scaling of 2%, for instance, was accomplished by offsetting each component upward by 0.0286 octave. Scale-factor axes in the illustrations are drawn on logarithmic scales to indicate the equal logarithmic steps in which scale factors were manipulated. Also, computation of correlation coefficients and of root-mean-squared (rms) differences between sets of scale factors utilized the base-2 logarithms of scale factors. The advantage of a logarithmic scale is that the logarithm of a factor and of its inverse are symmetrical about the origin. For instance, linear factors of 2 and $\frac{1}{2}$ correspond to +1 and -1 octave. For convenience to the reader, the results of computations conducted using logarithms are expressed as linear factors.

The set of tested sound-source locations sampled some regions of space somewhat more densely than others. To provide uniform sampling and to permit rotation of the coordinate sphere (needed for Sec. II B), a spatial interpolation procedure was employed. In that procedure, 393 sectors were defined on the coordinate sphere. Each sector had a solid angle of 0.0305 steradians. On the horizontal plane, the sectors were 10° wide in azimuth and 10° tall in elevation. At higher and lower elevations, the dimensions of sectors were adjusted to maintain a constant solid angle. The centers of sectors ranged through 360° of azimuth and from -60 to 90° in elevation. A DTF at the center of each sector was computed by forming a weighted average of the filter-bank-processed DTFs measured for all points within 16° of the

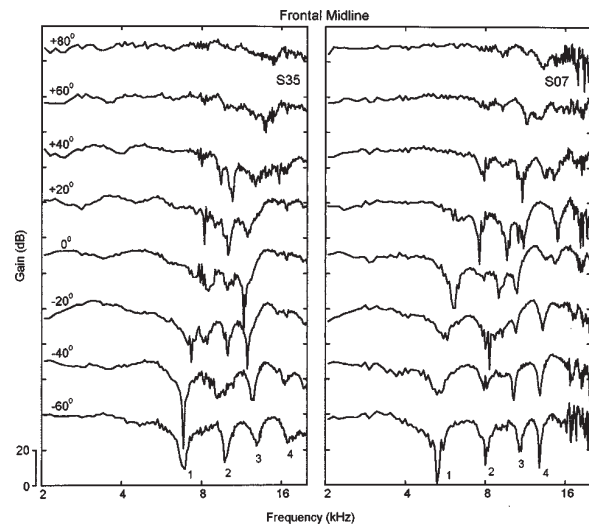


FIG. 1. Directional transfer functions (DTFs) for directions on the frontal midline at the elevations indicated relative to the horizontal plane. Left and right panels show DTFs measured from the right ears of subjects S35 and S07, respectively. Numbers 1 through 4 indicate spectral notches discussed in the text.

center of the sector. The spatial interpolation was described in more detail and its validity tested in a previous report (Middlebrooks, 1992). Except when stated otherwise, all computations that involved only one ear were conducted with DTFs that were measured from subjects' right ears and that were processed with the filter bank and spatially interpolated as described above.

II. RESULTS

A. Scaling of DTFs among subjects

Directional transfer functions (DTFs) were recorded from the ears of 45 subjects. Examples of DTFs from the right ears of two subjects are shown in Figs. 1, 2, and 3. Those DTFs were measured for sound sources at eight elevations along the frontal (Fig. 1) and rear (Fig. 2) midlines and at eight frontal azimuths in the horizontal plane (Fig. 3). The left and right columns of plots represent measurements from subjects S35 and S07, respectively. The DTFs of these subjects, as well as those of all other subjects in the sample, showed patterns of spectral features (e.g., spectral peaks, notches, and slopes) that varied systematically as a function of sound-source elevation and azimuth. As sound sources were shifted in elevation from low to high, the center frequencies of spectral features tended to shift from low to high frequencies (Figs. 1 and 2). Changes in sound-source azimuth resulted in upward shifts in the frequencies of spectral features as source azimuth shifted from the side opposite the recorded ear, past the frontal midline, to the side of the recorded ear (Fig. 3).

The DTFs shown in both columns of Figs. 1 through 3 contain similar patterns of spectral features, but those spectral features tend to lie at higher frequencies for subject S35 (left column) than for subject S07 (right column). For descriptive purposes in Fig. 1, numbers identify four notches in the DTFs for elevation -60°. In the left column, notches 1

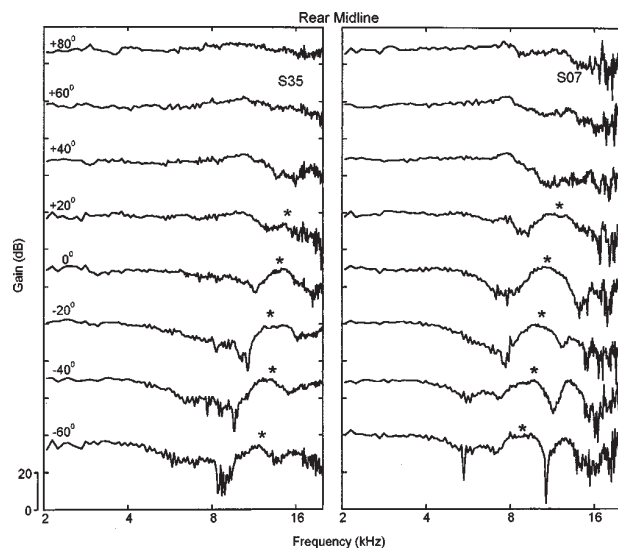


FIG. 2. DTFs for directions on the rear midline at the elevations indicated relative to the horizontal plane. The asterisks indicate spectral peaks discussed in the text.

through 4 lie between 6.9 and 16.8 kHz, whereas the corresponding notches lie about 25% lower in the right column, between 5.2 and 12.8 kHz. In the data for the rear midline (Fig. 2) asterisks indicate spectral peaks that can be seen in the DTFs for elevations -60 through $+20^\circ$. Again, those peaks lie at higher frequencies for subject S35 than for subject S07. In Fig. 3, asterisks indicate the lowest-frequency notches. As the sound-source location was shifted from azimuths 0° to right 80° , the center frequencies of those notches shifted from 8.4 to 10.5 kHz in S35's DTFs and from 6.1 to 8.0 kHz in the S07's DTFs.

Quantitative comparisons among DTFs from various subjects used DTFs that had been filtered with a bank of 85

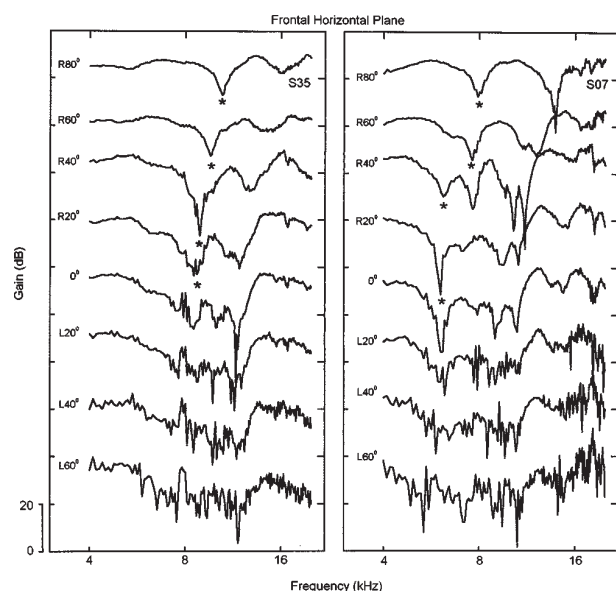


FIG. 3. DTFs for directions on the frontal horizontal plane at the azimuths indicated relative to the midline. L and R indicate left and right. The asterisks indicate spectral notches discussed in the text.

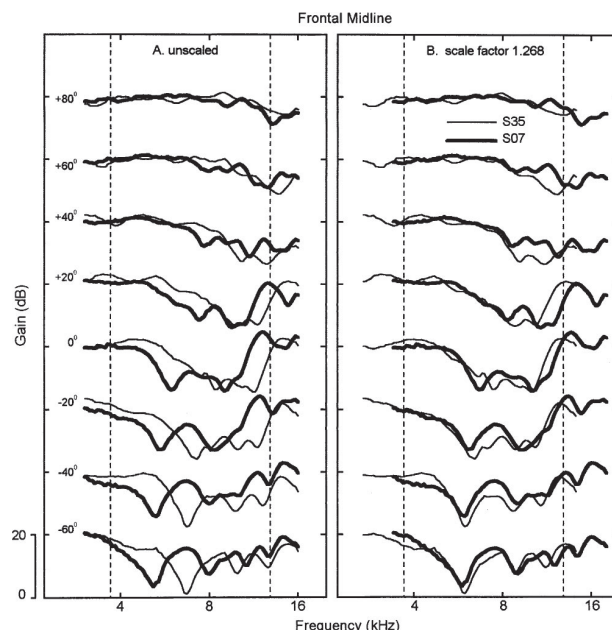


FIG. 4. DTFs for directions on the frontal midline at elevations indicated relative to the horizontal plane. These DTFs were processed with a bank of narrow-band filters, as described in Sec. 1D. Thin and thick lines represent measurements from subjects S35 and S07, respectively. In the left panel, the DTFs are shown unscaled, as they were measured, whereas in the right panel subject S07's DTFs are scaled frequency by 1.126 and subject S35's DTFs are scaled by $1/1.126$, a combined relative scaling of 1.28. The vertical grid lines indicate the frequency band over which quantitative comparisons were made.

bandpass filters (see Sec. I). Figure 4(A) shows DTFs from subjects S7 and S35 filtered in that way. One can see several instances in which a peak in the DTF from one subject corresponds in frequency to a notch in the DTF from the other subject. In Fig. 4(B), DTFs from subject S07 are scaled upward by a factor of 1.126, and those from subject S35 are scaled downward by 1.126. The combined inter-subject scaling by $1.268(1.126^2)$ resulted in a conspicuously better alignment of spectral features between the two subjects. Note that the DTFs that are represented in Fig. 4 contained frequency components in equal logarithmic steps. Frequency scaling upward or downward by a factor of 1.126 was accomplished by shifting components upward or downward by six 0.0286-octave steps.

Inter-subject differences between DTFs were quantified by a metric that will be referred to as the *inter-subject spectral difference*. That metric was computed across the 64 filter-bank components from 3.7 to 12.9 kHz. That frequency band is indicated in Fig. 4 by vertical dashed lines. The dB amplitudes of the 64 components of DTFs for each direction were subtracted, component by component, to form a difference spectrum. Then the variance across the 64 components of the difference spectrum was computed. Finally, the variance was averaged across 393 tested directions for each pair of subjects. The variance computation that was used to represent inter-subject spectral differences captured inter-subject differences that varied with frequency and eliminated differences, such as differences in overall gain, that were constant across frequency. For that reason, the variance com-

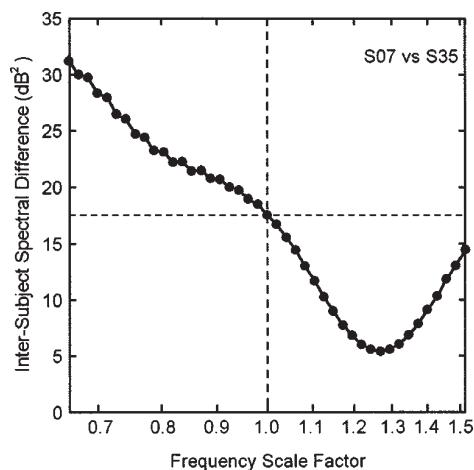


FIG. 5. Inter-subject spectral difference in sets of DTFs from subjects S07 and S35 as a function of relative frequency scale factor.

putation was preferred over a measure of the sum of the squared differences, or the square root thereof (the “Euclidean distance”), which might have emphasized differences in overall gain at the expense of differences in spectral shape.

Figure 5 shows the inter-subject spectral difference between subjects S07 and S35 as a function of relative frequency scale factor. DTFs were scaled incrementally by factors of 1.02 (or 1/1.02) by making alternating upward logarithmic frequency shifts of 0.0286 octave in one set of DTFs and downward log frequency shifts in the other set. As a set of DTFs was scaled upward, high-frequency samples passed out of the high end of the analysis band and low-

frequency amplitudes entered the low end of the band, as indicated Fig. 4(B). The plot in Fig. 5 shows a clear minimum spectral difference at a scale factor of 1.268. The inter-subject spectral difference dropped from 17.5 dB² prior to scaling to 5.4 dB² when subject S07’s DTFs were scaled upward by 1.268 relative to those of subject S35. That is a 69.1% reduction in inter-subject spectral difference. The scale factor that produced the minimum spectral difference between any given pair of subjects will be referred to as the *optimal scale factor*.

Although inter-subject spectral differences normally were measured across DTFs for all sound-source directions, the scale factor that minimized spectral difference usually was quite similar when measured for DTFs for individual source directions. Figure 6(A) shows inter-subject spectral differences for one pair of subjects as a function of scale factor. The thick line shows the spectral difference averaged across all sound-source directions (the usual procedure). The optimal scale factor was 1.24. The thin lines show the spectral differences computed for three individual directions. Two of the thin lines show minima that correspond closely to the minimum of the thick line. The third thin line showed a second, slightly deeper, minimum at a scale factor less than one (indicated by the arrow). The histogram in Fig. 6(B) shows the distribution of optimal scale factors for the same pair of subjects computed for each of the source directions. The seven tallest bars in the histogram encompass 77% of the directions, and optimal scale factors computed for those directions fell within a range of 1.17–1.32. The histogram shows a second peak near a scale factor of 0.79. That peak corresponds to measurements that were made at a few direc-

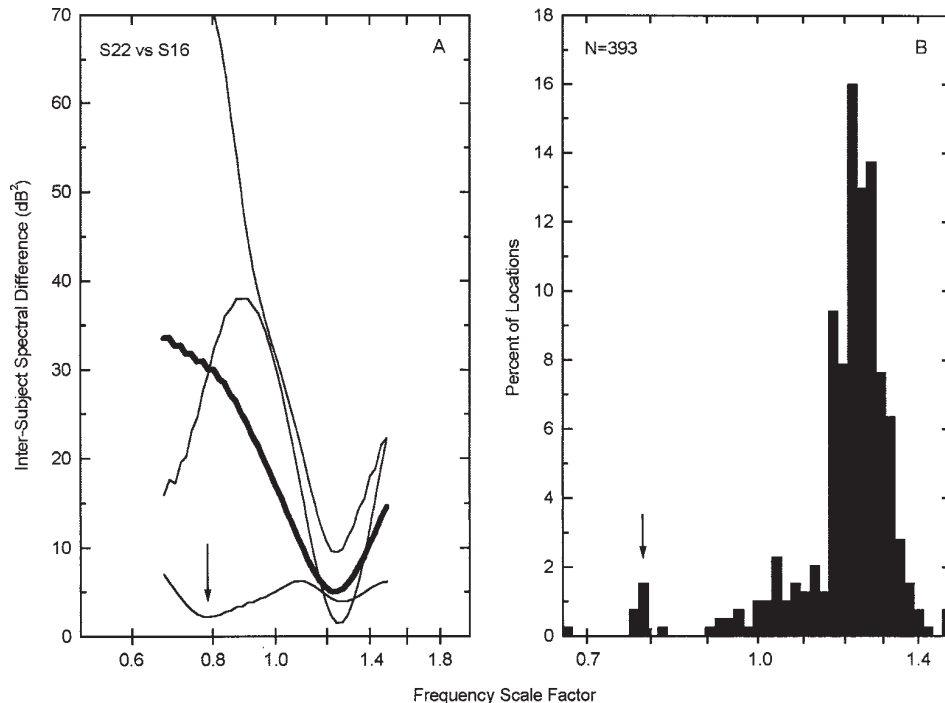


FIG. 6. (A) Inter-subject spectral difference between subjects S16 and S22 as a function of relative frequency scale factor. Values were computed for 3 individual sound-source directions (thin lines) or averaged across 393 sound-source directions (thick line). (B) Distribution of optimal scale factors for subjects S16 and S22, measured for 393 individual directions. The isolated mode around 0.79 represents 10 directions behind the subjects’ heads at which the signal-to-noise ratio was low.

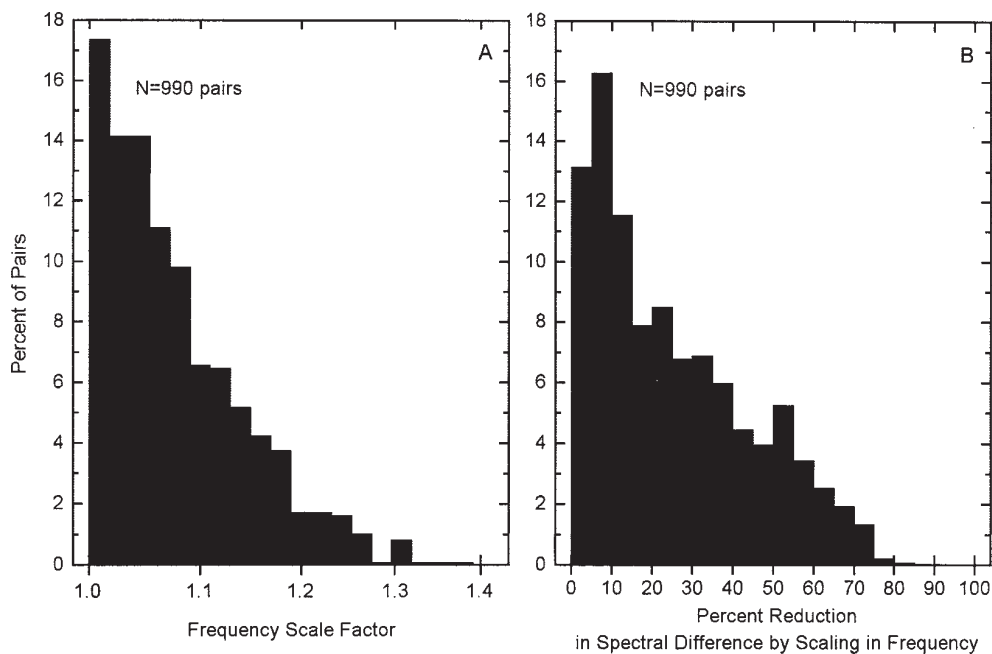


FIG. 7. Distributions of (A) optimal scale factors and (B) percent reduction of spectral difference by frequency scaling. The order of scaling was chosen so that all optimal scale factors were ≥ 1 . Values were computed for all 990 possible pairwise combinations of 45 subjects.

tions behind the head on the side opposite to the side of the recorded ear where signal-to-noise ratios were low. The spectral difference plot for one of those measurements is shown as a thin line in Fig. 6(A).

The distribution of optimal scale factors across all 990 pairs of subjects is shown in Fig. 7(A). The order of scaling was chosen so that optimal scale factors all were ≥ 1 , although for every scale factor X that scaled subject A to B, there was a complementary factor $1/X$ that scaled subject B

to A. The median and maximum optimal scale factors were 1.061 and 1.38. Figure 7(B) shows the distribution of the percent reduction of the inter-subject spectral difference that was accomplished by scaling DTFs in frequency. Frequency scaling reduced spectral differences by 15.5% or more for half of the subject pairs, and the spectral difference was reduced by more than half for 9.5% of subject pairs.

One can think of an optimal scale factor as the overall frequency disparity between subjects' DTFs. Figure 8 plots

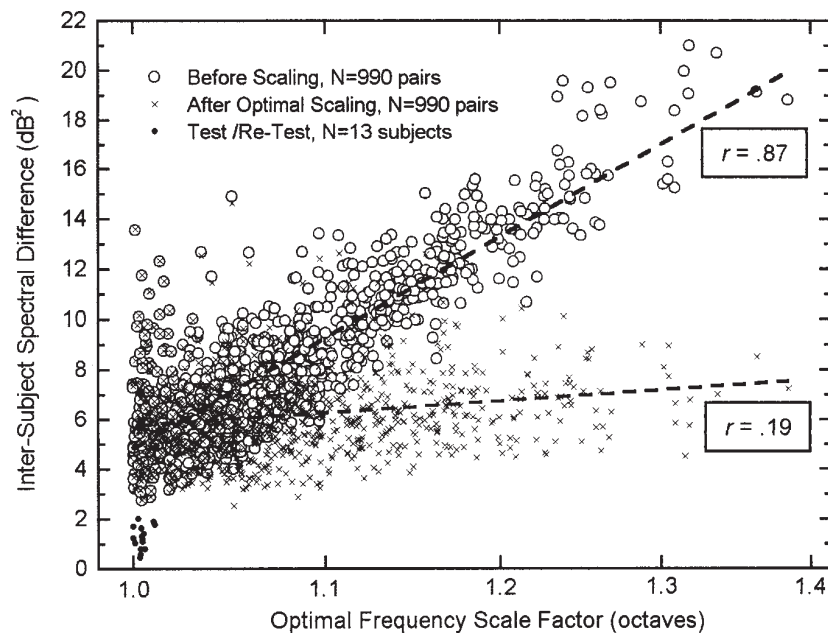


FIG. 8. Inter-subject spectral difference as a function of optimal scale factor. Symbols indicate the inter-subject spectral differences that were measured before (open circles) and after ($\times$'s) scaling DTFs by optimal scale factors. Small filled circles represent spectral differences and optimal scale factors for comparisons of pairs of successive measurements from 13 subjects. Dashed lines are regression lines. $N=990$ pairs.

the inter-subject spectral difference in DTFs, before and after frequency scaling, as a function of the optimal scale factor. Each symbol represents one pair of subjects. Before scaling (open circles), the spectral difference was roughly proportional to the frequency disparity between DTFs. After scaling ($\times$'s), the spectral difference was nearly independent of the magnitude of the optimal scale factor. This indicates that there was a component of the spectral difference that could be eliminated almost entirely by frequency scaling and that remaining components of the spectral difference were largely independent of the overall frequency disparity.

Optimal scale factors computed for subjects' left ears correlated highly with the optimal scale factors computed for their right ears. A correlation coefficient was computed between the base-2 logarithms of optimal frequency scale factors for 990 inter-subject pairs of right ears compared to 990 pairs of left ears. The correlation coefficient was 0.95. Given that observation, one would predict that interaural-level-difference (ILD) spectra would scale in frequency similarly to the DTFs of each ear. To test that prediction, ILD spectra were estimated by filtering right-ear and left-ear DTFs with the bank of band-pass filters, as described above, then subtracting left-ear DTFs in dB from right-ear DTFs in dB. Inter-subject ILD spectral differences for each pair of subjects were computed by subtracting ILD spectra between subjects for each direction, computing the variance of the difference distribution at each direction, then averaging those variances across all directions. The inter-subject ILD spectral differences were measured as a function of inter-subject frequency scale factor. The base-2-logarithms of optimal scale factors computed for ILD spectra correlated closely with those computed for right-ear DTFs ($r=0.96$).

B. Additional components of inter-subject DTF differences

An average of 6.18 dB^2 of inter-subject spectral difference remained after DTFs were scaled in frequency by an optimal factor. Several characteristics of measured DTFs might have accounted for that remaining difference. One possible component was the experimental variability inherent in the recording of DTFs. That variability was estimated by repeating the DTF measurements from 13 of the subjects. In each case, the second set of measurements was made one or more days after the first. The spectral differences and optimal scale factors that were measured for each pair of subjects are shown in Fig. 8 by small filled circles. The average spectral difference was only 1.33 dB^2 , which was approximately 20% of the spectral difference between subjects that remained after frequency scaling.

Another possible source of inter-subject differences in DTFs could have resulted from differences in the positions in which subjects held their heads during the DTF measurements and/or in the orientations of subjects' external ears on their heads. The effect of tilting a subject's head about the interaural axis was simulated *post hoc* by rotating the spatial coordinate system that was used to interpolate among DTFs at various locations (see Sec. ID). A relative tilt was introduced between a pair of subjects by rotating one subject's DTFs upward while rotating the other subject's DTFs a sym-

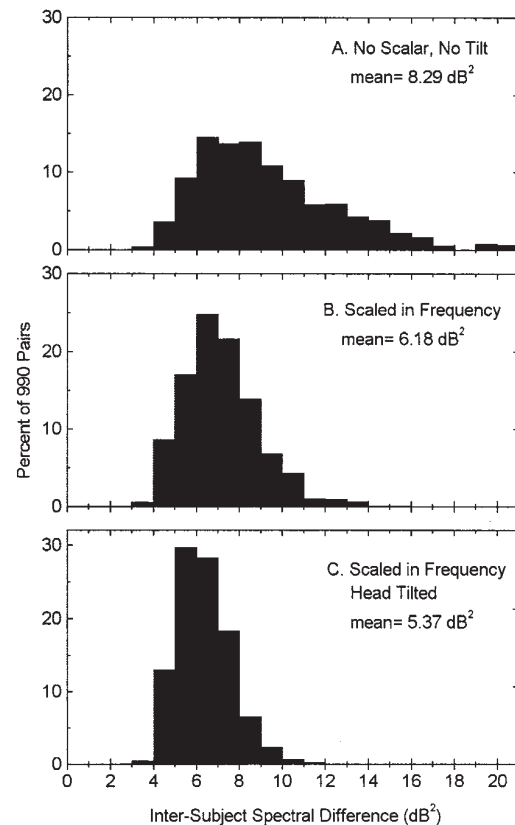


FIG. 9. Distributions of inter-subject spectral differences in DTFs. Distributions of inter-subject spectral differences (A) prior to any manipulation, (B) after scaling in frequency, and (C) after scaling in frequency and introducing relative coordinate-system tilts. $N=990$ subject pairs.

metrical amount downward. We computed for each pair of subjects the relative tilt that minimized spectral differences in DTFs. Across all 990 pairs of subjects, the unsigned magnitude of the optimal tilt averaged 9.0° , and the head-tilting procedure reduced spectral differences by an average of 13.7%.

Figure 9 shows the distribution of inter-subject spectral differences before manipulating DTFs [Fig. 9(A)], after scaling DTFs in frequency [Fig. 9(B)], and after scaling in frequency and adjusting the tilt [Fig. 9(C)]. The means of the distributions were reduced significantly after each procedure (paired t -test, $N=990$, $p<0.001$). Neither procedure had very much influence on the smallest spectral differences, but both reduced the largest differences. For example, frequency scaling brought the 95th percentile of inter-subject spectral difference from 14.3 to 9.3 dB^2 , and subsequent adjustment of relative tilts brought the 95th percentile to 7.6 dB^2 .

Visual inspection of DTFs from various subjects gives some insight into the characteristics of the inter-subject differences that remain after frequency scaling and correction for head and ear orientations. Figure 10 shows examples of DTFs from eight subjects for sound-source locations in the frontal midline at elevations of -60° (left column), 0° (middle column), and $+60^\circ$ (right column). In all cases, the coordinate systems were rotated (i.e., the heads were "tilted") to minimize spectral differences. The DTFs shown

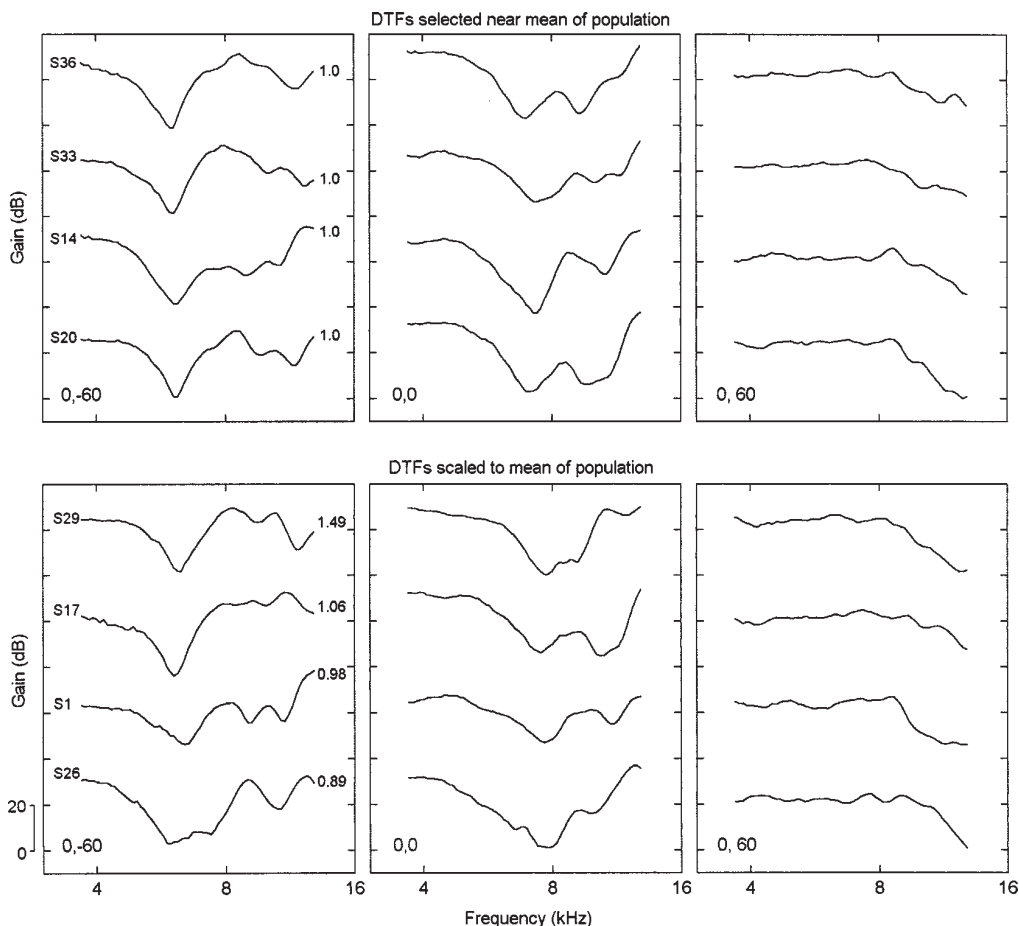


FIG. 10. DTFs from eight subjects. The top row of panels shows DTFs selected from four subjects among whom the optimal scale factors were near one. Those DTFs are illustrated unscaled. The bottom row of panels shows DTFs selected from four subjects that showed a broad range of optimal scale factors. Those DTFs have been scaled to minimize spectral difference with an average subject. The scale factors are shown next to the plots in the lower left panel. The left, middle, and right columns of plots represent DTFs from elevations of -60° , 0° , and 60° relative to the frontal horizontal plane.

in the upper row of panels were selected because the optimal scale factors among them were near unity. They are shown without any scaling. The DTFs shown in the lower row of panels represent a large range of relative frequency scale factors, and those DTFs are scaled in frequency to minimize the spectral difference among them. The number near each DTF indicates the scale factor that was applied. Although one can see inter-subject differences among the DTFs for each location, it appears that the most prominent spectral features are similar across subjects. In particular, the lowest-frequency spectral notch tends to be well aligned across all subjects. Because the lowest-frequency notch also tended to be the largest spectral feature, one assumes that it tended to dominate the measurements of inter-subject spectral difference and the determination of the optimal scale factor. Higher-frequency spectral features are somewhat less well aligned, and differences among them contribute to residual inter-subject spectral differences. One can see differences in the magnitudes of spectral peaks and notches, which were not influenced directly by any of the manipulations that were employed. Also, there are instances of spectral features that are unique to particular DTFs.

C. Relation of optimal scale factors to interaural delays

The difference in the travel time from a sound source to the two ears introduces an interaural time difference (ITD) that provides an important cue for sound localization. ITDs are produced by some of the same physical structures (i.e., head and external ears) that produce the amplitude component of DTFs. For that reason, one would expect inter-subject differences in ITDs to parallel inter-subject differences in DTFs. ITDs were computed from “raw” DTFs that had not been processed with a band-pass filter bank and had not undergone spatial interpolation. For each sound-source direction, a cross spectrum was computed from the complex DTFs from the right and left ears, then an inverse Fourier transform was used to generate the interaural cross-correlogram. The ITD was given by the delay of the peak of the correlogram.

The ITDs for one subject are plotted in Fig. 11(A) as a function of the angle of the sound source relative to the median plane (the *lateral angle*). The ITDs approximate the function that describes ITDs in high-frequency sounds around a spherical head (Woodworth, 1938). The function is

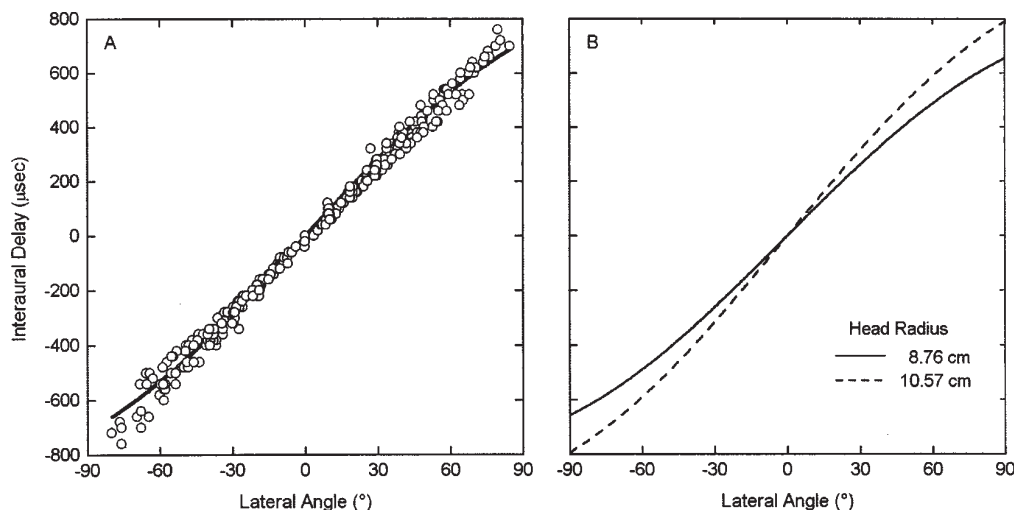


FIG. 11. (A) ITDs from subject S10 are plotted as a function of the angle of the sound source relative to the median plane. Measurements are shown for 400 sound-source directions. The solid line is a fit to the spherical-head ITD expression given in the text. (B) The spherical-head ITD expression is plotted here for two head radii that represent the smallest and largest radii in our sample of subjects.

$ITD = r/c[\theta + \sin(\theta)]$, where r is the radius of a spherical head, c is the speed of sound, and θ is the lateral angle in radians. The ITDs for each subject were summarized by fitting the data to this spherical-head ITD and adjusting r in the function to minimize the squared error. The function was then evaluated at a lateral angle of 90° . The result is an estimate of the maximum ITD for each subject. Maximum ITDs calculated in the 33 subjects from whom physical dimensions were available (see Sec. IID) ranged from 657 to 792 μs ($709 \pm 32 \mu s$). Figure 11(B) plots the fitted ITD functions for the subjects who showed the shortest and longest ITDs.

Inter-subject ratios of maximum ITDs were compared with the scale factors that minimized inter-subject differences in DTFs. For the purpose of computing correlation coefficients, inter-subject optimal scale factors and inter-subject ratios of maximum ITDs were expressed as logarithms so that logarithms of ratios and their inverses would be symmetrical on a linear scale. Log ratios of maximum ITD and log optimal scale factors were formed over all 1056 pairwise combinations of 33 subjects, including both subject A/subject B and subject B/subject A. Inter-subject log maximum ITD ratios correlated highly with inter-subject DTF log optimal scale factors ($r = 0.71$).

D. Relation of optimal scale factors and ITDs to physical dimensions

The spectral details of a DTF result from the physical interaction of an incident sound wave with the head and external ears. For that reason, one would expect the frequencies of spectral features to depend on the sizes of the head and ears. A set of physical dimensions was measured from 33 of the subjects; the other 12 subjects were not available for measurements. A micrometer was used to measure distances among the points on the external ear shown in Fig. 12. The measurements included inter-tragal notch to the rim of the helix (*pinna-cavity height*, I to A), *notch to anti-helix* (I

to B), *notch to crus* (I to C), *concha depth* (F to G), *tragus to antitragus* (G to H), and *pinna-cavity breadth* (G to E). A caliper was used to measure the *head width* from side to side at the point in front of the tragus that is defined by the condyle of the mandible (J). That point can be identified by palpating in front of the tragus while the subject opens and closes his or her jaw. Finally, as a global measure of each subject's physical size, each subject's *body height* was measured, from bare feet to the top of the head. Table I gives summary statistics of those measurements.

Inter-subject ratios of physical dimensions were compared with inter-subject optimal scale factors. Again, ratios were expressed as logarithms for the purpose of computation of correlation coefficients. Correlations of ratios of physical

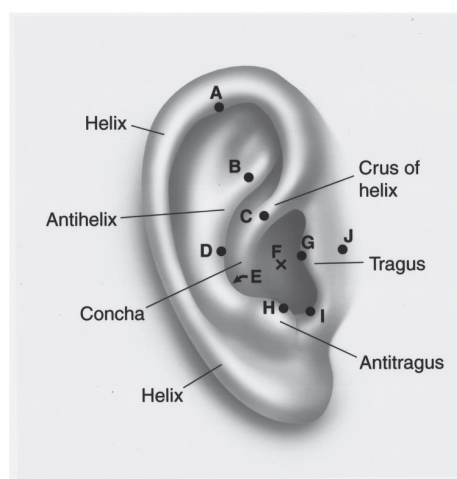


FIG. 12. Drawing of an external ear. Lettered dots indicate the points from which physical measurements were taken. Points A, B, and C are, respectively, the points on the rim of the helix, on the ridge of the anti-helix, and on the ridge of the crus that maximized the distance to the bottom of the inter-tragal notch (point I). Point E is the inside surface of the back of the concha, and point F is directly medial to the tip of the tragus (point G), superficial to the ear canal entrance.

TABLE I. Physical measurements (33 subjects).

	Minimum	Mean	Maximum	Std. Deviation
Pinna-cavity height (mm)	36.1	43.9	56.1	4.8
Notch to anti-helix (mm)	21.6	26.6	32.5	2.8
Notch to crus (mm)	11.5	16.4	21.4	2.4
Concha depth (mm)	6.5	10.4	14.7	2.0
Tragus to anti-tragus (mm)	3.0	6.8	12.9	2.1
Pinna-cavity breadth (mm)	8.2	11.9	16.7	1.9
Head width (mm)	122	134	153	8
Body height (mm)	1520	1690	1930	102
Maximum interaural delay (μ s)	657	709	792	32
Optimal scale factor (1056 subject pairs)	0.72		1.38	

dimensions with ratios of optimal scale factors were computed over all 1056 pairs of 33 subjects. Table II shows the correlation coefficients. All of the correlations in the table are significant at least at the $P < 0.01$ level. The physical dimensions that correlated most strongly with the inter-subject optimal scale factor were the pinna-cavity height ($r = 0.72$) and the head width ($r = 0.67$). The correlation of optimal frequency scale factors with all other dimensions that were measured from the pinna was $r \leq 0.41$. Correlations among physical dimension were computed directly from the dimensions in 33 subjects (i.e., not the logarithms of inter-subject ratios). The greatest correlation in Table I, $r = 0.85$, was between head width and maximum ITD. That is consistent with our previous report of a small sample of subjects that showed a correlation of $r = 0.91$ between head width and interaural envelope delay (Middlebrooks and Green, 1990). Body height also correlated highly with maximum ITD ($r = 0.67$), but that probably reflects the fact that body height correlated with head width itself ($r = 0.67$).

The set of optimal scale factors between each subject and every other subject could be summarized by the geometric mean of the optimal scale factors. Geometric means of optimal scale factors ranged from 0.833 to 1.143. Female subjects tended to have higher mean scale factors than males (i.e., spectral features lay at higher frequencies), but there was no indication of a gender effect that could not be accounted for by the tendency of females to be smaller than males. There was no conspicuous trend in scale factors detected across our relatively small sample of various racial groups.

Inter-subject optimal scale factors could be estimated with some accuracy from ratios of certain physical dimensions. The pinna-cavity height and the head width both showed strong correlations with optimal scale factors. Those

two measurements showed a correlation of only $r = 0.47$ with each other, so a combination of those two measurements should provide a better estimate than either one alone. A linear regression was performed on inter-subject log ratios of the two terms and the resulting regression equation was used to estimate log optimal scale factors. For the present data set, the regression equation to scale subject A to subject B using pinna-cavity height (pinna) and head width (head) was:

$$\log_2(\text{optimal scale factor}) = 0.340 * \log_2(\text{pinna}_A / \text{pinna}_B) + 0.527 * \log_2(\text{head}_A / \text{head}_B). \quad (1)$$

The sign convention was such that if subject A were larger than subject B, subject A's DTFs would be scaled upward in frequency to align with the DTFs of subject B. The scale factors estimated by this equation correlated highly with acoustically measured optimal scale factors ($r = 0.82$, $N = 1056$ subject pair). The rms difference between log optimal scale factors estimated in this way and logarithms of factors computed from acoustical measurements was 0.0813 octave, which corresponds to a factor of 1.058. For comparison, optimal scale factors in this set of 1056 subject pairs ranged between 0.72 and 1.38. Thus knowledge of the physical dimensions of the head and external ears of a pair of subjects substantially reduced the uncertainty in their optimal scale factor.

III. DISCUSSION

The results confirm, with some reservations, the two working hypotheses of the study. The first hypothesis speculated that DTFs differ systematically among subjects in regard to the position of spectral features along a frequency axis. Scaling of DTFs in frequency more than halved the spectral difference in 9.5% of subject pairs and produced a substantial, though lesser, reduction in spectral difference among the majority of other subject pairs. Some 20% of pairs of subjects showed little or no reduction in inter-subject spectral difference in DTFs, but those pairs involved subjects who showed relatively little spectral difference at the outset. A residue of spectral difference, averaging 6.2 dB² across all subject pairs, remained after scaling by an optimal factor. Possible components of that residual spectral difference included variability in experimental measurements and inter-subject differences in the orientations of subjects' heads and ears. The DTFs from individual subjects also contained direction-specific idiosyncrasies that could not be resolved by a global scale factor. Such idiosyncrasies included the

TABLE II. Correlation coefficients (r).

	Optimal scale factor (1056 pairs)	Maximum ITD (33 subjects)	Body height (33 subjects)	Head width (33 subjects)	Pinna-cavity height (33 subjects)
Maximum ITD	0.71				
Body height	0.59	0.67			
Head width	0.67	0.85	0.67		
Pinna-cavity height	0.72	0.54	0.51	0.47	
Pinna height to crus	0.41	0.43	0.58	0.49	0.48

presence or absence of particular spectral features and inter-subject differences in the magnitudes of such features. One's impression from visual inspection of DTF plots, however, is that frequency scaling eliminated the greatest, perhaps the only, systematic difference among individual subjects' DTFs.

The second working hypothesis speculated that the frequency difference between the DTFs from two subjects could be predicted by the relative physical sizes of the two subjects. Physical measurements of subjects identified several dimensions that correlated significantly with optimal scale factors and with maximum ITDs. Given measurements of pinna-cavity height and head width from a pair of subjects, it was possible to estimate the optimal scale factor between those subjects to within an rms error of 0.0813 octave, a factor of 1.058. That is less than three times the smallest increment in frequency scale factors that was tested. Some of the correlations that were observed among physical dimensions and optimal scale factors are likely to reflect a causal relationship. For instance, mathematical models of ITDs produced by a spherical head incorporate head radius as a parameter (Woodworth, 1938; Kuhn, 1977). Similarly, Rabinowitz and colleagues (1993) modeled the DTFs that would be produced by diffraction on the surface of a sphere. In the far-field condition (like that approximated in the present study), the DTF scaled linearly in frequency by $1/\beta$ when the sphere radius was scaled by β . Several investigators have modeled the directionality of the pinna in terms of a system of parallel delays and summations (e.g., Batteau, 1967; Hebrank and Wright, 1974; Watkins, 1978). As the size of the pinna increases, delays within the pinna increase, and notches and peaks in the delay-and-add spectra move downward in frequency. In contrast to these causally related parameters, the relative body heights of subjects probably have no direct influence on their optimal scale factors. Nevertheless, to the degree that the dimensions of subjects' heads and ears are proportional to subjects' overall physical sizes, ratios of subjects' heights might provide a rough estimate of their optimal scale factors.

A. Relation to previous results

Several previous reports have commented on inter-subject differences in the frequencies of spectral features in DTFs. Shaw (1966, 1982) illustrated DTFs from ten listeners. He noted that the "relative positions of the individual curves [in frequency]... are independent of the angle of incidence" (Shaw, 1966, p. 469). This observation is in keeping with the notion that DTFs are similar among subjects but tend to be offset in frequency. Shaw (1982) also noted inter-subject differences in the number of spectral "minima," the elevation dependence of the frequencies of minima, and the tendency of minima to cluster along the frequency axis. Those are the types of individual differences that were not influenced by any of the manipulations in the present study. Mehrgardt and Mellert (1977) were concerned with preserving spectral detail when averaging DTFs among 20 subjects. They observed that DTFs from various individual subjects for particular directions overlapped more closely when they

were scaled in frequency by as much as 8%. A scale factor of 8% is slightly larger than the median value of optimal scale factors, 6%, in the present study. Mehrgardt and Mellert optimized frequency scale factors on a direction-by-direction basis, whereas the procedure in the present study attempted to find a single scale factor that would minimize inter-subject spectral differences across all source directions.

In a previous study (Middlebrooks *et al.*, 1989), we described the directional sensitivity of amplitude gain of the ear at a several individual frequencies. Directional sensitivity was represented in that report primarily in the form of frequency-specific plots of sound amplitude in the ear canal as a function of two-dimensional sound-source direction. We observed great similarity in directional sensitivity among subjects, but noted that the frequency at which we observed a particular directional pattern tended to vary among subjects. That is, the directional pattern measured at 8 kHz for subject A might correspond more closely to the pattern measured at 8.9 kHz for subject B than to subject B's pattern measured at 8 kHz. We computed, in effect, the optimal scale factor among 5 subjects at 21 reference frequencies across 356 sound-source directions. There was a general tendency, with some variation, for frequency scale factors to be constant across reference frequencies. Moreover, in that study, scale factors tended to rank according to the physical sizes of subjects. In another related study, we analyzed DTFs using principal components analysis (Middlebrooks and Green, 1992). When subjects were ranked according to body height, features of the principal-component basis vectors computed from DTFs of the shorter subjects lay at higher frequencies than those of the taller subjects.

B. Significance of frequency scaling for virtual localization

The DTFs measured from a listener can be used to filter sounds that are presented to the listener through headphones (Wightman and Kistler, 1989b). In the most successful cases, a realistic virtual acoustic environment results, in which acoustic targets appear to occupy well-defined positions in the space surrounding the listener. The main practical motivation of the current study was to develop a means of tailoring one or a few generic sets of DTFs to specific listeners. Scaling DTFs in frequency might accomplish a great part of that goal, and the companion paper reports the results of tests of the perceptual consequences of frequency scaling (Middlebrooks, 1999). Considered here are some of the implications of the magnitudes of optimal scale factors, the significance of head position in measurements of DTFs, and the utility of estimation of optimal scale factors from physical dimensions.

It is instructive to compare the magnitudes of optimal scale factors with perceptual measures of frequency sensitivity. The median optimal scale factor in the present sample was 1.061, and the largest optimal scale factor between any pair of subjects was 1.38. A factor of 1.061 corresponds to 0.086 octave, which is slightly larger than a musical semitone. That value is roughly half the equivalent rectangular bandwidth estimated for an auditory filter around 7 kHz (Moore and Glasberg, 1983). Moore and colleagues (1989)

found that frequency-discrimination thresholds around 8 kHz were of the order of .015 octave for tones or band-pass noise, but that frequency-discrimination thresholds for a 9-dB peak or notch imposed on a broadband base were of the order of 0.08 to 0.10 octave. That means that, if two subjects were to be chosen randomly, there would be a roughly 50% chance that individual spectral features in the DTFs from one subject would fall a detectable frequency increment away from the corresponding features in the other subject's DTFs.

The spectral difference between sets of DTFs from a given pair of subjects often could be reduced by rotating about the interaural axis the spatial coordinate system of one set of DTFs relative to the other. That result suggests either that subjects held their heads at slightly different angles during the measurements or that subjects' external ears are positioned on their heads at somewhat different angles. That variability in head and/or ear position complicates inter-subject comparisons of DTFs. Nevertheless, we are inclined to regard the inter-subject variance in head orientation as a nuisance factor that can be largely ignored for the purposes of virtual localization testing. In a psychophysical localization task there can be considerable variation in nominally upright head position even in a single subject. For instance, when subjects were asked to orient their heads toward a continuous broadband noise source located near the horizontal plane, the head orientations by individual subjects varied with standard deviations of about 3° (Makous and Middlebrooks, 1990; also observed in the companion study). In a situation in which a subject localized using DTFs that were recorded from a head that was tilted downward by 10°, one might expect to see a 10° vertical offset in localization judgements. That offset would be small, however, compared to the front/back confusions that often are observed when a subject listens through another subject's DTFs.

The observation that the optimal scale factor between a pair of subjects can be estimated from physical measurements of their pinna-cavity heights and head widths offers the hope that a generic set of DTFs might be tailored to fit any subject. Estimates of frequency scale factors based on physical measurements [Eq. (1)] differed from scale factors computed from acoustical measurements by an rms error of 5.8%. The magnitude of that error can be appreciated by referring to the plot of inter-subject spectral difference versus frequency scale factor shown in Fig. 5. A scaling up or down by 2 products of 1.06 (scale factors of 0.89 to 1.12) encompassed 95% of errors in estimates (as measured from the error distribution). That range encompassed fewer than six sample points on either side of the minimum value in the figure. In the illustrated example, frequency scaling by any factor in that range reduced the spectral difference by more than 40% or more. It might be the case that a optimal scale factor estimated only from physical dimensions is adequate to produce considerable reduction in inter-subject differences in DTFs. Alternatively, an optimal scale factor estimated from physical dimensions could be used as a starting point for an adaptive psychophysical procedure that would seek an optimal scale factor or that would select from among a library of sets of DTFs. The companion paper explores the perceptual consequences of scaling nonindividualized DTFs

and demonstrates the sensitivity of localization performance to changes in the values of frequency scale factors.

ACKNOWLEDGMENTS

I thank Zekiye Onsan for expert technical assistance. Li Xu helped to record the DTFs. Shigeto Furukawa, Ewan Macpherson, Brian Mickey, and Li Xu provided stimulating discussion and commented on an early draft of the manuscript. I thank Professors Edgar Shaw, Frederic Wightman, and a third anonymous reviewer for their constructive criticisms. This research was supported by ONR Grant No. N00014-94-1-0481 and NIH Grant No. DC000420.

- Batteau, D. W. (1967). "The role of the pinna in human localization: Theoretical and physiological consequences," *Proc. R. Soc. London, Ser. B* **168**, 158–180.
- Blauert, J. (1969–1970). "Sound localization in the median plane," *Acustica* **22**, 205–213.
- Bronkhorst, A. W. (1995). "Localization of real and virtual sound sources," *J. Acoust. Soc. Am.* **98**, 2542–2553.
- Butler, R. A. (1987). "An analysis of the monaural displacement of sound in space," *Percept. Psychophys.* **41**, 1–7.
- 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.
- Hartmann, W. M., and Wittenberg, A. (1996). "On the externalization of sound images," *J. Acoust. Soc. Am.* **99**, 3678–3688.
- 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.
- Kuhn, G. F. (1977). "Model for the interaural time differences in the azimuthal plane," *J. Acoust. Soc. Am.* **62**, 157–167.
- Makous, J. C., and Middlebrooks, J. C. (1990). "Two-dimensional sound localization by human listeners," *J. Acoust. Soc. Am.* **87**, 2188–2200.
- Mehrgardt, S., and Mellert, V. (1977). "Transformation characteristics of the external human ear," *J. Acoust. Soc. Am.* **61**, 1567–1576.
- Middlebrooks, J. C. (1992). "Narrow-band sound localization related to external ear acoustics," *J. Acoust. Soc. Am.* **92**, 2607–2624.
- Middlebrooks, J. C. (1999). "Virtual localization improved by scaling non-individualized external-ear transfer functions in frequency," *J. Acoust. Soc. Am.* **106**, 1493–1510.
- Middlebrooks, J. C., and Green, D. M. (1990). "Directional dependence of interaural envelope delays," *J. Acoust. Soc. Am.* **87**, 2149–2162.
- Middlebrooks, J. C., and Green, D. M. (1991). "Sound localization by human listeners," *Annu. Rev. Psychol.* **42**, 135–159.
- Middlebrooks, J. C., and Green, D. M. (1992). "Observations on a principal components analysis of head-related transfer functions," *J. Acoust. Soc. Am.* **92**, 597–599.
- Middlebrooks, J. C., Makous, J. C., and Green, D. M. (1989). "Directional sensitivity of sound-pressure levels in the human ear canal," *J. Acoust. Soc. Am.* **86**, 89–108.
- Møller, H., Sørensen, M. F., Jensen, C. B., and Hammershoi, D. (1996). "Binaural technique: Do we need individual recordings?," *J. Aud. Eng. Soc.* **44**, 451–469.
- Moore, B. C. J., and Glasberg, B. R. (1983). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," *J. Acoust. Soc. Am.* **74**, 750–753.
- Moore, B. C. J., Oldfield, S. R., and Dooley, G. J. (1989). "Detection and discrimination of spectral peaks and notches at 1 and 8 kHz," *J. Acoust. Soc. Am.* **85**, 820–836.
- Morimoto, M., and Ando, Y. (1980). "On the simulation of sound localization," *J. Acoust. Soc. Jpn. (E)* **1**, 167–174.
- Oppenheim, A. V., and Schaffer, R. W. (1975). *Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ).

- Rabinowitz, W. R. *et al.* (1993). "Sound localization cues for a magnified head: Implications from sound diffraction about a rigid sphere," *Presence* **2**, 125–129.
- Shaw, E. A. G. (1966). "Earcanal pressure generated by a free sound field," *J. Acoust. Soc. Am.* **39**, 465–470.
- Shaw, E. A. G. (1974). "Transformation of sound pressure level from the free field to the eardrum in the horizontal plane," *J. Acoust. Soc. Am.* **56**, 1848–1861.
- Shaw, E. A. G. (1982). "External ear response and sound localization," in *Localization of Sound: Theory and Applications* (Amphora, Groton, CT), pp. 30–41.
- 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.
- Woodworth, R. S. (1938). *Experimental Psychology* (Holt, Rinehart, and Winston, New York), pp. 349–361.
- 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.