Hearing Research 420 (2022) 108500

Contents lists available at ScienceDirect

Hearing Research

journal homepage: www.elsevier.com/locate/heares

Auditory filter shapes derived from forward and simultaneous masking at low frequencies: Implications for human cochlear tuning

John Leschke[a,1], Gerardo Rodriguez Orellana[b,1], Christopher A. Shera[c], Andrew J. Oxenham[d,*]

[a] Department of Neurosurgery, University of Minnesota, Minneapolis, MN 55455, USA
[b] Department of Biomedical Engineering, University of Minnesota, Minneapolis, MN 55455, USA
[c] Auditory Research Center, Caruso Department of Otolaryngology, University of Southern California, Los Angeles, CA 90033, USA
[d] Department of Psychology, University of Minnesota, Minneapolis, MN 55455, USA

ARTICLE INFO

Article history:
Received 17 April 2021
Revised 8 March 2022
Accepted 28 March 2022
Available online 31 March 2022

Keywords:
Frequency selectivity
Auditory masking
Nonsimultaneous masking
Cochlear apex

ABSTRACT

Behavioral forward-masking thresholds with a spectrally notched-noise masker and a fixed low-level probe tone have been shown to provide accurate estimates of cochlear tuning. Estimates using simultaneous masking are similar but generally broader, presumably due to nonlinear cochlear suppression effects. So far, estimates with forward masking have been limited to frequencies of 1 kHz and above. This study used spectrally notched noise under forward and simultaneous masking to estimate frequency selectivity between 200 and 1000 Hz for young adult listeners with normal hearing. Estimates of filter tuning at 1000 Hz were in agreement with previous studies. Estimated tuning broadened below 1000 Hz, with the filter quality factor based on the equivalent rectangular bandwidth (Q_{ERB}) decreasing more rapidly with decreasing frequency than predicted by previous equations, in line with earlier predictions based on otoacoustic-emission latencies. Estimates from simultaneous masking remained broader than those from forward masking by approximately the same ratio. The new data provide a way to compare human cochlear tuning estimates with auditory-nerve tuning curves from other species across most of the auditory frequency range.

1. Introduction

Frequency-to-place mapping, or tonotopy, is a primary organizing principle of the auditory system. It is established in the cochlea along the length of the basilar membrane (Russell and Nilsen, 1997; von Békésy, 1960), and is maintained throughout the auditory pathways up to and including auditory cortex (Allen et al., 2022; De Martino et al., 2013; Formisano et al., 2003; Moerel et al., 2012; Narayan et al., 1998; Reale and Imig, 1980; Schreiner and Langner, 1988). The degree to which different frequencies stimulate different sensory cells, avoiding overlap and interference, has been termed frequency selectivity. Perceptually, frequency selectivity underlies our ability to detect one sound in the presence of another competing sound with different frequency content. This ability has been quantified in a variety of ways (Helmholtz, 1885; Plomp, 1964; Zwicker et al., 1957), but the most widespread method involves perceptual masking of a tone in the presence of noise, with its spectral characteristics defined by bandpass or bandstop filtering (Fletcher, 1940; Patterson, 1974, 1976; Wegel and Lane, 1924).

The most influential estimates of human behavioral frequency selectivity, published by Glasberg and Moore (1990), resulted in an equation that provides the estimated equivalent rectangular bandwidth (ERB) of the auditory filters at moderate sound levels as a function of center frequency (CF) between 100 and 10 kHz. This equation differs somewhat from the earlier estimates of the critical bands (CBs; Bark scale) of Zwicker (1961) in two important ways. First, the ERBs are somewhat narrower than the equivalent CBs; second, the absolute ERB values continue to decrease at CFs below 500 Hz, albeit at a slower rate (Glasberg and Moore, 1990; Moore et al., 1990), whereas the CB remains roughly constant at around 100 Hz for CFs below 500 Hz (Scharf, 1970; Zwicker, 1961). For frequencies at and above about 1000 Hz, both scales suggest bandwidths that are roughly proportional to CF, implying a constant quality factor (Q).

The ERB estimates of Glasberg and Moore (1990) have gained wide acceptance, and provide good estimates of frequency selectivity at moderate sound levels under conditions of simultaneous

* Corresponding author.
E-mail address: oxenham@umn.edu (A.J. Oxenham).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.heares.2022.108500
0378-5955/

masking. However, more recent work has suggested that their estimates may not be appropriate for direct comparisons with physiological estimates of cochlear tuning (Shera et al., 2002). Cochlear filtering is highly nonlinear in nature, resulting in level-dependent tuning and interactions between simultaneously presented sounds (Rhode and Cooper, 1993; Sachs and Kiang, 1968). Because of these nonlinearities, psychophysical estimates of filter bandwidth depend on many factors, including stimulus level, whether the masker is simultaneous or non-simultaneous, and whether the masker or signal vary in level and/or frequency (e.g., Eustaquio-Martin and Lopez-Poveda, 2011; Lopez-Poveda and Eustaquio-Martin, 2013; Verschuure, 1981a, 1981b). To make human behavioral estimates of frequency selectivity as comparable as possible to those of auditory-nerve tuning curves, non-simultaneous (forward) masking can be used to avoid direct cochlear interactions, such as suppression, between the masker and target (Houtgast, 1973; Moore, 1978; Shannon, 1976), and the target tone can be presented at a fixed low level, with the masker level adaptively varied to more closely resemble the paradigm used in physiological studies. Although this procedure provides more similarity to the procedure used to measure auditory-nerve tuning curves, there still remain potentially important differences, such as the use of noise-bands (instead of just pure tones), which may interact nonlinearly with themselves (Houtgast, 1972). Nevertheless, when combining these steps to estimate human frequency selectivity (Glasberg and Moore, 1982; Oxenham and Shera, 2003; Oxenham and Simonson, 2006; Shera et al., 2002), the resulting bandwidths are narrower than those reported by Glasberg and Moore (1990) and demonstrate a different relationship with CF, becoming increasingly sharply tuned for CFs between 1 and 8 kHz, rather than having a roughly constant *Q*.

The observed increasing filter sharpness with increasing CF, quantified as the quality factor based on the ERB (Q_{ERB}) or CF/ERB, is similar to that found in the auditory-nerve tuning-curve data from other species, including cat, guinea pig, chinchilla, and ferret, with one important difference: human tuning is sharper across all CFs measured by a factor of between 2 and 3 (Shera et al., 2002, 2010; Sumner et al., 2018). The sharper tuning estimated for humans is consistent with the correspondingly longer latencies of stimulus-frequency otoacoustic emissions (Shera et al., 2002, 2010). Supporting the idea that these estimates reflect underlying cochlear tuning, behavioral estimates for ferrets, obtained using the forward-masking notched-noise method, align well with tuning estimates from both auditory-nerve tuning curves and otoacoustic emissions from the same species (Sumner et al., 2018). A new equation to describe human frequency selectivity was proposed by Oxenham and Shera (2003) as follows:

$$Q_{ERB} = 11.1F^{0.27}, \quad (1)$$

where F is the filter CF in kHz. Although this equation provides a good fit to existing human behavioral data, and provides a closer parallel to existing auditory-nerve data in other species, it is limited to CFs between 1 and 8 kHz, as 1 kHz was the lowest CF used in the behavioral task. Thus, it is unknown whether this equation can be extrapolated to lower frequencies. This question is important for several reasons: First, earlier estimates of frequency selectivity diverged most at CFs below 1 kHz (Glasberg and Moore, 1990; Zwicker, 1961), making this a frequency region where additional clarity is most needed. Second, it is unknown whether the differences in tuning estimates observed between forward and simultaneous masking (Oxenham and Shera, 2003) extend to CFs below 1 kHz, or whether the estimates converge at lower CFs. Convergence is a possibility, because the differences between forward and simultaneous masking are thought to be depend on cochlear nonlinearities, which may be reduced, or at least differ in nature, in the lower-CF apical regions of the cochlea (Cooper and Rhode, 1995; Recio-Spinoso and Oghalai, 2017, 2018; Versteegh et al., 2011). Third, there is considerable evidence from laboratory animals that tuning and other response characteristics differ significantly between the basal and apical regions of the cochlea with a transition zone, the location of which varies across species (Shera et al., 2010). Specifically, based on auditory-nerve and otoacoustic-emission data, it has been proposed that the apical-basal transition occurs near the 3-4 kHz place in cat, chinchilla, and ferret; in humans, based solely on otoacoustic-emission data, the transition appears to be closer to the 1-kHz place (see Table 1 of Shera et al., 2010).

The purpose of this study was to characterize human cochlear tuning, as estimated via behavioral notched-noise forward masking, at CFs between 0.2 and 1 kHz. Estimates using simultaneous masking were collected for comparison. The results suggest a relative broadening of tuning at CFs below 1 kHz, beyond that predicted by the original power law proposed by Oxenham and Shera (2003). Combined with the earlier data, these behavioral estimates of cochlear tuning are consistent with the predicted apical-basal transition point of around 1 kHz. The relationship between CF and Q_{ERB} can be described by different power laws above and below that transition point, or by a single logarithmic function.

2. Methods

2.1. Listeners

A total of 23 young normal-hearing listeners participated in these experiments. Thirteen listeners (6 female, 7 male, ages 18–38 years, mean age 25.2 years) took part in the forward-masking experiment and the remaining ten (4 female, 6 male, ages 21–27 years, mean age 22.9 years) took part in the simultaneous-masking experiment. Due to a technical error in their absolute-threshold measures, data from two of the listeners in the simultaneous-masking experiment were excluded, so the reported data consist of thirteen listeners under forward masking and eight listeners under simultaneous masking. Normal hearing was defined as having audiometric thresholds of 10 dB hearing level (HL) or better at octave frequencies between 0.25 and 8 kHz. All listeners provided written informed consent and were paid for their participation. All protocols were approved by the Institutional Review Board of the University of Minnesota.

2.2. Stimuli

All stimuli were generated digitally at a sampling rate of 48 kHz and were converted via an E22 soundcard (Lynx Studio Technologies, Costa Mesa, CA) at 24-bit resolution. The stimuli were presented to listeners seated individually in a sound-attenuating chamber monaurally to the left ear via HD650 headphones (Sennheiser, Old Lyme, CT).

2.2.1. Forward masking

The signal was a tone burst with a total duration of 25 ms, including 12.5-ms raised-cosine onset and offset ramps (no steady-state portion). The signal frequency (f_s) was 200, 350, 500, 750, or 1000 Hz. The masker was a Gaussian noise with a bandwidth extending from 20 to 2000 Hz. The masker duration was 200 ms, also including 12.5-ms raised-cosine onset and offset ramps. The signal was presented at 12 dB sensation level (SL), as measured for each listener individually, directly after the masker, with no gap between the end of the masker and the beginning of the signal, resulting in a half-amplitude gap between the masker and signal of 12.5 ms.

The noise maskers were generated with a spectral notch around the signal frequency. In the five symmetric notch conditions, the

normalized deviation of the notch's lower and upper spectral edge and the signal frequency, $\Delta f/f_s$, was set to 0 (no notch), 0.1, 0.2, 0.3, or 0.4, resulting in notch widths of 0, 0.2, 0.4, 0.6, and $0.8f_s$, with the signal spectrally centered (on a linear scale) within the notch. In the four asymmetric conditions, the respective lower and upper normalized deviations of the edge of the spectral notch from the signal frequency were 0.1 and 0.3, 0.2 and 0.4, 0.3 and 0.1, and 0.4 and 0.2. The noise was generated in the spectral domain by setting the magnitude of components outside the desired passband to zero and was then converted to the time domain via an inverse Fourier transform.

2.2.2. Simultaneous masking

The signal was a tone burst with a total duration of 50 ms, including 25-ms raised-cosine onset and offset ramps (no steady-state portion). A longer signal duration was used, since the very short duration needed for forward masking was not necessary here. The noise masker was again 200 ms in total duration, also including 25-ms raised-cosine onset and offset ramps. The signal was temporally centered within the masker. All other properties of the signal and masker were the same as in the forward-masking conditions, with 5 symmetric and 4 asymmetric notches tested at each of the 5 signal frequencies.

When the signal frequency was 200 Hz, additional conditions were run with the signal duration and ramps reduced to the same durations as used in the forward-masking conditions, to test whether the signal and ramp durations (and hence the bandwidth of the signal) affected estimates of frequency selectivity. No effect was expected, but if there were an effect, it was reasoned that it would be greatest at the lowest CF, where the signal's bandwidth would be the greatest proportion of the CF, and thus most likely to have a significant effect.

2.3. Procedure

All thresholds were measured using a two-interval, two-alternative forced-choice task. Initially, the absolute thresholds for the signal were measured at all tested frequencies for each listener's left ear. The total signal durations were 25 and 50 ms, for use in the forward-masking and simultaneous-masking conditions, respectively. Once absolute thresholds had been determined, the signal level was fixed at 12 dB above the measured absolute threshold for each listener.

In each trial, two intervals were presented, separated by a 500-ms inter-stimulus interval, with the intervals marked by lights on a virtual button box presented on a monitor. The listeners had to select the interval that contained the signal (chosen at random on each trial). Feedback was provided after each trial. When measuring absolute thresholds, the signal level was varied adaptively to track threshold using a 3-down 1-up procedure; when measuring thresholds in notched noise, the signal level was kept fixed at 12 dB above absolute threshold and the spectrum level of the notched noise was varied adaptively using a 3-up 1-down procedure (Levitt, 1971). At the beginning of each run, the initial level of the variable stimulus was selected to ensure that the signal was clearly audible. The initial step size of the adaptive procedure was 8 dB. This was reduced to 4 dB after two reversals in the direction of the change in the level, and was reduced to a final value of 2 dB after a further four reversals. The run continued for another six reversals at the smallest step size, and the threshold for the run was defined as the mean of the levels at the last six reversal points. Each condition was measured three times for each listener. In some rare instances, when ten reversals had not occurred within 100 trials, the run was abandoned. For the listeners in the forward-masking condition, those conditions were rerun. For listeners in the simultaneous-masking condition (involving 12 out of a total of 1296 runs, or 1.1%), only the remaining two runs were recorded. In some other rare instances, individual runs produced masker levels at threshold that were far below the other two runs, suggesting that the participant had lost concentration. To avoid including such runs in the mean estimates, the following rule was adopted: If the masker spectrum level for an individual run was below −5 dB SPL (−12 dB under simultaneous masking) and if that level was at least four standard deviations below the mean of the other two runs for that condition, then the individual run was discarded. This occurred for 24 runs (1.37%) under forward masking and 24 runs (1.85%) under simultaneous masking. The mean of the remaining valid runs (at least two in each case) defined threshold for each participant in each condition. All the stimuli and procedures were created in Matlab (The Mathworks, Natick, MA), using the AFC toolbox (Ewert, 2013).

2.4. Fitting auditory filter shape

Auditory filters were estimated from the individual as well as the mean data. The frequency response of Sennheiser HD 650 headphones is relatively flat in the frequency region of interest (20 to 2000 Hz), so no headphone compensation was included in the model. The model incorporated a transfer function designed to simulate the effect of the middle ear, as described by Moore et al. (1997). The middle-ear function plays a particularly important role at frequencies below about 500 Hz, where the roll-off in sensitivity with decreasing frequency is relatively steep (Puria et al., 1997). As was found in an earlier study (Oxenham and Shera, 2003), eliminating the simulation of the middle-ear transfer function had no systematic effect on the estimated filter bandwidths, but did affect the estimated asymmetry of the filters.

The assumed shape of the auditory filter was a variant of the rounded exponential (roex) function (Patterson and Nimmo-Smith, 1980), referred to in previous studies as the roex(p_l,w,t,p_u) filter (Oxenham and Shera, 2003). This shape provides an initial steep slope to the lower side of the filter (determined by slope parameter p_l), along with a shallower "tail" to the filter slope (characterized by t). The transition between the steeper and shallower slopes is determined by a weight that ranges from 0 to 1 (w). The upper side of the filter is defined by a single slope parameter (p_u). This shape has been shown to provide good fits to existing notched-noise datasets (Glasberg and Moore, 2000; Rosen et al., 1998) and it approximates the general shape of auditory-nerve tuning curves at high frequencies (> 1 kHz), with a sharp tip and a broader low-frequency tail (Kiang et al., 1967). The equation for the lower side of the filter is:

$$W(g) = (1 - w)(1 + p_l g)e^{-p_l g} + w(1 + p_l g/t)e^{-p_l g/t}, \quad (2)$$

where g is the normalized absolute frequency deviation (Δf) from the filter's CF (|CF-Δf|/CF). The upper side of the filter does not have the shallower slope and so is defined simply by:

$$W(g) = (1 + p_u g)e^{-p_u g}. \quad (3)$$

To determine the best-fitting parameters, a least-squares minimization routine was implemented with Matlab's built-in multidimensional unconstrained nonlinear minimization routine that uses the Nelder-Mead numerical method (fminsearch). This method adjusts the filter parameters to minimize the sum of squared differences in the filter output signal-to-masker ratio (in dB) across all conditions at a given signal frequency, with the signal and masker levels fixed at the threshold levels determined empirically in each condition. Each fit was repeated at least 10 times with different starting parameter values to avoid basing final fits on local minima found by the fitting routine. The fit with the lowest variance in signal-to-noise ratio (SNR) at the output of the filter (in dB) was

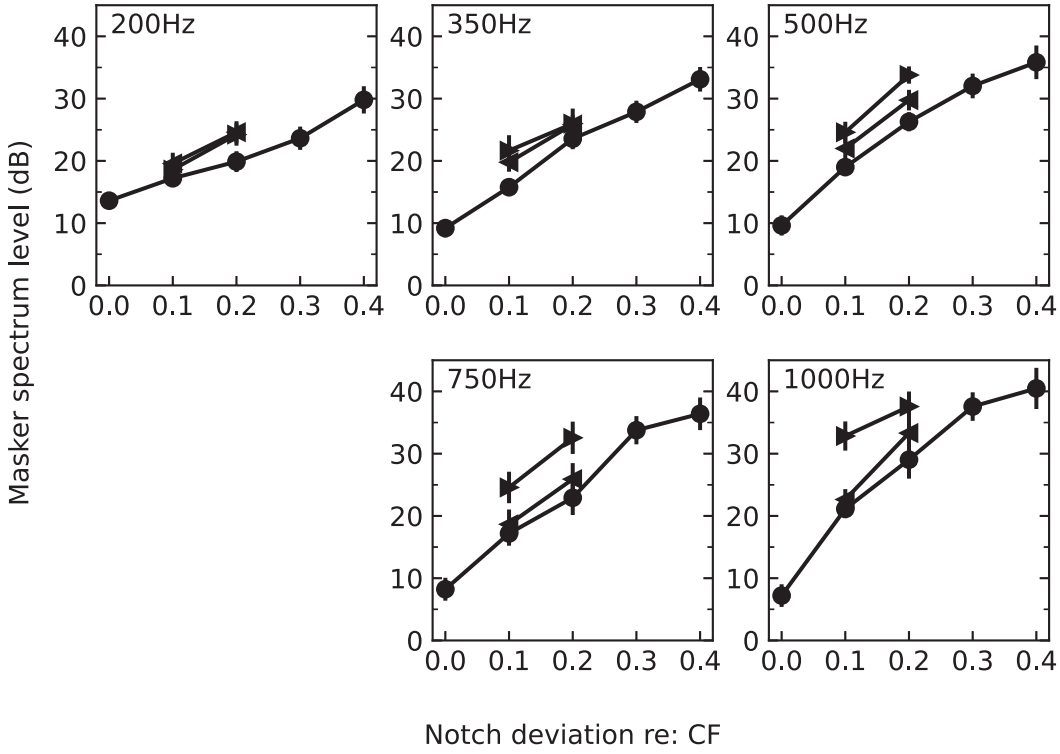

Fig. 1. Mean data from the forward-masking experiment ($N = 13$). Masker spectrum level (re: 20 μPa) at threshold is plotted as a function of the normalized frequency distance (relative to the signal frequency) between the signal and the spectral edge of the noise. Circles denote thresholds in the symmetric notch conditions, where the notch width is twice the notch deviation. Triangles denote asymmetric conditions and are positioned relative to the spectral edge of the notch closest to the signal frequency. Right-pointing triangles correspond to conditions in which the upper edge of the notch was set to 0.3 or 0.4f_s, while the lower edge was set to 0.1 or 0.2 f_s, respectively. Left-pointing triangles correspond to conditions in which the upper edge of the notch was set to 0.1 or 0.2f_s, and the lower edge of the notch was set to 0.3 or 0.4 f_s, respectively. Error bars represent $\pm$1 standard error of the mean (SEM) and are shown when they exceed the symbol size.

selected in each condition. It was assumed that the signal was detected via the filter with the best SNR, even if the filter was not centered directly on the signal frequency ("off-frequency listening"); all filters with CFs within $\pm$10% of the signal frequency were considered in this calculation but in practice the best SNR was usually found for a filter with a CF at or very close to the signal frequency. The resulting best-fitting parameters were used to derive the ERB and Q_{ERB} values for each individual and the individual estimates were geometrically averaged to determine the mean values at each signal frequency. The mean Q_{ERB} across listeners was very similar to the Q_{ERB} of the filter fitted to the data pooled across all listeners, as reported in earlier studies (Oxenham and Shera, 2003; Oxenham and Simonson, 2006).

3. Results

3.1. Masked thresholds

3.1.1. Forward masking

Mean thresholds (and standard deviations) in quiet for the 25-ms signal were 26.6 (2.2), 21.4 (2.8), 18.4 (2.8), 15.1 (3.5), and 13.2 (2.7) dB SPL at 200, 350, 500, 750, and 1000 Hz, respectively. Mean masker spectrum levels at threshold for the different notch widths with the signal presented at 12 dB sensation level (SL) are shown in Fig. 1, with the results from each signal frequency shown in a separate panel. Thresholds in the presence of the symmetric notches are shown with circles, while thresholds in the asymmetric notch conditions are shown with triangles. The direction of the triangle represents the direction of the asymmetry of the notch, relative to the signal frequency, so that left-pointing triangles represent conditions with the center of the notch lower than the signal frequency and the right-pointing triangles represent conditions with the center of the notch higher than the signal frequency. Symmetric filters (ignoring the effects of the middle ear transfer function), would result in left- and right-pointing triangles being roughly coincident. The fact that this appears to be the case at the lowest signal frequencies of 200 and 350 Hz suggests that the combination of middle-ear and auditory filtering results in a roughly symmetric function, implying that the auditory filters themselves are somewhat asymmetric, with a shallower lower than upper slope. An asymmetry emerges at the higher frequencies, with the right-pointing triangles being generally higher than the left-pointing triangles. This asymmetry again implies a shallower lower slope than upper slope, consistent with previous studies at signal frequencies of 1 kHz and above (Oxenham and Shera, 2003). Considering the symmetric notch conditions (circles), the slope relating masker level at threshold to the symmetric notch deviation seems to become steeper with increasing signal frequency, implying sharper filter tuning with increasing CF between 200 and 1000 Hz, as expected.

The within-listener variability in the threshold estimates was quite large. Across all conditions and listeners, the estimated within-listener standard deviation (across the three threshold estimates per condition, estimated by averaging the estimated variance across conditions and taking the square root of the estimated variance) was 7.3 dB. This value varied considerably between individual listeners, with a range from 2.3 to 12.2 dB across the 13 listeners (median value 6.9 dB). However, no clear patterns emerged regarding whether specific conditions (notch widths or signal frequencies) produced more or less variable threshold estimates than others.

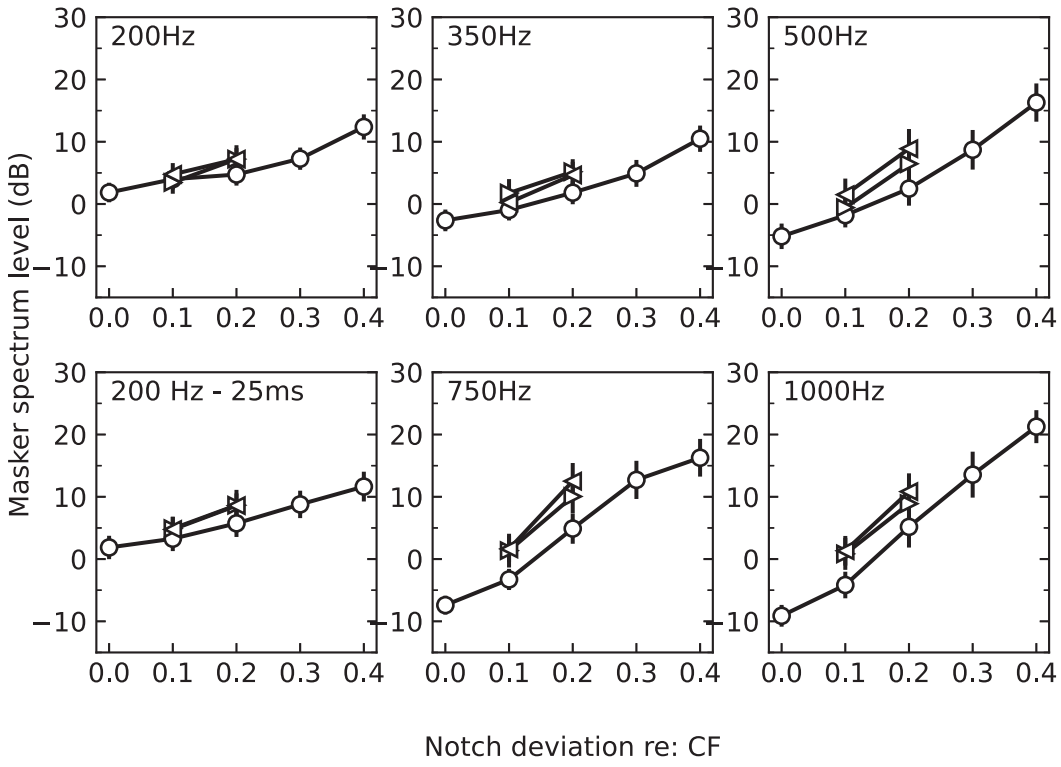

Fig. 2. Mean data from the simultaneous-masking experiment ($N = 8$). Masker spectrum level (re: 20 μPa) at threshold is plotted as a function of normalized notch distance from signal frequency. Open symbols are used here and elsewhere to denote simultaneous masking. Otherwise the symbols and error bars have the same meaning as in Fig. 1.

3.1.2. Simultaneous masking

Mean thresholds (and standard deviations) in quiet for the 50-ms signal were 20.3 (5.1), 15.7 (4.9), 13.4 (5.4), 9.8 (5.8), and 10.1 (4.2) dB SPL at 200, 350, 500, 750, and 1000 Hz, respectively. The additional threshold measurements for the 200-Hz 25-ms signal in quiet for the listeners tested under simultaneous masking yielded a mean threshold of 23.7 (5.8) dB SPL, which was not significantly different from the threshold measured for the same signal with listeners in the forward-masking experiment [independent-samples t-test: $t(19) = 1.66$, $p = 0.11$]. The within-listener variability across repetitions of a given condition was considerably less than under forward masking, with an estimated standard deviation across the three estimates per listener of 3.3 dB. This value also varied across listeners, ranging from 1.9 to 5.0 dB (median value 3.2 dB). Mean masker spectrum levels at threshold are shown in Fig. 2, with the same format as in Fig. 1, but under simultaneous masking. The general pattern of results was similar, with increasing masking spectrum level at threshold as a function of the notch width. The asymmetries in masking seem less pronounced under simultaneous masking than under forward masking; if anything, the left-pointing triangles were slightly above the right-pointing triangles, in contrast to what was observed at the higher signal frequencies under forward masking. Results with the 25-ms signal were very similar to those obtained with the 50-ms signal (cf. upper and lower left panels of Fig. 2), both in terms of absolute masker levels and changes in masker level as a function of notch width, suggesting that the choice of signal duration did not affect estimates of frequency selectivity.

3.2. Auditory filter shape and bandwidth as a function of signal frequency

3.2.1. Filters derived from pooled data

To obtain a general impression of how filter shape and bandwidth varied with CF under both forward and simultaneous masking, the data were pooled across listeners in each of the two experiments, and the roex(p,w,t,p) filters were fitted to the pooled datasets. The filters derived from the data pooled across all listeners are shown in Fig. 3 for both forward-masking (solid lines) and simultaneous-masking (dashed lines) conditions. The filters derived from the forward-masking data had sharper tuning around the filter tip, with steeper upper slopes, compared to those derived from simultaneous-masking data, but also tended to have shallower "tails" on the lower side of the filter. The filters increased in sharpness as the frequency increased. In the simultaneous-masking condition, the lower and upper slopes appear quite symmetric, whereas the forward-masking filters show a more pronounced asymmetry at the higher frequencies, as expected from the raw data shown in Figs. 1 and 2.

3.2.2. Filters derived from individual data

Individual data were used to derive filter functions for each listener, and a bootstrapping approach was used to estimate within-listener variability in filter bandwidth estimates. The bootstrapping involved resampling the individual threshold estimates from each notch width with replacement and then fitting the filter based on the resampled data. This was done 100 times for each listener's data to obtain an estimate of the distribution of Q_{ERB} values. In most cases, the mean of the distribution of the 100 Q_{ERB} estimates for each listener and frequency was very close to the Q_{ERB} estimated directly from the listener's raw data (with no resampling) in each condition. However, the variability differed substantially between individual listeners. This pattern is illustrated in Fig. 4, which shows examples of Q_{ERB} distributions from the bootstrapped data of three sample listeners whose data exhibited low, medium, and high variability, from both forward and simultaneous masking. The estimated standard deviations of the individual bootstrapped log-transformed Q_{ERB} distributions (across listeners and frequencies) were factors (i.e., mean $\times$ /$\div$ SD) of 1.4 and 1.2 for the forward- and simultaneous-masking conditions, respec-

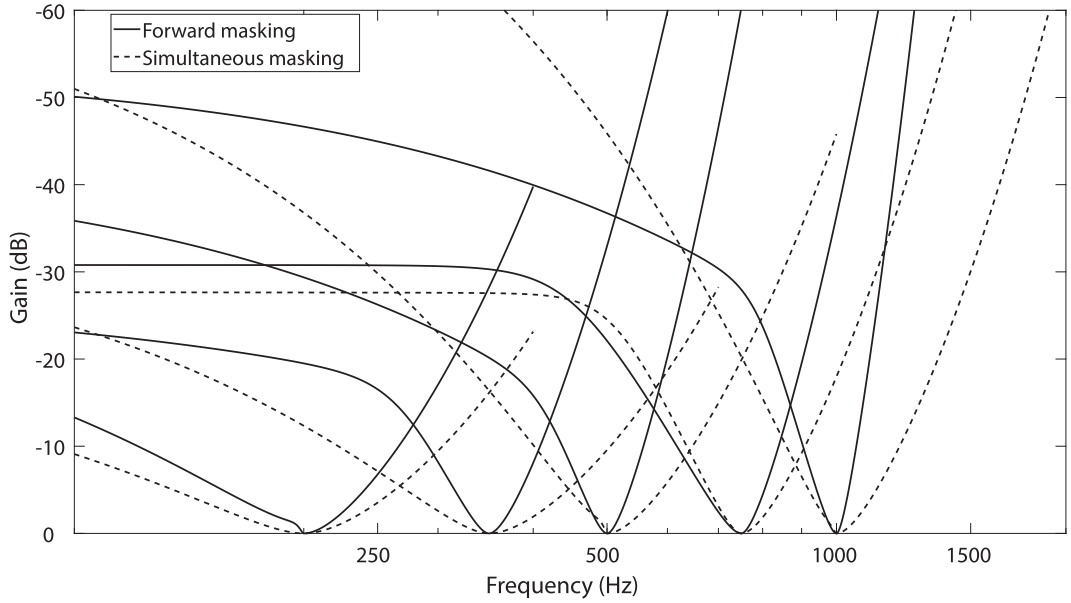

Fig. 3. Auditory filter shapes with center frequencies between 200 and 1000 Hz, derived from the pooled data in each experiment. Solid lines correspond to filters derived from data under forward masking; dashed lines correspond to filters derived from data under simultaneous masking.

tively. Across listeners, the standard deviations ranged from factors of 1.1 to 1.6 (median 1.3) for forward masking and from factors of 1.1 to 1.3 (median 1.2) for simultaneous masking.

The individual (geometric) mean bootstrapped Q_{ERB} estimates are shown as a function of frequency in the top panel of Fig. 5, along with the geometric means across listeners, for both forward- and simultaneous-masking conditions (filled and open symbols, solid and dashed lines, respectively). The middle panel shows the corresponding ERB values (i.e., CF/Q_{ERB}), and the bottom panel shows the ratio between ERB or Q_{ERB} values under forward and simultaneous masking, with values greater than 1 indicating sharper tuning under forward masking than under simultaneous masking. Overall, Q_{ERB} values under forward masking increased with increasing frequency from around 4 at 200 Hz to around 13 at 1 kHz. The Q_{ERB} values under simultaneous masking followed a similar pattern but were lower (implying broader tuning) by a factor of between 1.5 and 3 across all signal frequencies.

At 200 Hz, signal durations of both 25 and 50 ms were tested under simultaneous masking. The 25-ms signal yielded a Q_{ERB} estimate that was not significantly different from that for the 50-ms signal [paired t-test on the log-transformed values: $t(7) = 0.69$, $p = 0.51$]. This null result suggests that the signal bandwidth did not significantly affect the estimates of filter tuning, confirming that the estimates from forward and simultaneous masking can be meaningfully compared, despite the difference in signal duration.

3.3. Combining results across studies to derive a general function for human cochlear tuning

3.3.1. Comparison of current and previous estimates of tuning at 1 kHz

A previous study (Oxenham and Shera, 2003) used stimuli similar to the ones employed here to estimate auditory filter tuning between 1 and 8 kHz. We compared the estimates from that earlier study with the current one at the single frequency they had in common (1 kHz). For the forward-masking conditions with the roex(p,w,t,p) filter ($N = 8$ and $N = 13$ for the previous and present study, respectively), there was no significant difference in estimated Q_{ERB} value between studies [independent-samples t-test on the log-transformed values: $t(19) = -1.4$, $p = 0.17$]. Similarly, for simultaneous masking ($N = 8$, pooling across the two signal levels tested in the previous study, and $N = 8$ in the present study), the difference was not significant [$t(14) = 1.94$, $p = 0.07$]. These comparisons confirm that there were no significant differences between the tuning estimates across the two studies, as expected given the similarity of stimuli and conditions.

3.3.2. A general function to relate tuning to center frequency

Since no significant differences in tuning estimates were observed between the present data and the earlier data of Oxenham and Shera (2003), the forward-masking data from the two studies were combined to provide estimates of tuning from 0.2 to 8 kHz. The combined data from the two studies are shown in the left panel of Fig. 6, in terms of geometric mean Q_{ERB} estimates, with the present data in black and the previous data in gray. Plotted with the data are the functions proposed by Oxenham and Shera (2003) for forward masking (solid curve) and the function proposed by Glasberg and Moore (1990) for simultaneous masking at moderate levels to describe filter tuning as a function of CF. The Oxenham and Shera function is a power law relating filter tuning (Q_{ERB}) to filter CF (F, in kHz) as follows:

$$Q_{ERB} = 11.1F^{0.27} \tag{4}$$

The Glasberg and Moore equation, originally for ERB, is rewritten here in terms of Q_{ERB}:

$$Q_{ERB} = \frac{1000F}{24.7(4.37F + 1)} \tag{5}$$

The Glasberg and Moore equation was devised to describe auditory filter tuning under simultaneous masking at moderate levels with a fixed masker level and variable signal level. Given that the present data were collected under forward masking with a variable masker level and a low fixed signal level, it is not surprising that the curve does not provide a good fit. The curve derived from Oxenham and Shera's study was obtained from forward-masking data, but only for CFs of 1 kHz and above. As expected, it provides a good fit to the data from that study, but fails to capture the more rapid decrease in tuning sharpness with decreasing CF below 1 kHz, particularly at the lowest CF of 200 Hz.

The more rapid decrease in the apical segment of the cochlea mirrors the pattern of stimulus-frequency otoacoustic emission de-

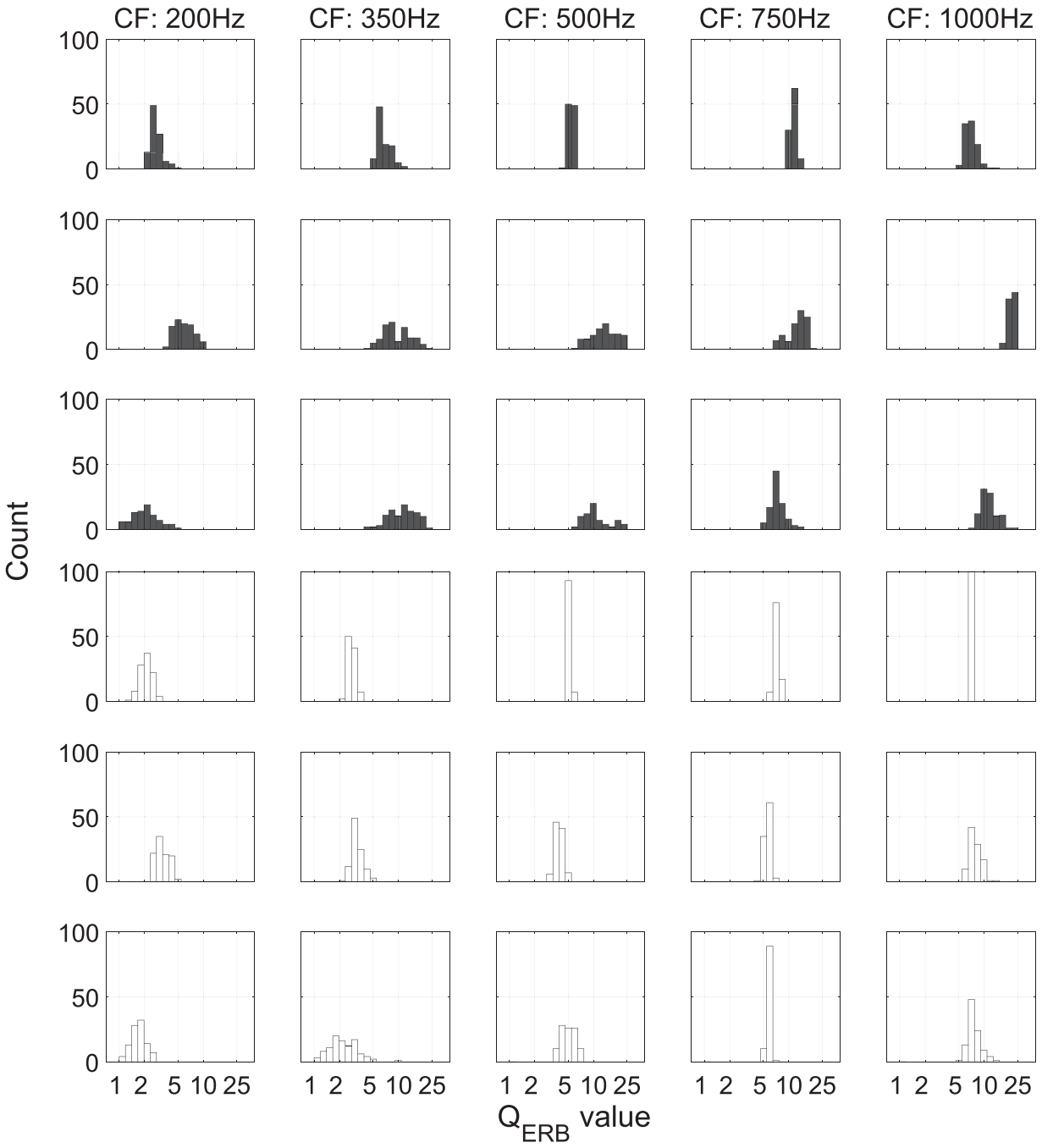

Fig. 4. Distributions of the Q_{ERB} estimates from the bootstrapped data from three sample listeners, representing low, medium, and high variability in the top, middle, and bottom rows, respectively, for forward masking (upper three rows, filled bars) and simultaneous masking (lower three rows, open bars). Each column corresponds to a different signal frequency. Each histogram represents the distribution obtained across 100 bootstrapped samples. For ease of visualization, rare Q_{ERB} values above 20 were allocated to the bin at $Q_{ERB} = 20$.

lays (Shera and Guinan, 2003) and can be captured with a two-segment power-law function, as described in the Discussion in the context of comparisons with other species (Section 4.2). However, inspection of Fig. 6 also suggests that the data could also be reasonably well described with a single logarithmic function across the entire frequency range of CFs tested here and in the previous study (Oxenham and Shera, 2003), as follows:

$$Q_{ERB} = \alpha \log_{10} F + \beta \quad (6)$$

The best-fitting function (in a least-squares sense) for the forward-masking Q_{ERB} estimates are shown in the right panel of Fig. 6, along with the data averaged from the present study and Oxenham and Shera (2003). The best-fitting function had parameter values (with 95% confidence intervals) of $\alpha = 8.40 \pm 1.35$ and $\beta = 10.67 \pm 0.74$, with a correlation coefficient of $r = 0.97$.

4. Discussion

4.1. Relationship between ERBs under forward and simultaneous masking

It has long been known that forward masking results in sharper tuning estimates than simultaneous masking (Glasberg and Moore, 1982; Moore, 1978; Vogten, 1978). This finding has been

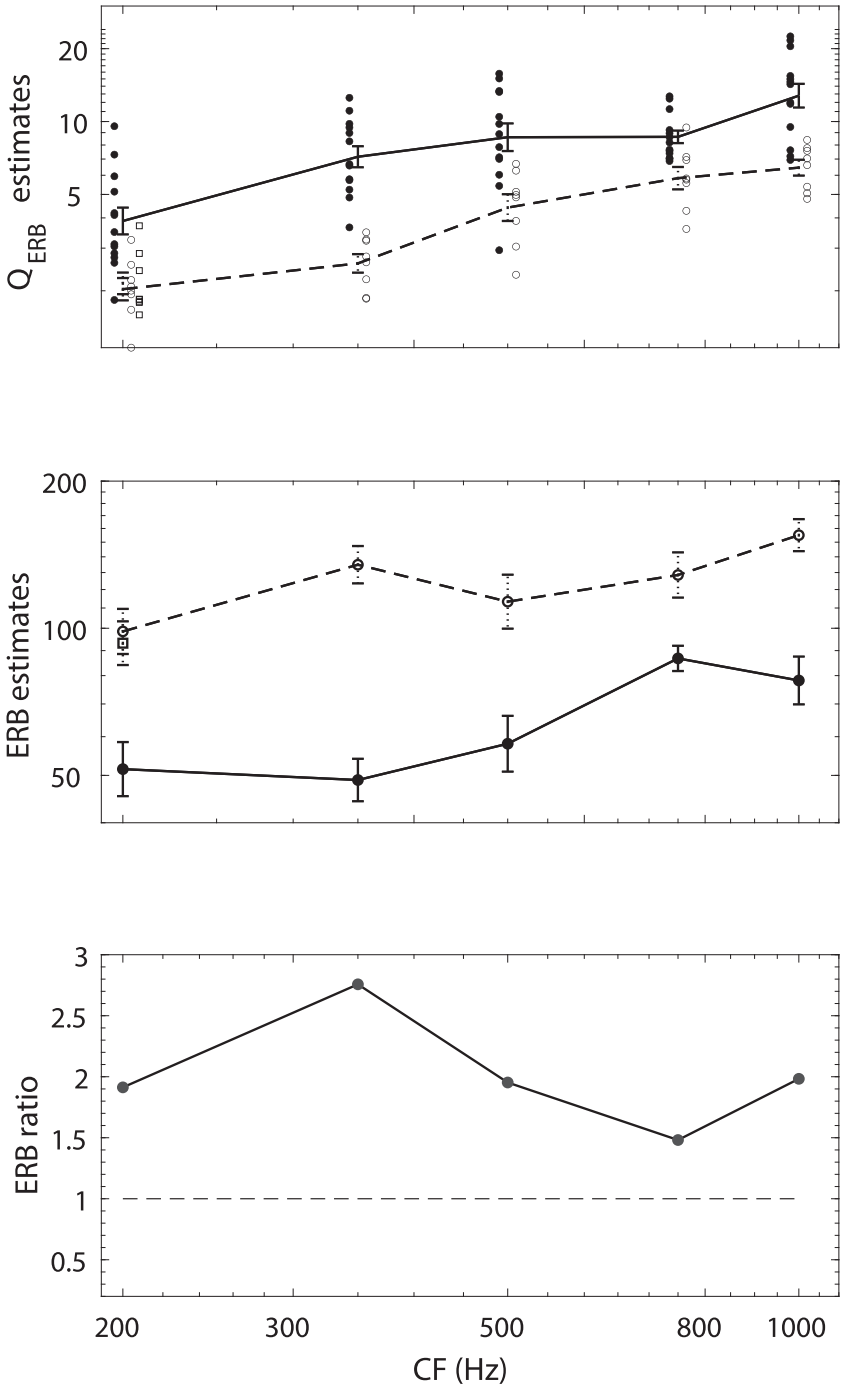

Fig. 5. Mean Q_{ERB} (top panel, along with individual values) and ERB (middle panel) estimates for the bootstrapped individual data under forward masking (filled circles, solid lines) and simultaneous masking (open circles, dashed lines). The values derived from the 200-Hz, 25-ms condition are shown as open squares. The bottom panel shows the ratio of the mean simultaneous-masking to forward-masking ERB values; values greater than 1 indicate sharper tuning under forward masking. Error bars represent ±1 SEM.

explained in terms of the underlying cochlear nonlinearities, with results under simultaneous masking reflecting the effects of both excitatory masking and nonlinear suppression effects, due to the interaction of the masker and signal, and results from forward masking reflecting just excitatory masking (e.g., Delgutte, 1990b; Oxenham and Plack, 1998). The results from forward masking are thus more likely to reflect those from direct neural tuning curves, which also avoid suppressive effects by presenting just a single stimulus at a time (Sumner et al., 2018). Although nonlinear interaction effects may occur within each of the noise maskers, any such effects do not seem to result in systematic discrepancies in overall tuning, as indicated by the good correspondence between auditory-nerve tuning curves and behavioral estimates using notched noise under forward masking in the ferret (Sumner et al., 2018).

The present results confirm that sharper tuning is observed under forward masking than under simultaneous masking at 1000 Hz (Moore, 1978; Oxenham and Shera, 2003), and extend that finding to frequencies as low as 200 Hz. Taken together with earlier data (Oxenham and Shera, 2003), our data suggest that the ratio of the Q_{ERB} under simultaneous and forward masking does not vary systematically across the CF range from 200 to 8000 Hz. In

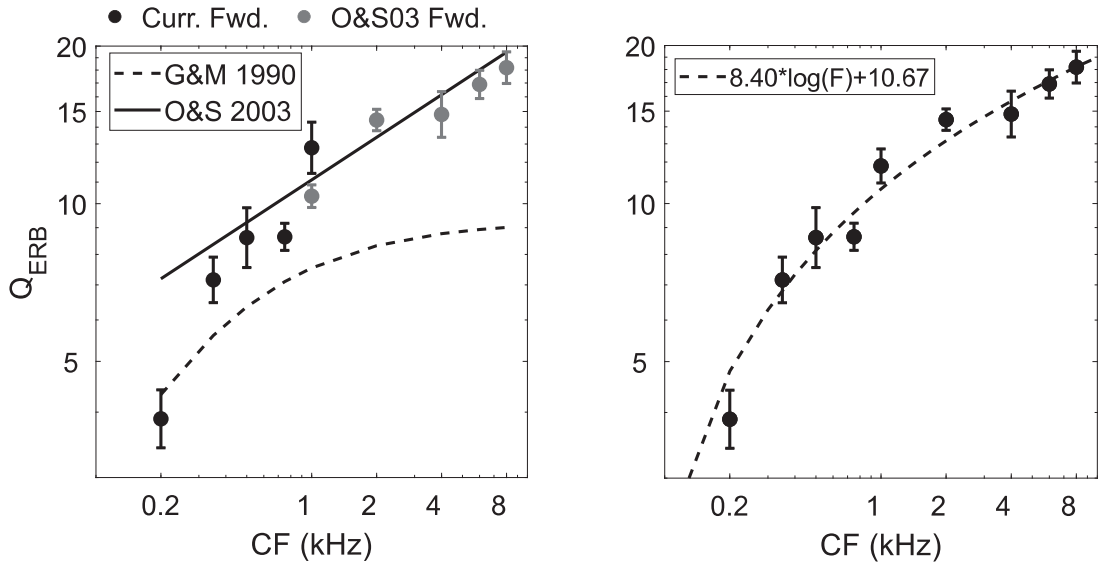

Fig. 6. Q_{ERB} estimates from forward masking compared with various curve fits. The left panel shows predictions of the Glasberg and Moore (1990) equation (dashed curve) and the Oxenham and Shera (2003) equation (solid line); data in black were obtained in the current study ($N = 13$), while data in gray were obtained in the Oxenham and Shera (2003) study (O&S03, $N = 8$). The right panel shows a logarithmic fit to tuning estimates combined from the present study (0.2–1 kHz) and from Oxenham and Shera (2003) (1–8 kHz). Data were averaged across studies at 1 kHz. Fitting equation is shown in the legend. Error bars represent ±1 SEM.

some ways this result is surprising, as it is generally thought that cochlear mechanics are altered in some ways in the apical regions of the cochlea for CFs below about 1000 Hz (Shera et al., 2010). Auditory-nerve two-tone suppression (Abbas and Sachs, 1976), which is generally thought to reflect nonlinear cochlear mechanics (Ruggero et al., 1992), is reduced or absent at CFs below about 1 kHz (e.g., Delgutte, 1990a), suggesting the possibility of reduced mechanical nonlinearities. In addition, more recent studies of apical cochlear regions have suggested less frequency-selective gain, a decoupling of basilar-membrane and auditory-nerve tuning curves (Recio-Spinoso and Oghalai, 2017), and generally different mechanical properties from those found in the base (Recio-Spinoso and Oghalai, 2018). Given the potential decoupling between basilar-membrane and auditory-nerve tuning at very low CFs, our behavioral method is more likely to reflect auditory-nerve than basilar-membrane tuning, as perception must be derived from auditory-nerve responses.

Although the relationship between ERB estimates under forward and simultaneous masking did not vary systematically across the range of CFs studied, our tuning estimates under simultaneous masking at the lowest CFs differ from estimates in earlier studies. For instance at 200 Hz our average observed Q_{ERB} was around 2 (ERB ≈ 100 Hz), whereas the Glasberg and Moore (1990) function's Q_{ERB} is 4 (ERB = 50 Hz). Glasberg and Moore's function was based on data with maskers presented at a fixed moderate level, as opposed to a low-level fixed signal. However, other data obtained with more comparable fixed low-level signals also suggest sharper tuning than we observed (Baker et al., 1998; Jurado et al., 2011). This apparent discrepancy warrants further study, as it remains unclear how it can be explained. In any case, our focus was primarily on the results under forward masking, as those have been shown to provide more accurate estimates of the underlying auditory-nerve tuning (Sumner et al., 2018).

4.2. Comparison with auditory-nerve tuning curves in other species

4.2.1. Bandwidths

Previous studies comparing cochlear tuning across species have suggested that there may be a different dependence of tuning on CF in the apical and basal regions (e.g., Shera et al., 2010; Temchin et al., 2008a). This is illustrated in Fig. 7. In addition to the human data (based on our forward-masking results), Fig. 7 shows estimates derived from auditory-nerve recordings in other mammals (cat, macaque, chinchilla, guinea pig, and ferret). Although human tuning is significantly sharper overall (e.g., by a factor of 2–3 compared to cat), its variation with CF is similar to that found in the other animals. In all species shown, Q_{ERB} increases monotonically with CF and typically increases more rapidly at low than at high CFs. The data in other species have also been approximated by a two-segment power law, indicating

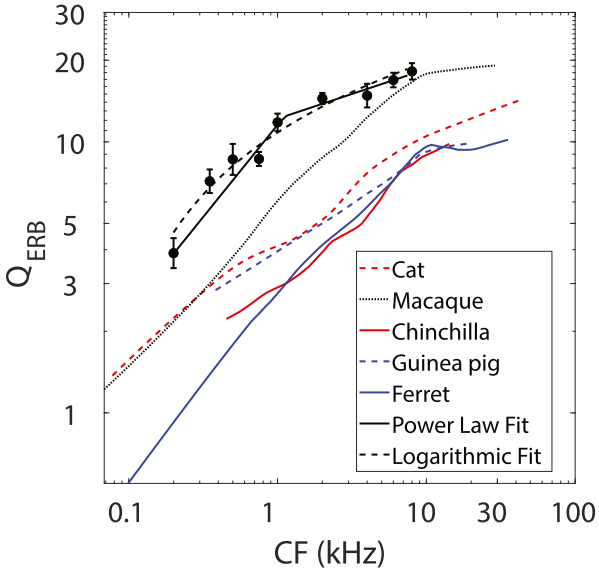

Fig. 7. Human tuning compared to auditory-nerve tuning in other mammals. Solid black curves show the segmented power-law fits to the human data; dashed black curves replot the logarithmic fit shown in the right panel of Fig. 6. The power-law exponent (slope of the solid black lines) changes near 1 kHz, decreasing from a value of 0.7 at low frequencies to 0.2 at higher frequencies. Other curves represent data from different species, as shown in the legend. Animal data are taken from Shera et al. (2010), Joris et al. (2011), and Sumner et al. (2018).

that they have low- and high-frequency asymptotes of different but roughly constant slope. The black solid lines in Fig. 7 represent a segmented power-law fit to the human data; the slope changes near 1 kHz, consistent with a similar change in the frequency dependence of otoacoustic-emission delay. The power-law slope decreases from a value of about 0.7 in the apical region of the cochlea (low CFs) to about 0.2 in the base. For comparison, the dashed black curves replot the logarithmic function used in the right panel of Fig. 6. Although both functions provide a good description of the data between 0.2 and 8 kHz, the segmented power-law function may generalize better across species and reflects an apical-basal transition, generally most prominent in otoacoustic-emissions data, with a location that varies across species (Joris et al., 2011; Shera et al., 2010; Sumner et al., 2018).

4.2.2. Filter shapes and asymmetries

There is empirical support for the idea that tuning estimates from forward-masked notched-noise behavioral thresholds correspond well to those from auditory-nerve tuning curves in the same species (e.g., Sumner et al., 2018). However, no studies have compared estimates of filter shape or asymmetry across these two methods. The roex filter shape assumed by our filter-fitting procedure has some properties that can capture the shape of auditory-nerve tuning curves observed at high CFs (> 1 kHz), such as a sharp tip and a shallow low-frequency tail. The filter shape itself is, however, flexible, in that it can accommodate asymmetries in either direction and, if the tail factor is close to 1 or the weighting of the tail is low, the tail can effectively disappear (e.g., see simultaneous-masking curves below 500 Hz in Fig. 3).

Aside from a different dependence of auditory-nerve tuning on CF in the apex and base of the cochlea (e.g., Temchin et al., 2008b), other aspects of the tuning curve are also different. For instance, the well-known shallow tail of tuning curves at high CFs is typically absent at CFs below 2-3 kHz. Also, the asymmetry of the high-CF tuning curves, with a steeper upper than lower slope, is absent around 1 kHz, and may reverse at CFs lower than 1 kHz (e.g., Temchin et al., 2008a). The reversal in asymmetry of auditory-nerve tuning curves below about 1 kHz is also observed in the phase responses of the auditory-nerve fibers in cat, which show an upward glide in the instantaneous frequency of the impulse response for CFs above 1 kHz, but a downward glide for CFs below 1 kHz (Carney et al., 1999).

Qualitatively similar patterns can be observed in our human behavioral estimates with forward masking, but with an apparently lower transition CF: At 1 kHz, the asymmetry and tail of the filter remain apparent, but the asymmetry is reduced or absent at CFs lower than 750 Hz, and the tail is less prominent at the lowest CF of 200 Hz (Figs. 1 and 3). Interestingly, a similar pattern of decreasing asymmetry with decreasing CF has also been observed in human masking studies using linear frequency sweeps as maskers, where an asymmetry in filtering consistent with a rising instantaneous frequency sweep in the impulse response has been found at frequencies of 1 kHz and higher, but a symmetric response has been found at lower frequencies of 125 and 250 Hz (Oxenham and Dau, 2001).

Overall, the dependence of filter asymmetry and prominence of the filter tail on CF in the present estimates derived from forward-masking data are qualitatively in line with existing auditory-nerve data in other species. However, as with estimates of bandwidth (Shera et al., 2010), the transition between basal and apical properties seems to occur at a lower CF (around 1 kHz) in humans than in several other species (e.g., around 3-4 kHz in smaller mammals such as cat, guinea pig, and chinchilla).

4.3. A new function for describing human cochlear tuning across CF

The primary novel finding here is that behavioral estimates of cochlear tuning (from forward masking) become considerably broader below 1 kHz, with the estimated Q_{ERB} of around 12 at 1 kHz decreasing to around 4 at 200 Hz – a factor of about 3. Poorer tuning at low CFs is predicted by Glasberg and Moore's (1990) function, but not to same extent, with a change in predicted Q_{ERB} across the same frequency range of a factor of about 1.7. The transition in tuning dependence on CF in humans around 1 kHz is consistent with the predictions from human otoacoustic-emissions data by Shera et al. (2010), and can be captured by a segmented power-law function with an exponent of 0.7 in the apical region (< 1 kHz) and an exponent of around 0.2 in the basal region (> 1 kHz). Alternatively, a logarithmic function can describe the Q_{ERB} estimates derived from forward masking as a function of F (in kHz) between 0.2 and 8 kHz, as follows:

$$Q_{ERB} = 8.40\log_{10}F + 10.67 \quad (7)$$

The segmented power-law function, as shown in Fig. 7, provides a similarly good fit over this range, and may be more appropriate for very low frequencies, where the logarithmic function decreases rapidly (reaching zero at 18.6 Hz, just below the traditional lower limit of human hearing). Tuning estimates at 100 Hz and below have been obtained using simultaneous masking with either a fixed-masker level and notched noise (Jurado and Moore, 2010) or with psychophysical tuning curves (Jurado et al., 2011). Both produced a range of estimated ERB values spanning 25 to 50 Hz for CFs between 31.5 and 100 Hz, with no clear increasing or decreasing trend in that range, implying Q_{ERB} values from around 4 at 100 Hz to 0.6 at 31.5 Hz. To our knowledge, no such estimates exist using forward masking. Another benefit of the segmented power-law function is that it provides a more ready comparison to available auditory-nerve tuning estimates in other species (Fig. 7). Finally, in cases where estimates of effective frequency selectivity in humans are required under higher sound levels and under simultaneous masking, the function of Glasberg and Moore (1990) may remain more appropriate.

Funding sources

This work was supported by National Institutes of Health grants R01 DC012262 (AJO) and R01 DC003687 (CAS) and by NSF graduate training grant NRT-UtB 1734815 (supporting GRO).

CRediT authorship contribution statement

John Leschke: Conceptualization, Methodology, Investigation. **Gerardo Rodriguez Orellana:** Investigation, Formal analysis, Visualization, Software, Writing – review & editing. **Christopher A. Shera:** Conceptualization, Writing – review & editing, Visualization. **Andrew J. Oxenham:** Conceptualization, Methodology, Software, Writing – original draft, Supervision, Funding acquisition.

Acknowledgments

Enrique Lopez-Poveda, Brian C. J. Moore, and the Associate Editor, Laurel Carney, provided many helpful comments and suggestions during the review process.

References

Abbas, P.J., Sachs, M.B., 1976. Two-tone suppression in auditory-nerve fibers: extension of a stimulus-response relationship. J. Acoust. Soc. Am. 59, 112–122.
Allen, E.J., Mesik, J., Kay, K.N., Oxenham, A.J., 2022. Distinct representations of tonotopy and pitch in human auditory cortex. J. Neurosci. 42, 416–434.

Baker, R.J., Rosen, S., Darling, A.M., 1998. An efficient characterisation of human auditory filtering across level and frequency that is also physiologically reasonable. In: Palmer, A.R., Rees, A., Summerfield, A.Q., Meddis, R. (Eds.), Psychophysical and Physiological Advances in Hearing. Whurr, London, pp. 81–87.
Carney, L.H., McDuffy, M.J., Shekhter, I., 1999. Frequency glides in the impulse responses of auditory-nerve fibers. J. Acoust. Soc. Am. 105, 2384–2391.
Cooper, N.P., Rhode, W.S., 1995. Nonlinear mechanics at the apex of the guinea-pig cochlea. Hear. Res. 82, 225–243.
De Martino, F., Moerel, M., van de Moortele, P.F., Ugurbil, K., Goebel, R., Yacoub, E., Formisano, E., 2013. Spatial organization of frequency preference and selectivity in the human inferior colliculus. Nat. Commun. 4, 1386.
Delgutte, B., 1990a. Two-tone suppression in auditory-nerve fibers: dependence on suppressor frequency and level. Hear. Res. 49, 225–246.
Delgutte, B., 1990b. Physiological mechanisms of psychophysical masking: observations from auditory-nerve fibers. J. Acoust. Soc. Am. 87, 791–809.
Eustaquio-Martin, A., Lopez-Poveda, E.A., 2011. Isoresponse versus isoinput estimates of cochlear filter tuning. J. Assoc. Res. Otolaryngol. 12, 281–299.
Ewert, S., 2013. AFC: a modular framework for running psychoacoustic experiments and computational perception models. In: Proceedings of the International Conference on Acoustics AIA-DAGA, Merano, Italy, pp. 1326–1329.
Fletcher, H., 1940. Auditory patterns. Rev. Mod. Phys. 12, 47–65.
Formisano, E., Kim, D.S., Di Salle, F., van de Moortele, P.F., Ugurbil, K., Goebel, R., 2003. Mirror-symmetric tonotopic maps in human primary auditory cortex. Neuron 40, 859–869.
Glasberg, B.R., Moore, B.C.J., 1982. Auditory filter shapes in forward masking as a function of level. J. Acoust. Soc. Am. 71, 946–949.
Glasberg, B.R., Moore, B.C.J., 1990. Derivation of auditory filter shapes from notched-noise data. Hear. Res. 47, 103–138.
Glasberg, B.R., Moore, B.C.J., 2000. Frequency selectivity as a function of level and frequency measured with uniformly exciting notched noise. J. Acoust. Soc. Am. 108, 2318–2328.
Helmholtz, H.L.F., 1885/1954. On the Sensations of Tone. Dover, New York.
Houtgast, T., 1972. Psychophysical evidence for lateral inhibition in hearing. J. Acoust. Soc. Am. 51, 1885–1894.
Houtgast, T., 1973. Psychophysical experiments on "tuning curves" and "two-tone inhibition". Acustica 29, 168–179.
Joris, P., Bergevin, C., Kalluri, R., McLaughlin, M., Michelet, P., Van der Heijden, M., Shera, C.A., 2011. Frequency selectivity in old-world monkeys corroborates sharp cochlear tuning in humans. Proc. Natl. Acad. Sci. U. S. A. 108, 17516–17520.
Jurado, C., Moore, B.C.J., 2010. Frequency selectivity for frequencies below 100 Hz: comparisons with mid-frequencies. J. Acoust. Soc. Am. 128, 3585–3596.
Jurado, C., Pedersen, C.S., Moore, B.C.J., 2011. Psychophysical tuning curves for frequencies below 100Hz. J. Acoust. Soc. Am. 129, 3166–3180.
Kiang, N.Y., Sachs, M.B., Peake, W.T., 1967. Shapes of tuning curves for single auditory-nerve fibers. J. Acoust. Soc. Am. 42, 1341–1342.
Levitt, H., 1971. Transformed up-down methods in psychoacoustics. J. Acoust. Soc. Am. 49, 467–477.
Lopez-Poveda, E.A., Eustaquio-Martin, A., 2013. On the controversy about the sharpness of human cochlear tuning. J. Assoc. Res. Otolaryngol. 14, 673–686.
Moerel, M., De Martino, F., Formisano, E., 2012. Processing of natural sounds in human auditory cortex: tonotopy, spectral tuning, and relation to voice sensitivity. J. Neurosci. 32, 14205–14216.
Moore, B.C.J., 1978. Psychophysical tuning curves measured in simultaneous and forward masking. J. Acoust. Soc. Am. 63, 524–532.
Moore, B.C.J., Peters, R.W., Glasberg, B.R., 1990. Auditory filter shapes at low center frequencies. J. Acoust. Soc. Am. 88, 132–140.
Moore, B.C.J., Glasberg, B.R., Baer, T., 1997. A model for the prediction of thresholds, loudness, and partial loudness. J. Aud. Eng. Soc. 45, 224–240.
Narayan, S.S., Temchin, A.N., Recio, A., Ruggero, M.A., 1998. Frequency tuning of basilar membrane and auditory nerve fibers in the same cochleae. Science 282, 1882–1884.
Oxenham, A.J., Plack, C.J., 1998. Suppression and the upward spread of masking. J. Acoust. Soc. Am. 104, 3500–3510.
Oxenham, A.J., Dau, T., 2001. Towards a measure of auditory-filter phase response. J. Acoust. Soc. Am. 110, 3169–3178.
Oxenham, A.J., Shera, C.A., 2003. Estimates of human cochlear tuning at low levels using forward and simultaneous masking. J. Assoc. Res. Otolaryngol. 4, 541–554.
Oxenham, A.J., Simonson, A.M., 2006. Level dependence of auditory filters in nonsimultaneous masking as a function of frequency. J. Acoust. Soc. Am. 119, 444–453.
Patterson, R.D., 1974. Auditory filter shape. J. Acoust. Soc. Am. 55, 802–809.
Patterson, R.D., 1976. Auditory filter shapes derived with noise stimuli. J. Acoust. Soc. Am. 59, 640–654.
Patterson, R.D., Nimmo-Smith, I., 1980. Off-frequency listening and auditory filter asymmetry. J. Acoust. Soc. Am. 67, 229–245.
Plomp, R., 1964. The ear as a frequency analyzer. J. Acoust. Soc. Am. 36, 1628–1636.
Puria, S., Peake, W.T., Rosowski, J.J., 1997. Sound-pressure measurements in the cochlear vestibule of human-cadaver ears. J. Acoust. Soc. Am. 101, 2754–2770.
Reale, R.A., Imig, T.J., 1980. Tonotopic organization in auditory cortex of the cat. J. Comp. Neurol. 192, 265–291.
Recio-Spinoso, A., Oghalai, J.S., 2017. Mechanical tuning and amplification within the apex of the guinea pig cochlea. J. Physiol. 595, 4549–4561.
Recio-Spinoso, A., Oghalai, J.S., 2018. Unusual mechanical processing of sounds at the apex of the Guinea pig cochlea. Hear. Res. 370, 84–93.
Rhode, W.S., Cooper, N.P., 1993. Two-tone suppression and distortion production on the basilar membrane in the hook region of the cat and guinea pig cochleae. Hear. Res. 66, 31–45.
Rosen, S., Baker, R.J., Darling, A., 1998. Auditory filter nonlinearity at 2kHz in normal hearing listeners. J. Acoust. Soc. Am. 103, 2539–2550.
Ruggero, M.A., Robles, L., Rich, N.C., 1992. Two-tone suppression in the basilar membrane of the cochlea: mechanical basis of auditory-nerve rate suppression. J. Neurophysiol. 68, 1087–1099.
Russell, I.J., Nilsen, K.E., 1997. The location of the cochlear amplifier: spatial representation of a single tone on the guinea pig basilar membrane. Proc. Nat. Acad. Sci. 94, 2660–2664.
Sachs, M.B., Kiang, N.Y.S., 1968. Two-tone inhibition in auditory nerve fibers. J. Acoust. Soc. Am. 43, 1120–1128.
Scharf, B., 1970. Critical bands. In: Tobias, J.V. (Ed.), Foundations of Modern Auditory Theory. Academic, New York.
Schreiner, C.E., Langner, G., 1988. Periodicity coding in the inferior colliculus of the cat. II. Topographical organization. J. Neurophysiol. 60, 1823–1840.
Shannon, R.V., 1976. Two-tone unmasking and suppression in a forward masking situation. J. Acoust. Soc. Am. 59, 1460–1470.
Shera, C.A., Guinan, J.J., 2003. Stimulus-frequency-emission group delay: a test of coherent reflection filtering and a window on cochlear tuning. J. Acoust. Soc. Am. 113, 2762–2772.
Shera, C.A., Guinan, J.J., Oxenham, A.J., 2002. Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. Proc. Natl. Acad. Sci. U.S.A. 99, 3318–3323.
Shera, C.A., Guinan, J.J., Oxenham, A.J., 2010. Otoacoustic estimation of cochlear tuning: validation in the chinchilla. J. Assoc. Res. Otolaryngol. 11, 343–365.
Sumner, C.J., Wells, T.T., Bergevin, C., Sollini, J., Kreft, H.A., Palmer, A.R., Oxenham, A.J., Shera, C.A., 2018. Mammalian behavior and physiology converge to confirm sharper cochlear tuning in humans. Proc. Natl. Acad. Sci. U.S.A. 115, 11322–11326.
Temchin, A.N., Rich, N.C., Ruggero, M.A., 2008a. Threshold tuning curves of chinchilla auditory-nerve fibers. I. Dependence on characteristic frequency and relation to the magnitudes of cochlear vibrations. J. Neurophysiol. 100, 2889–2898.
Temchin, A.N., Rich, N.C., Ruggero, M.A., 2008b. Threshold tuning curves of chinchilla auditory nerve fibers. II. Dependence on spontaneous activity and relation to cochlear nonlinearity. J. Neurophysiol. 100, 2899–2906.
Verschuure, J., 1981a. Pulsation patterns and nonlinearity of auditory tuning. II. Analysis of psychophysical results. Acustica 49, 296–306.
Verschuure, J., 1981b. Pulsation patterns and nonlinearity of auditory tuning. I. Psychophysical results. Acustica 49, 288–295.
Versteegh, C.P., Meenderink, S.W., van der Heijden, M., 2011. Response characteristics in the apex of the gerbil cochlea studied through auditory nerve recordings. J. Assoc. Res. Otolaryngol. 12, 301–316.
Vogten, L.L.M., 1978. Low-level pure-tone masking: a comparison of "tuning curves" obtained with simultaneous and forward masking. J. Acoust. Soc. Am. 63, 1520–1527.
von Békésy, G., 1960. Experiments in Hearing. McGraw-Hill, New York.
Wegel, R.L., Lane, C.E., 1924. The auditory masking of one sound by another and its probable relation to the dynamics of the inner ear. Phys. Rev. 23, 266–285.
Zwicker, E., 1961. Subdivision of the audible frequency range into critical bands (Frequenzgruppen). J. Acoust. Soc. Am. 33, 248.
Zwicker, E., Flottorp, G., Stevens, S.S., 1957. Critical bandwidth in loudness summation. J. Acoust. Soc. Am. 29, 548–557.