

# Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements

Christopher A. Shera<sup>\*†‡</sup>, John J. Guinan, Jr.<sup>\*†</sup>, and Andrew J. Oxenham<sup>§</sup>

<sup>\*</sup>Eaton-Peabody Laboratory of Auditory Physiology, Massachusetts Eye and Ear Infirmary, 243 Charles Street, Boston, MA 02114; <sup>†</sup>Department of Otolaryngology, Harvard Medical School, Boston, MA 02115; and <sup>§</sup>Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139

Communicated by George Zweig, Los Alamos National Laboratory, Los Alamos, NM, December 17, 2001 (received for review July 13, 2001)

**We develop an objective, noninvasive method for determining the frequency selectivity of cochlear tuning at low and moderate sound levels. Applicable in humans at frequencies of 1 kHz and above, the method is based on the measurement of stimulus-frequency otoacoustic emissions and, unlike previous noninvasive physiological methods, does not depend on the frequency selectivity of masking or suppression. The otoacoustic measurements indicate that at low sound levels human cochlear tuning is more than twice as sharp as implied by standard behavioral studies and has a different dependence on frequency. New behavioral measurements designed to minimize the influence of nonlinear effects such as suppression agree with the emission-based values. A comparison of cochlear tuning in cat, guinea pig, and human indicates that, contrary to common belief, tuning in the human cochlea is considerably sharper than that found in the other mammals. The sharper tuning may facilitate human speech communication.**

The mammalian cochlea acts as an acoustic prism, mechanically separating the frequency components of sound so that they stimulate different populations of sensory cells. As a consequence of this frequency separation, or filtering, each sensory cell within the cochlea responds preferentially to sound energy within a limited frequency range. In its role as a frequency analyzer, the cochlea has been likened to a bank of overlapping bandpass filters, often referred to as “cochlear filters.” The frequency tuning of these filters plays a critical role in our ability to distinguish and perceptually segregate different sounds. For instance, hearing loss is often accompanied by a degradation in cochlear tuning, or a broadening of the cochlear filters. Although quiet sounds can be restored to audibility with appropriate hearing-aid amplification, the loss of cochlear tuning leads to pronounced, and as yet largely uncorrectable, deficits in the ability of hearing-impaired listeners to extract meaningful sounds from background noise (1).

The bandwidths of cochlear filters have been measured directly in anesthetized, non-human mammals by recording from the auditory-nerve fibers that contact the sensory cells (2). Filter bandwidths in humans, however, must be determined indirectly from noninvasive measurements. Traditionally, such studies have relied on psychophysical (i.e., behavioral) measures of filter bandwidth based on the phenomenon of masking; that is, the ability of one sound to interfere with, or “mask,” the perception of another. Strong masking is interpreted as indicating that frequency components of the masker fall within the passband of the cochlear filter whose output is used to detect the signal. Interference then occurs because both signal and masker stimulate an overlapping group of sensory cells. Since the pioneering work of Harvey Fletcher (3), filter bandwidths have been obtained by measuring listeners’ thresholds for detection of a pure tone in background noises with particular spectral characteristics. These tone thresholds are then used to infer filter bandwidths, by using the assumptions of the “power spectrum” model of masking (4). Unfortunately, the assumptions underlying this model are of uncertain validity; indeed, some (e.g., that the cochlear filters are independent of sound intensity) are

known to be false. Furthermore, the psychophysical detection tasks used in the threshold measurements depend not only on the characteristics of the peripheral filters, but also to an unknown extent on neural processing in the central nervous system. Although both masking models and psychophysical procedures have been varied and refined over the years, there has been, until now, no convincing physiological means of validating behavioral measures of human cochlear tuning. By developing a noninvasive measure of cochlear tuning based on otoacoustic emissions, we aimed to compare cochlear tuning across species and to test the correspondence between physiological and behavioral measures of auditory frequency selectivity.

## Methods

**Otoacoustic.** We measured stimulus-frequency otoacoustic emission (SFOAE) phase in guinea pigs ( $n = 9$ ) and humans ( $n = 9$ ) by using the acoustic suppression method (5), and in cats ( $n = 7$ ) by using both acoustic and efferent suppression (6). In both methods, the emission is obtained as the complex (or vector) difference between the ear-canal pressure at the probe frequency measured first with the probe tone alone and then with the addition of a “suppressor.” The suppressor was either (i) a tone at a nearby frequency or (ii) olivocochlear efferent stimulation by electrical shocks. Both acoustic and efferent suppression are assumed to reduce the SFOAE at the probe frequency substantially. In all cases, the probe sound-pressure level (SPL) was approximately 40 dB. Details of the animal preparation and measurement methods can be found elsewhere (5–7). Treatment of animal and human subjects accorded with protocols established at the Massachusetts Eye and Ear Infirmary.

We calculated SFOAE group delays,  $\tau_{\text{SFOAE}}$ —defined as the negative of the slope of the emission-phase (in cycles) versus frequency function—from unwrapped phase responses and expressed them in dimensionless form as the equivalent number,  $N_{\text{SFOAE}}$ , of stimulus periods. To augment our human data at frequencies above 10 kHz, we included SFOAE group delays from a study by Dreisbach *et al.* (8) in the analyzed data set.

**Psychophysical.** We tested eight young normal-hearing subjects, three of whom were also subjects in the SFOAE measurements, in a double-walled sound-attenuating chamber, using an Etymotic ER-2 insert earphone. Treatment of subjects accorded with protocols established at the Massachusetts Institute of Technology. The subjects’ task was to detect a sinusoidal signal, presented at a level 10 dB above its threshold in quiet. The 10-ms signal was gated on and off with 5-ms raised-cosine ramps (no steady-state portion) and was presented 5 ms after the offset of

Abbreviations: BM, basilar membrane; CF, characteristic frequency; ERB, equivalent rectangular bandwidth; SFOAE, stimulus-frequency otoacoustic emission.

<sup>†</sup>To whom reprint requests should be addressed at: Eaton-Peabody Laboratory, 243 Charles Street, Boston, MA 02114. Email: shera@epl.mei.harvard.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

a burst of masking noise (the forward masker). The signal frequency ( $f_s$ ) was 1, 2, 4, 6, or 8 kHz. The 400-ms forward masker (gated with 5-ms ramps) consisted of two spectral bands of Gaussian noise, each  $0.25f_s$  wide. The spectral edges of the noise closest to the signal were placed symmetrically at intervals 0, 0.1, 0.2, 0.3, or  $0.4f_s$  below and above the signal frequency. Two asymmetric conditions were also tested, with the cut-off frequencies set to 0.2 and  $0.4f_s$ , or 0.4 and  $0.2f_s$ . We measured thresholds using a three-alternative forced-choice procedure. The masker level was varied adaptively with a 2-up 1-down tracking procedure, which estimates the 71%-correct point on the psychometric function. The masker level was initially varied in steps of 8 dB. After every two level reversals, the step size was halved until it reached the final step size of 2 dB. The mean of the remaining eight reversals defined the threshold level of the run. We repeated every condition at least three times for each subject. For the mean data, noise spectrum levels at threshold were between 2.6 and 12.2 dB SPL for the no-notch condition and were between 42.1 and 52.4 dB SPL for the widest notch condition ( $0.4f_s$ ). At a given notch width, there was a tendency for the masker level to increase with increasing signal frequency.

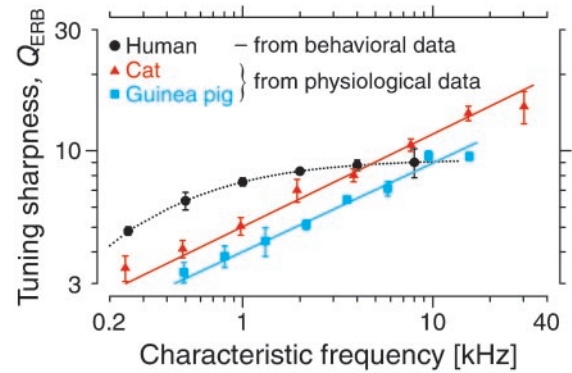
We used the individual and mean data to derive cochlear filter magnitude responses using the roexp(pwt) model (9, 10). In this analysis, a certain generic, rounded-exponential filter shape is varied by using four free parameters to best fit the data according to a least-squares error criterion. To make the results as comparable as possible to the neural-tuning data from the animal studies, we applied no correction for the frequency dependence of middle-ear transmission. The inclusion of a middle-ear correction (11) had only a slight effect on the estimated bandwidths, and no effect on our conclusions. The model provides a good description of the data (rms error = 0.76 dB for the mean data). The results presented here are derived from the mean bandwidths computed from fits to the data in individual subjects.

### Comparing Cochlear Tuning Across Species

In the psychophysical literature the bandwidth of tuning is conventionally characterized by a parameter-free quantity known as the equivalent rectangular bandwidth (ERB), also often called the “critical bandwidth” (3). For any filter, the corresponding ERB is simply the bandwidth of the rectangular filter with the same peak response that passes the same total power when driven by white noise. Here, we represent cochlear frequency selectivity by using a related dimensionless measure of tuning, the  $Q_{\text{ERB}}$ , defined as  $Q_{\text{ERB}}(\text{CF}) \equiv \text{CF}/\text{ERB}(\text{CF})$ , where the characteristic frequency, CF, is the center frequency of the filter (e.g., the frequency for which the auditory neuron is most sensitive).  $Q_{\text{ERB}}$  is a measure of the “sharpness” (i.e., frequency selectivity) of tuning: the smaller the bandwidth, the larger the  $Q_{\text{ERB}}$ .<sup>1</sup>

Fig. 1 compares the human  $Q_{\text{ERB}}$  derived from behavioral measurements (15) with the  $Q_{\text{ERB}}$  obtained from physiological measurements (12, 13) in two laboratory animals widely studied as models of mammalian hearing (cat and guinea pig). In its variation with characteristic frequency, the human behavioral  $Q_{\text{ERB}}$  in Fig. 1 differs qualitatively from the two physiological measures: Whereas the cat and guinea-pig  $Q_{\text{ERB}}$  values generally increase with CF throughout the measured range, the human  $Q_{\text{ERB}}$  increases only at low frequencies, remaining essentially constant in the basal, high-frequency half of the cochlea. Although rough agreement between physiological and behavioral measures of tuning has been reported in cat and guinea pig (17, 18), independent physiological measures that might corroborate,

<sup>1</sup>In the physiological literature, the sharpness of tuning is often measured using the  $Q_{10}$ , defined as  $\text{CF}/\text{BW}_{10}$ , where  $\text{BW}_{10}$  is the bandwidth 10 dB below the peak. Although we adopt the ERB-based measure to facilitate comparisons with behavioral measurements, we obtain similar conclusions when using  $Q_{10}$ .



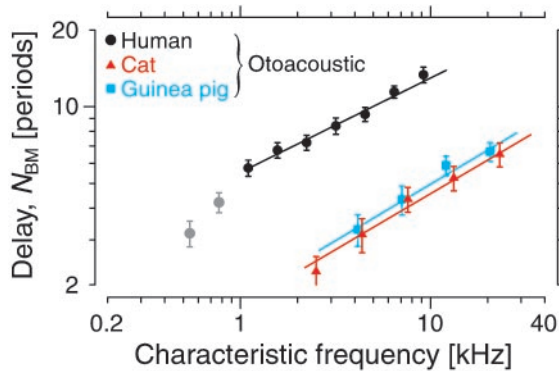
**Fig. 1.** Sharpness of cochlear tuning in three species derived from previous measurements.  $Q_{\text{ERB}}$  is the ratio  $\text{CF}/\text{ERB}(\text{CF})$ , where CF is the characteristic (or center) frequency of the filter, and ERB is the equivalent rectangular bandwidth. The cat and guinea-pig  $Q_{\text{ERB}}$  values were computed from threshold frequency tuning curves of single auditory-nerve fibers (12, 13) by using standard algorithms (14). The human  $Q_{\text{ERB}}$  curve was computed from Glasberg and Moore’s polynomial fit to a variety of standard psychophysical masking data (15, 16). Their formula gives  $Q_{\text{ERB}}(\text{CF}) = (Q_{\infty} \cdot \text{CF}/\text{CF} + \text{CF}_{1/2})$ , where the parameters  $Q_{\infty} = 1,000[\text{Hz}]/(4.37 \cdot 24.7[\text{Hz}]) = 9.26$  and  $\text{CF}_{1/2} = 1,000[\text{Hz}]/4.37 = 230$  Hz. The symbols with error bars represent the means and standard errors computed from the original data in logarithmically spaced frequency bins. The straight lines show power-law fits to the animal data (see Table 1). The flattening of the  $Q_{\text{ERB}}$  at the highest CFs—visible in the guinea pig above 15 kHz and perhaps in the cat above 20 kHz—may well be a measurement artifact. Mechanical responses in the high-frequency region of the cochlea are extremely labile, and cochlear tuning at high CFs can easily be damaged (e.g., as a result of trauma caused by surgically opening the auditory bulla).

or contradict, the human behavioral results are lacking. So, do the differing trends in Fig. 1 reflect genuine species differences, or are the human behavioral measurements—or their standard interpretation as measures of cochlear tuning comparable to neural threshold tuning curves—somehow in error?

### An Otoacoustic Measure of Cochlear Tuning

To address these questions we developed an objective, noninvasive measure of low-level cochlear tuning by exploiting the fact that the ear makes sound while listening to sound. Evoked otoacoustic emissions (OAEs) are sounds, recordable in the ear canal with low-noise microphones, that originate within the cochlea (19). OAEs can be evoked with a variety of stimuli, but the easiest to interpret (although hardest to measure) are those evoked by a pure tone—stimulus-frequency OAEs (SFOAEs), so-called because they occur at the frequency of stimulation. At low and moderate sound levels, these emissions can be explained quantitatively as resulting from the coherent scattering of cochlear traveling waves off small, random perturbations in the mechanical properties of the cochlea (20).

The theory for reflection emissions successfully relates emission characteristics measurable in the ear canal to the mechanical responses of the inner ear (20, 21). We focus here on a quantity that can be related to the frequency selectivity of cochlear tuning: SFOAE group delay ( $\tau_{\text{SFOAE}}$ ), defined as the negative slope of the emission-phase versus frequency function. The theory implies that  $\tau_{\text{SFOAE}}$  is equal to twice the group delay of the basilar membrane (BM) mechanical transfer function ( $\tau_{\text{BM}}$ ), evaluated at the cochlear location with CF equal to the stimulus frequency (20). To facilitate comparison with  $Q_{\text{ERB}}$ , we express BM group delays in the dimensionless form  $N_{\text{BM}}$ , obtained by measuring time in periods of the local CF; thus,  $N_{\text{BM}} \equiv \tau_{\text{BM}} \cdot \text{CF}$ . According to the emission theory,  $N_{\text{BM}} \approx \frac{1}{2} N_{\text{SFOAE}}$ , where  $N_{\text{SFOAE}} \equiv \tau_{\text{SFOAE}} \cdot \text{CF}$ . In this equation, the delay  $\tau_{\text{SFOAE}}(f)$  represents the measured emission group delay at frequency  $f$ , and the stimulus frequency has been identified with CF. Comparisons in laboratory animals between BM group delays



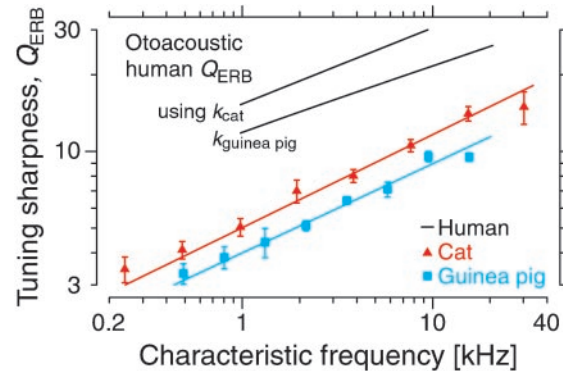
**Fig. 2.** Dimensionless basilar-membrane group delay,  $N_{BM}$ , measured using SFOAEs in three species.  $N_{BM}$  is equal to one half of SFOAE group delay, expressed in periods of the stimulus frequency (identified with CF). The symbols with error bars represent the means and standard errors computed from the original data in logarithmically spaced frequency bins (22, 23). The lines show power-law fits to the original data (see Table 1) at CFs where the otoacoustic measure of  $N_{BM}$  is believed accurate (i.e., roughly the basal 60% of the cochlea). The gray circles show an extension of the human data to lower frequencies.

measured directly and those obtained noninvasively at comparable stimulus levels by using SFOAE phase indicate that the otoacoustic measures are accurate in roughly the basal-most 60% of the cochlea (22, 23); in humans, this region corresponds to  $CF \approx 1$  kHz.

We relate BM group delay to cochlear-filter bandwidth by noting that at low levels BM transfer functions manifest many of the characteristics of minimum-phase-shift filters (24). In particular, their bandwidths and phase slopes are reciprocally related, with smaller bandwidths (i.e., larger  $Q_{ERB}$ ) corresponding to steeper phase slopes (i.e., longer delays and larger  $N_{BM}$ ). Because the tuning of the basilar membrane at low sound levels appears nearly identical to the tuning of corresponding auditory-nerve fibers at frequencies near the CF (25), values of  $N_{BM}$  and  $Q_{ERB}$  are physically related. Mathematically, the two are related by a simple proportionality factor,  $k$ , defined by  $Q_{ERB} \equiv kN_{BM}$ . The function  $k$ , equivalent to the reciprocal of the product of filter bandwidth and group delay, is a dimensionless measure of filter shape. [In the gammatone filter, for example, the value of  $k$  determines the filter order and thus controls the asymmetry (or skewness) of the impulse-response envelope about its maximum. For the gammatone, smaller values of  $k$  correspond to higher orders, and thus to more symmetrical impulse responses.] Although  $k$  depends, in general, on CF, we expect this dependence to be relatively weak: Because bandwidth and group delay are inversely related, the product of the two (i.e.,  $1/k$ ) is likely to vary more slowly in the cochlea than does either factor by itself.<sup>||</sup> We therefore expect  $Q_{ERB}$  and  $N_{BM}$  to vary in almost constant proportion.

Fig. 2 shows our otoacoustic measurements of BM group delay  $N_{BM}$  versus CF in the same three species illustrated in Fig. 1. Comparison of the trends in the two figures demonstrates that our otoacoustic measure of low-level cochlear tuning renders the human behavioral  $Q_{ERB}$  even more anomalous than is indicated by Fig. 1. For example, the animal  $Q_{ERB}$  and  $N_{BM}$  both increase similarly with CF, exhibiting the slowly varying proportionality factor  $k$  expected on theoretical grounds. By contrast, the two human curves depend very differently on CF: Whereas the

<sup>||</sup>To illustrate this with an explicit calculation, consider a minimum-phase band-pass filter of center frequency  $f_c$  (e.g., a gammatone filter) and denote the filter bandwidth by  $\Delta f$ . If the filter phase changes by an amount  $\Delta\phi$  over the bandwidth  $\Delta f$ , then the filter group delay is approximately  $\tau_g \approx -\Delta\phi/\Delta f$ . The value of  $1/k$  is therefore  $N/Q = (\tau_g f_c)/(\tau_g \Delta f) = \Delta f \tau_g \approx -\Delta\phi$ , which is approximately constant in filters of fixed order. Thanks to Jont Allen for suggesting this formulation.



**Fig. 3.** Otoacoustic-emission-based human  $Q_{ERB}$  compared with physiological values in cat and guinea pig. The human  $Q_{ERB}$  was obtained from the human  $N_{BM}$  using Eq. 1 and values of  $k$  calculated from cat and guinea pig. Computations were performed using power-law fits to the original data (see Table 1). The physiological  $Q_{ERB}$  curves for cat and guinea pig are taken from Fig. 1.

behavioral  $Q_{ERB}$  is almost constant at high frequencies,  $N_{BM}$  continues to increase, implying that the human  $k$  function varies strongly with CF. In addition, Fig. 1 indicates that at CFs greater than 1 kHz, the human behavioral  $Q_{ERB}$  is roughly similar in value to the physiological  $Q_{ERB}$  measured in cats and guinea pigs. By contrast, Fig. 2 indicates that BM group delays are roughly a factor of three larger in humans—and thus, if the behavioral  $Q_{ERB}$  is correct, the human  $k$  is roughly a factor of three smaller—than in the two laboratory animals. These differences remain large if the comparisons between species are made at constant relative cochlear location (e.g., at the midpoint of each cochlea), rather than at constant CF.<sup>\*\*</sup> Thus, if the human behavioral measurements of  $Q_{ERB}$  correctly characterize low-level cochlear tuning, we face the additional discrepancy that the human  $k$  function must be very different, both from theoretical expectations and from its counterparts in cat and guinea pig.

The more parsimonious assumption that the function  $k$  is generally similar across species enables us to estimate the human  $Q_{ERB}$  from our measurements of  $N_{BM}$ . Empirical values of  $k$  are obtained by combining the measurements in Figs. 1 and 2. Using  $k$  from cat, for example, yields

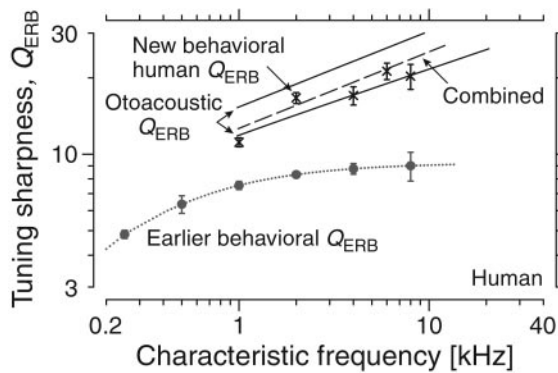
$$Q_{ERB}^{\text{human}} \approx k_{\text{cat}} N_{BM}^{\text{human}}, \quad [1]$$

where the function  $k$  has been assumed to be approximately equal at corresponding cochlear locations in human and cat. Approximate species-invariance of  $k$  is suggested by the general similarity in the shapes of neural tuning curves in those mammalian species for which measurements have been made. Fig. 3 shows the otoacoustic human  $Q_{ERB}$  obtained from Eq. 1 by using power-law fits to the human  $N_{BM}$  (see Fig. 2) and  $k$  functions from both cat and guinea pig.<sup>††</sup> Differences between the two

<sup>\*\*</sup>Comparisons between human and cat at constant relative cochlear location can be simulated by shifting the human curve toward higher CFs by about  $1\frac{1}{2}$  octaves (26). When the two are compared at constant relative location, the human behavioral  $Q_{ERB}$  is considerably less than the animal  $Q_{ERB}$  in the base of the cochlea. The ratio of group delays decreases somewhat but remains substantial: At constant relative location, the human group delays average roughly a factor of 2 (or 2.5) greater than the delays in cat (or guinea pig).

<sup>††</sup>To compensate for the differing frequency ranges of hearing in the three species, the predicted  $Q_{ERB}$  curves shown in Fig. 3 were computed at corresponding cochlear locations, rather than at constant CF, by first transforming the independent variable by using the corresponding frequency-position maps  $CF(\chi)$  (13, 26, 27). In other words, we used the equation  $Q_{ERB}^{\text{human}}[CF_h(\chi)] \approx k_c[CF_c(\chi)]N_{BM}^{\text{human}}[CF_h(\chi)]$ , where the single-letter sub- and superscripts identify the species and  $\chi$  is the fractional distance from the stapes. Predictions made at constant CF are quite similar.





**Fig. 4.** New human behavioral and otoacoustic  $Q_{\text{ERB}}$ . Data points ( $\times$ ) give the mean behavioral and its standard error measured at 1, 2, 4, 6, and 8 kHz by using nonsimultaneous masking and other modified psychophysical procedures. The otoacoustic  $Q_{\text{ERB}}$  values from Fig. 3 are shown for comparison. The dashed line shows a power-law fit to the combined otoacoustic and behavioral values (see Table 1). The earlier behavioral  $Q_{\text{ERB}}$ , reproduced from Fig. 1, appears in gray.

otoacoustic  $Q_{\text{ERB}}$ 's reflect differences between the functions  $k_{\text{cat}}$  and  $k_{\text{pig}}$ , which are similar but not identical. At 1 kHz, the animal  $k$  values differ by a factor of  $1.3 \pm 0.4$ , where the spread represents the 95% confidence interval arising from uncertainty in the estimates of the parameters of the power-law fits. Because the two slopes differ by an amount indistinguishable from zero ( $0.02 \pm 0.13$ ), the frequency-dependence of the cat and guinea-pig  $k$  functions is essentially equivalent. Unless the human  $k$  differs substantially from the range set by these animal values, our otoacoustic measures of  $Q_{\text{ERB}}$  imply that the bandwidths of human cochlear filters differ markedly—in both overall magnitude and dependence on CF—from standard behavioral values.

### New Behavioral Measurements of Cochlear Tuning

Our otoacoustic measure of the human  $Q_{\text{ERB}}$  suggests that previous behavioral measures of frequency selectivity greatly overestimate the bandwidths of peripheral cochlear filters, at least in the base of the cochlea and at the relatively low stimulus levels used here. Because the apparent discrepancy increases at higher frequencies, where nonlinearities in cochlear mechanics appear more significant (28), it is natural to suspect that suppression by energy in the masker, compression at the signal frequency [or “self-suppression” (29)], and/or other nonlinear effects may be at least partly responsible. Because the effective tuning of masker and signal together is broader than the tuning to the signal alone, it has long been known that nonlinear effects such as suppression can result in overestimates of filter bandwidth (30–33). However, no systematic measurements of the frequency dependence of  $Q_{\text{ERB}}$  made using procedures modified to correct these shortcomings have been reported.

To obtain the most accurate behavioral measures of cochlear tuning that psychophysics can currently provide, we therefore measured human cochlear filters by using psychophysical paradigms specifically designed both to limit the effects of nonlinear compression and suppression and to mimic more closely the procedures used in the measurement of neural tuning curves. Our procedures include the use of (i) relatively low (i.e., near-threshold) signal levels, as in the measurement of neural tuning curves and SFOAEs; (ii) a noise masker extending spectrally both above and below the signal frequency, to avoid effects of “off-frequency listening” (10) and “confusion” between the masker and signal (34); (iii) nonsimultaneous rather than simultaneous masking, to minimize suppressive interactions between the masker and the signal (30); and (iv) constant signal level rather than constant masker level, to mimic the constant-response paradigm used in neural

**Table 1.** Parameters of power-law fits to the functions  $Q_{\text{ERB}}$ ,  $N_{\text{BM}}$ , and  $k$  used to characterize cochlear tuning

		Cat	Guinea pig	Human
$Q_{\text{ERB}}$	$\alpha$	0.37 (10)	0.35 (4)	0.30 (6)
	$\beta$	5.0 (1.1)	4.0 (3)	12.7 (1.0)
$N_{\text{BM}}$	$\alpha$	0.44 (5)	0.44 (3)	0.37 (7)
	$\beta$	1.66 (18)	1.78 (13)	5.5 (6)
$k$	$\alpha$	-0.07 (12)	-0.09 (5)	-0.07 (6)
	$\beta$	3.0 (8)	2.25 (23)	2.30 (16)

Power-law fits (i.e., straight-line approximations on log-log axes) are an excellent approximation at high CFs. For each species, the parameters  $\{\alpha, \beta\}$  characterizing the CF dependence of  $Q_{\text{ERB}}$ ,  $N_{\text{BM}}$ , and  $k$  in the high-frequency region of the cochlea were determined by linear regression using power-law fits of the form  $y = \beta x^\alpha$ , where  $y$  is the dependent variable and  $x = \text{CF}/[\text{kHz}]$  (i.e., CF in kHz). Parameters for the human  $Q_{\text{ERB}}$  were obtained by averaging the otoacoustic and behavioral estimates. The numbers in parentheses give the approximate uncertainty (i.e., 95% confidence interval) in the final digit(s) estimated from the fits [thus,  $0.37(10) = 0.37 \pm 0.1$  and  $0.35(4) = 0.35 \pm 0.04$ ]; when the uncertainty is 1 or greater, the position of the decimal point is shown for clarity. The uncertainties in  $\alpha$  and  $\beta$  are strongly correlated; the typical correlation coefficient between  $\alpha$  and  $\log \beta$  is roughly  $-0.9$ . Note that the equation  $Q_{\text{ERB}} = kN_{\text{BM}}$  implies that  $\beta_Q = \beta_k \beta_N$  and  $\alpha_Q = \alpha_k + \alpha_N$ .

threshold measurements. Earlier studies that measured psychophysical tuning curves often used procedures *i*, *iii*, and *iv*, but not *ii*. Most studies using the notched-noise technique adopt only procedure (ii), although some recent studies also employ *iv* (35, 36). Altogether, only two previous studies have combined all these techniques to measure human auditory-filter bandwidths (37, 38).<sup>‡‡</sup> Neither study, however, provides a firm basis for estimating the frequency dependence of  $Q_{\text{ERB}}$ .

Fig. 4 shows the human behavioral  $Q_{\text{ERB}}$ -versus-frequency function obtained using our modified procedures. The new values differ substantially from standard measurements of cochlear tuning obtained using simultaneous masking (15, 16). The quantitative agreement with the independent, otoacoustic  $Q_{\text{ERB}}$  is especially striking. Table 1 summarizes our results with power-law fits to the functions  $Q_{\text{ERB}}$ ,  $N_{\text{BM}}$ , and  $k$  used to describe human and animal cochlear tuning.

### Discussion

Both our physiological and behavioral measures indicate that human cochlear filters are substantially sharper—by a factor of two or more,<sup>§§</sup> depending on frequency—than commonly believed.<sup>¶¶</sup> In addition, their variation with CF is rather different.

<sup>‡‡</sup>The first study to combine the four techniques found a  $Q_{\text{ERB}}$  value of approximately 10 when using notched-noise at 1 kHz (37). This value falls much closer to the otoacoustic  $Q_{\text{ERB}}$  than does the accepted behavioral value (cf. Figs. 1 and 3), derived from simultaneous-masking studies (15). Although the second study (38) measured forward-masked psychophysical tuning curves at 0.5, 1, 2, and 4 kHz, the two subjects showed substantial differences. At 4 kHz, for example, the bandwidth estimates differed by about a factor of 2.

<sup>§§</sup>The power-law fit to the combined otoacoustic and behavioral measurements implies that at low to moderate sound levels the number of ERBs spanning the frequency range 1–20 kHz is approximately 62, more than twice the number (namely, 26) implied by the standard fit to previous behavioral measurements (15, 16).

<sup>¶¶</sup>Fletcher, in his reports first introducing the power-spectrum model (3, 39), used simultaneous masking and found substantially sharper tuning than suggested by later studies using the same technique. At 1 kHz, for example, Fletcher found a  $Q_{\text{ERB}}$  value of approximately 15.9 [for comparison, the standard behavioral value is 7.5 (15); our estimate, from Table 1, is  $12.7 \pm 1$ ]. Although Fletcher found relatively narrow bandwidths, the frequency dependence of his results generally resembles that found in subsequent simultaneous-masking studies (and therefore differs from the frequency dependence reported here). Much of the discrepancy over the size of the ERB in simultaneous-masking studies has been attributed to Fletcher's assumption of a threshold detection criterion corresponding to a signal-to-noise ratio (SNR) of 1 at the output of the filter; subsequent measurements found typical SNR values (or “critical ratios”) closer to 0.4 (40, 41).

**Table 2. Estimated parameters characterizing the variation of  $Q_{\text{ERB}}$ ,  $N_{\text{BM}}$ , and  $k$  in the basal region of a “generic” mammalian cochlea**

	$\alpha$	$\beta$	$\mu$	$\nu$
$Q_{\text{ERB}}$	0.34 (5)	$\beta_k \beta_N$	1.78 (25)	$\nu_k \nu_N$
$N_{\text{BM}}$	0.42 (3)	$\beta_N$	2.20 (16)	$\nu_N$
$k$	-0.08 (6)	2.5 (3)	-0.4 (3)	1.9 (2)

When the parameters for cat, guinea pig, and human suggest little interspecies variation, the estimates are numerical averages across the three species. When the parameters vary more widely (e.g.,  $\beta$  for  $Q_{\text{ERB}}$  and  $N_{\text{BM}}$ ), we give algebraic formulae indicating how to compute the value from appropriate measurements in the species of interest. For example, the  $\beta$  entry for the mammalian  $Q_{\text{ERB}}$  indicates that  $\beta_Q$  can be obtained from a measurement of  $\beta_N$  for  $N_{\text{BM}}$  and the estimate  $\beta_k \approx 2.5 \pm 0.3$  that follows from the assumption of approximate species-invariance of the ratio  $k$ . Because the optimal estimation procedure is currently unknown, we give two sets of parameters. The parameters  $\{\alpha, \beta\}$  characterize power-law variation of the form indicated in Table 1 and were obtained by averaging across species at corresponding values of CF. The parameters  $\{\mu, \nu\}$  characterize exponential variation of the form  $y = \nu e^{-\mu x}$ , where  $x$  is distance from the stapes normalized by total cochlear length. Values were obtained by averaging across species at corresponding values of  $x$  computed using the cochlear map (13, 26, 27). The numbers in parentheses give the approximate uncertainty (i.e., 95% confidence interval) in the final digit(s) estimated from the fits. As before, the uncertainties are strongly correlated.

Because earlier measurements found near-constancy of  $Q_{\text{ERB}}$  in the base of the cochlea, the frequency analysis performed by the human cochlea at CFs above 500 Hz has been likened to that of a bank of constant- $Q$  filters, or to a continuous wavelet transform (42). Evidently, however, the bandwidths of human cochlear filters do not simply increase in direct proportion to CF (see Table 1), at least at low levels. Instead, filter bandwidths increase at a rate rather less than linear. Specifically, they vary as  $\text{CF}^\alpha$  with  $\alpha = 0.7 \pm 0.06$ , an exponent similar to that in cat ( $0.63 \pm 0.1$ ) and guinea pig ( $0.65 \pm 0.04$ ). At low sound levels and for frequencies above 1 kHz, our physiological and behavioral measurements thus effectively reverse previous characterizations of human peripheral tuning: Whereas earlier measurements suggested that at corresponding cochlear locations the human and cat  $Q_{\text{ERB}}$  have generally similar values (4) but depend differently on CF, the new measures indicate that human tuning is considerably sharper than cat, but varies similarly with CF.

The mutual consistency between our physiological and behavioral measures of tuning supports the assumption of approximate species-invariance of the ratio  $k$  underlying the otoacoustic prediction (e.g., Eq. 1). Our evident success at predicting human cochlear tuning from OAE measurements encourages applications to other mammals, both to explore the method’s generality when independent measures of cochlear tuning exist and to provide noninvasive measures of cochlear tuning in species for which no such measurements are available. We apply this idea in Table 2, which gives parameters characterizing the values of  $Q_{\text{ERB}}$ ,  $N_{\text{BM}}$ , and  $k$  for a “generic mammal.” When the parameters for cat, guinea pig, and human suggest little interspecies variation, we averaged the values across the three species; otherwise we give algebraic formulae based on the assumption of approximate species invariance of the function  $k$ . Unlike previous noninvasive physiological measures of tuning (e.g., those based on evoked potentials or distortion-product OAEs), the results obtained here do not rely on the frequency selectivity of masking or suppression. The development of an objective, noninvasive measure of cochlear tuning solves a long-standing problem in the hearing sciences.

Our two independent measures of cochlear tuning derive from two completely different kinds of measurements interpreted using very different theoretical frameworks involving different

potential sources of uncertainty. Qualitative correspondence, let alone quantitative agreement, between the two measures was by no means assured. The concordance we find is therefore significant in itself. The mutual agreement between physiological and behavioral measures of tuning supports the notion that human auditory frequency selectivity is determined at the level of the periphery (18), at least under the listening conditions probed here.

Cochlear filtering is a dynamic process that no set of static linear filters can completely represent; measurements of filter characteristics therefore need to be understood as applicable within only a limited range of stimulus levels and configurations. Both our physiological and behavioral measurements were obtained at low to moderate signal levels, where cochlear mechanical tuning appears roughly linear, and our conclusions therefore apply in this regime. At higher levels, the “effective” filter bandwidths may be broader. Although our confidence in the otoacoustic measures of  $N_{\text{BM}}$  limits the discussion here to CFs of 1 kHz and above, our general conclusions may also apply at lower frequencies. For example, if the human  $Q_{\text{ERB}}$  resembles those in cat and guinea pig in having an almost constant slope on log-log axes (cf. Fig. 1), extrapolating the power-law behavior of the new human  $Q_{\text{ERB}}$  to lower frequencies suggests that human filter bandwidths are considerably narrower than previously believed throughout the cochlea.

Our findings raise new questions with important consequences for theories of hearing. The revised estimates of human cochlear tuning inform issues as basic as our understanding of the cochlear frequency-position map and the spatial significance of the critical band. If the human cochlear map is logarithmic in the base of the cochlea, as suggested by physiological measurements in cat and guinea pig (13, 27), the evident nonconstancy of  $Q_{\text{ERB}}$  above 1 kHz implies that the ERB does not correspond to a constant distance along the basilar membrane, as is often suggested (4, 43). Because the constant-distance assumption underlies the derivation of the human cochlear map from behavioral measurements (26, 43), our results imply that the relation between the cochlear map and the spatial correlate of the critical band needs reexamination. For example, Allen provides evidence for Fletcher’s suggestion that the spatial correlate of the ERB (the “equivalent rectangular spread”) depends on the width the basilar membrane and therefore varies with position in the cochlea (44).

Although further exploration is beyond the scope of this paper, our finding that cochlear tuning appears markedly sharper in humans than in cats or guinea pigs raises fundamental questions about the mechanical, biophysical, and evolutionary origins of these prominent species differences.<sup>||</sup> One might, for instance, speculate that sharper cochlear tuning facilitates speech-like acoustic communication. Although fine frequency selectivity is not required for basic speech reception in quiet (46), greater selectivity is required both for reception in a background of noise (47) and for the perceptual segregation of sounds in complex acoustic environments.

Many models of auditory perception are based on initial filtering algorithms that mimic the broader selectivity found in earlier studies. In some cases, such as the models of simultaneous masking used in digital-audio compression algorithms (e.g., MPEG-1 Layer 3, commonly known as MP3), these algorithms remain justified at sound levels where the “effective” filter shapes are determined by many of the nonlinear interactions that we sought to eliminate. However, our results imply that numerous models of loudness and frequency perception (e.g., refs. 11,

<sup>||</sup>Measurements of SFOAEs in rhesus monkeys indicate that  $N_{\text{SFOAE}}$  at frequencies of 1–2 kHz is nearly the same as that in humans (45), suggesting that the sharpness of tuning is generally similar among primates.

48, and 49), which explicitly assume that standard behavioral results accurately describe cochlear tuning, are in error. A revision of these models to take into account sharper human cochlear tuning at low levels appears necessary.

We gratefully acknowledge the help of Leslie Liberman, who provided invaluable assistance with animal care and preparation. We thank our subjects—the psychophysical for their hard work, the otoacoustic for

their patience—and M. Charles Liberman for generously sharing his data for analysis. Finally, we thank Jont Allen, Laurel Carney, Paul Fahey, Michael Heinz, Brian C.J. Moore, and William Peake for their helpful comments on the manuscript. A.J.O. thanks the Hanse Institute for Advanced Study in Delmenhorst, Germany for the warm hospitality extended during his tenure as a fellow. This work was supported by Grants DC03687, DC00235, and DC03909 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health.

1. Peters, R. W., Moore, B. C. J. & Baer, T. (1998) *J. Acoust. Soc. Am.* **103**, 577–587.
2. Kiang, N. Y. S., Watanabe, T., Thomas, E. C. & Clark, L. F. (1965) *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve* (MIT Press, Cambridge, MA).
3. Fletcher, H. (1938) *Proc. Natl. Acad. Sci. USA* **24**, 265–274.
4. Moore, B. C. J. (1986) *Scand. Audiol. Suppl.* **25**, 139–152.
5. Shera, C. A. & Guinan, J. J. (1999) *J. Acoust. Soc. Am.* **105**, 782–798.
6. Guinan, J. J. (1986) *Scand. Audiol. Suppl.* **25**, 53–62.
7. Shera, C. A., Talmadge, C. L. & Tubis, A. (2000) *J. Acoust. Soc. Am.* **108**, 2933–2948.
8. Dreisbach, L. E., Siegel, J. H. & Chen, W. (1998) *Assoc. Res. Otolaryngol. Abs.* **21**, 349.
9. Patterson, R. D., Nimmo-Smith, I., Weber, D. L. & Milroy, R. (1982) *J. Acoust. Soc. Am.* **72**, 1788–1803.
10. Glasberg, B. R., Moore, B. C. J., Patterson, R. D. & Nimmo-Smith, I. (1984) *J. Acoust. Soc. Am.* **76**, 419–427.
11. Moore, B. C. J., Glasberg, B. R. & Baer, T. (1997) *J. Aud. Eng. Soc.* **45**, 224–240.
12. Liberman, M. C. (1990) *Hear. Res.* **49**, 209–224.
13. Tsuji, J. & Liberman, M. C. (1997) *J. Comp. Neurol.* **381**, 188–202.
14. Evans, E. F. & Wilson, J. P. (1973) in *Basic Mechanisms in Hearing*, eds. Møller, A. R. & Boston, P. (Academic, New York), pp. 519–551.
15. Glasberg, B. R. & Moore, B. C. J. (1990) *Hear. Res.* **47**, 103–138.
16. Moore, B. C. J. (1997) *An Introduction to the Psychology of Hearing*. (Academic, New York).
17. Pickles, J. O. (1979) *J. Acoust. Soc. Am.* **66**, 1725–1732.
18. Evans, E. F., Pratt, S. R., Spenner, H. & Cooper, N. P. (1992) in *Auditory Physiology and Perception*, eds. Cazals, Y., Horner, K. & Demany, L. (Pergamon, Oxford), pp. 159–169.
19. Kemp, D. T. (1978) *J. Acoust. Soc. Am.* **64**, 1386–1391.
20. Zweig, G. & Shera, C. A. (1995) *J. Acoust. Soc. Am.* **98**, 2018–2047.
21. Talmadge, C. L., Tubis, A., Long, G. R. & Piskorski, P. (1998) *J. Acoust. Soc. Am.* **104**, 1517–1543.
22. Shera, C. A. & Guinan, J. J. (2000) in *Recent Developments in Auditory Mechanics*, eds. Wada, H., Takasaka, T., Ikeda, K., Ohyama, K. & Koike, T. (World Scientific, Singapore), pp. 381–387.
23. Shera, C. A. & Guinan, J. J. (2000) *Assoc. Res. Otolaryngol. Abs.* **23**, 545.
24. Zweig, G. (1976) *Cold Spring Harbor Symp. Quant. Biol.* **40**, 619–633.
25. Narayan, S. S., Temchin, A. N., Recio, A. & Ruggero, M. A. (1998) *Science* **282**, 1882–1884.
26. Greenwood, D. D. (1990) *J. Acoust. Soc. Am.* **87**, 2592–2605.
27. Liberman, M. C. (1982) *J. Acoust. Soc. Am.* **72**, 1441–1449.
28. Cooper, N. P. & Yates, G. K. (1994) *Hear. Res.* **78**, 221–234.
29. Kanis, L. J. & de Boer, E. (1993) *J. Acoust. Soc. Am.* **94**, 3199–3206.
30. Houtgast, T. (1973) *Acustica* **29**, 168–179.
31. Delgutte, B. (1990) *J. Acoust. Soc. Am.* **87**, 791–809.
32. Oxenham, A. J. & Plack, C. J. (1998) *J. Acoust. Soc. Am.* **104**, 3500–3510.
33. Heinz, M. G., Colburn, H. S. & Carney, L. H. (2002) *J. Acoust. Soc. Am.*, in press.
34. Neff, D. L. (1985) *J. Acoust. Soc. Am.* **78**, 1966–1976.
35. Rosen, S. R., Baker, R. J. & Darling, A. (1998) *J. Acoust. Soc. Am.* **103**, 2539–2350.
36. Glasberg, B. R. & Moore, B. C. J. (2000) *J. Acoust. Soc. Am.* **108**, 2318–2328.
37. Glasberg, B. R. & Moore, B. C. J. (1982) *J. Acoust. Soc. Am.* **71**, 946–949.
38. Moore, B. C. J., Glasberg, B. R. & Roberts, B. (1984) *J. Acoust. Soc. Am.* **76**, 1057–1066.
39. Fletcher, H. (1940) *Rev. Mod. Phys.* **12**, 47–65.
40. Scharf, B. (1970) in *Foundations of Modern Auditory Theory*, ed. Tobias, J. (Academic, New York), pp. 159–202.
41. Patterson, R. D. & Moore, B. C. J. (1986) in *Frequency Selectivity in Hearing*, ed. Moore, B. C. J. (Academic, London), pp. 123–177.
42. Irino, T. & Patterson, R. D. (1997) *J. Acoust. Soc. Am.* **101**, 412–419.
43. Greenwood, D. D. (1961) *J. Acoust. Soc. Am.* **33**, 1344–1356.
44. Allen, J. B. (1996) *J. Acoust. Soc. Am.* **99**, 1825–1839.
45. Lonsbury-Martin, B. L., Martin, G. K., Probst, R. & Coats, A. C. (1988) *Hear. Res.* **33**, 69–94.
46. Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J. & Ekelid, M. (1995) *Science* **270**, 303–304.
47. Dorman, M. F., Loizou, P. C., Fitzke, J. & Tu, Z. (1998) *J. Acoust. Soc. Am.* **104**, 3583–3585.
48. Florentine, M. & Buus, S. (1981) *J. Acoust. Soc. Am.* **70**, 1646–1654.
49. Zwicker, E. (1956) *Acustica* **6**, 356–381.