Title (89 char with spaces).

Mammalian behavior and physiology converge to confirm sharper cochlear tuning in humans

Short title (48 char with spaces; max 50):

Convergent measures of mammalian cochlear tuning

Authors

Christian J. Sumner[1]*, Toby T. Wells[1], Christopher Bergevin[2], Joseph Sollini[1+], Heather A. Kreft[3], Alan R. Palmer[1], Andrew J. Oxenham[3], Christopher A. Shera[4].

Affiliations

[1]Medical Research Council Institute of Hearing Research, School of Medicine, The University of Nottingham, University Park, Nottingham, NG7 2RD, UK

[2] Dept. of Physics & Astronomy, Centre for Vision Research, York University, Toronto, ON Canada

[3]Department of Psychology and Department of Otolaryngology, University of Minnesota, 75 East River Road, Minneapolis, MN 55455, U.S.A.

[4]Caruso Department of Otolaryngology and Department of Physics and Astronomy; University of Southern California, Los Angeles, CA, U.S.A.

[+]Current address: UCL Ear Institute, Grays Inn Road, London, UK

* Corresponding author.

Orchid IDs:

Sumner: 0000-0002-2573-7418; Bergevin: 0000-0002-4529-399X; Sollini: 0000-0002-1974-4291; Palmer: 0000-0001-5893-4525; Oxenham: 0000-0002-9365-1157; Shera: 0000-0002-5939-2710

Classification:

BIOLOGICAL SCIENCES; Neuroscience; Psychological and Cognitive Science.

Figures

3 in main text. 3 Supplementary Figures.

Keywords

cochlear tuning, auditory nerve, psychoacoustics, otoacoustic emissions

Abstract (184 words; max 250).

Frequency analysis of sound by the cochlea is the most fundamental property of the auditory system. Despite its importance, the resolution of this frequency analysis in humans remains controversial. The controversy persists because the methods used to estimate tuning in humans are indirect and have not all been independently validated in other species. Some data suggest that human cochlear tuning is considerably sharper than that of laboratory animals, while others suggest little or no difference between species. We show here in a single species (ferret) that behavioral estimates of tuning bandwidths obtained using perceptual masking methods, and objective estimates obtained using otoacoustic emissions, both also employed in humans, agree closely with direct physiological measurements from single auditory-nerve fibers. Combined with new human behavioral data, this outcome indicates that the frequency analysis performed by the human cochlea is of significantly higher resolution than found in common laboratory animals. This finding raises important questions about the evolutionary origins of human cochlear tuning, its role in the emergence of speech communication, and the mechanisms underlying our ability to separate and process natural sounds in complex acoustic environments.

Significance statement (max 150 words)

Sound consists of a dynamic stream of energy at different frequencies. Auditory processing of sound frequency is critical in determining our ability to interact and communicate in a complex acoustic world, yet fundamental gaps remain in our understanding of how this is achieved. Indeed, the resolving power of the system, how best to measure it, and the mechanisms that underlie it are all still debated. Here we provide critical evidence demonstrating that humans can resolve the frequency components of competing sounds better than other commonly studied mammals. This finding raises important questions both for theories of auditory perception and for our understanding of the evolutionary relationships between the auditory system and acoustic communication, including speech.

\body

Introduction

The cochlea within the inner ear acts like an acoustic prism to decompose sound into its constituent frequency components, creating a frequency-to-place map along its length. This decomposition establishes the tonotopic encoding of sound frequency that remains a fundamental organizing principle of the auditory system from the cochlea to the auditory cortex (1-4). The resolution with which the cochlea performs this frequency analysis influences our ability to perceptually separate different sounds and to communicate in complex acoustic environments. The loss of cochlear frequency resolution, through damage or disease, underlies some of the most troublesome problems associated with hearing impairment, including difficulty understanding speech in noise (5).

For many years a consensus existed that cochlear tuning was similar across a wide range of mammalian species, including humans. That conclusion was based on the relatively good correspondence between indirect behavioral estimates of human tuning (6, 7) and direct measures of cochlear tuning taken from the auditory nerve of smaller laboratory animals (8, 9). Very few physiological human data existed, and those that did were not sufficient in number or did not deviate sufficiently from animal data to suggest any fundamental differences between species (10). However, more recent studies have suggested that human cochlear tuning may be sharper, by a factor of two or more, than cochlear tuning in typical laboratory animals, such as cat and guinea pig. The latest estimates from humans combined more refined behavioral measures and new non-invasive objective measures based on otoacoustic emissions (OAEs)—sounds that are emitted by the cochlea and can be recorded in the ear canal (11).

Knowledge of any interspecies differences in the frequency resolution of the cochlea is critical to our understanding of a diverse range of issues (12). For example, the claimed disparities in estimates between animal and human tuning are sufficiently large to substantially affect the

neural coding and representation of speech and other critical natural sounds (13-15). Quantification of species differences is also important for understanding the mechanisms underlying frequency analysis. For instance, it has been claimed that the cortical representation of frequency results from neural sharpening by the central auditory system from a less sharply tuned representation in the cochlea (16). This claim hinges critically on the assumption that human cochlear tuning is similar to that of small mammals.

In large part, claims of sharper tuning in the human cochlea remain controversial (17-19) because of a lack of commensurate measures across species. Direct measures of tuning from single-unit recordings in the auditory nerve (ANF in Fig. 1) have been obtained in laboratory animals, but are too invasive to be performed in humans. Conversely, the more recent psychophysical methods (PSY) used in humans, involving the masking of a probe tone by spectrally notched noise under forward masking (PSY-F; Fig. 1) have not yet been tested in animals. Estimates based on OAE measurements have been obtained in both humans and smaller mammals, and are consistent with the claim of sharper tuning in humans (11, 18). However, uncertainty surrounding the mechanisms by which OAEs are generated, and their relationship to cochlear tuning, leave room for doubt (20, 21). In summary, three types of measure have been used to estimate cochlear tuning—behavioral, otoacoustic, and neural—but have never all been measured and compared in the same species. To resolve this problem, we used ferrets to examine all three measures within the same species. We reasoned that if the two indirect measures (OAE and PSY) provide accurate estimates of cochlear tuning, then they should both agree with the direct neural (ANF) measures. By employing all three methods in the same species, our experiments provide the strongest test to date of the validity of the indirect measures used to assess cochlear frequency tuning in humans.

Results

We estimated ferret frequency tuning perceptually using a psychophysical notched-noise masking paradigm (Fig. 1 PSY; Supplementary Fig. S1). This paradigm measures the effectiveness of noises with various spectral shapes at masking a narrowband signal, such as a pure tone. By varying the frequency extent of a spectral notch in the masking noise, the shape and bandwidth of the effective auditory filter can be derived (see Supplementary Methods). We applied this method in ferrets performing behavioral detection tasks, and from the results derived the equivalent rectangular bandwidths (ERBs)—and a corresponding dimensionless measure of tuning sharpness, Q_{ERB} (center frequency/ERB)—of the filters. For any filter shape, the ERB is the bandwidth of the rectangular filter with the same peak height that passes the same total power.

Because of cochlear nonlinearities, the exact stimulus conditions employed can influence the measured bandwidths. These include whether the masking noise is presented simultaneously with the signal (PSY-S) or directly precedes the signal (PSY-F), thereby avoiding physical interactions between the stimuli within the cochlea (22-24). The estimated bandwidths can also depend on whether the intensity of the tone is kept constant and the threshold is found by varying the intensity of the masker, or *vice versa*. We estimated filter bandwidths in ferrets using all of these variants. Consistent with results in humans (22-24), we observed that forward-masking (PSY-F) produces significantly sharper estimates of tuning than simultaneous masking (Q_{ERB}(PSY-S) = 0.72 x Q_{ERB}(PSY-F); p=0.04; see Fig. 2b, 3 and Supplementary Fig. S3). We found no significant effect of whether thresholds are derived by varying the level of the masker or target tone (p=0.2), contrary to expectations (19, 25, 26). The absence of a significant effect may be partly due to our use of low stimulus levels (< 40 dB SPL), which are generally below the onset level of the compressive cochlear non-linearity in ferrets (27), and partly due to the relatively small number of estimates in each condition (n = 5 for the fixed signal and n = 3 for the fixed masker), providing limited statistical power to detect a difference. Therefore, we only distinguish between forward and simultaneous masking in our further comparisons.

Next, we recorded stimulus-frequency otoacoustic emissions (SFOAEs) from the ears of sedated ferrets and inferred cochlear bandwidths using the emission group delay (Fig. 1; OAE, Supplementary Methods and Supplementary Fig. S2). The OAE-based method estimates the sharpness (Q_{ERB}) of the cochlear filters using the assumption of approximate species-invariance of the "tuning ratio". The tuning ratio is the empirical relationship between emission-delay and auditory-nerve-fiber tuning trends obtained from independent measurements in other species. To estimate the ferret Q_{ERB} trend from the SFOAE delays, we followed Joris et al (28) and used a tuning ratio obtained by averaging those previously derived for cats, guinea pigs, and chinchillas—species whose tuning ratios are all similar (18). Figure 2a shows the trend of auditory filter sharpness inferred from the emission delays (data points are shown in Supplementary Fig. S2).

Finally, we compared the estimates from the two indirect measures with our previously published responses of single auditory-nerve fibers in anesthetized ferrets to short (50 ms) tone pips varying in frequency and sound level (27). The spike counts in response to these tones allowed us to map out the receptive field of each fiber (see Fig. 1; ANF) (i.e., the range of stimulus conditions over which the nerve fibers responded). From the lowest (threshold; Fig. 1, ANF; grey line) sound level that produced a response at each frequency we modeled the shape of the auditory filter in each nerve fiber by fitting a rounded-exponential function (Fig. 1; ANF; brown line, 29), and derived its Q_{ERB}, in the same manner as was done with the behavioral estimates.

Figure 2 shows that all three measures of Q_{ERB}—those derived from auditory-nerve responses (ANF), from otoacoustic emissions (OAEs), and from psychophysical forward-masking (PSY-F)—are in good agreement. The agreement includes both the overall sharpness of tuning as well as its approximate power-law dependence on frequency. The agreement is especially remarkable given the very different natures of the three measures employed.

To compare the measurements quantitatively, we fitted the data (log-transformed frequency and Q_{ERB}) with a linear model. With respect to overall tuning sharpness, the agreement among the different measures is most apparent when the data are expressed relative to the mean auditory-nerve tuning at the same frequency (i.e., residuals of the linear model; Fig. 3). Although the mean OAE-based estimates of Q_{ERB} are similar to those obtained directly from auditory-nerve tuning curves, their ratio is less than unity (Q_{ERB}(OAE) = 0.82 x Q_{ERB}(ANF) ; Fig. 3), and this difference is statistically significant (sandwich-test, p<0.001; see Supplementary Methods), in part due to the very large sample size of the OAE data (n~1500). The difference in means implies that the tuning ratio in ferrets derived from these data is somewhat larger than the average of those previously obtained for cat, guinea pig, and chinchilla. For comparison, the variation among the tuning ratios for these three species is shown in Fig. 9B of reference (16); the approximate "invariance" of the tuning ratio typically holds to within 5-15%, with the largest variations occurring in the apical regions of the cochlea.

Consistent with findings in humans, psychophysical estimates of tuning using simultaneous masking (PSY-S) are significantly broader (Q_{ERB}(PSY-S) = 0.72 x Q_{ERB}(PSY-F); Fig. 3) than the tuning estimates derived from both auditory-nerve fiber responses and OAEs (sandwich test, p<0.01; Cohen's d ~1). In order to adapt the behavioral experiments to animal use, we necessarily modified some procedures used in previous human experiments. To explore the possible effects of these modifications, we tested a new set of human listeners using methods (stimuli and task) directly comparable to those used in our ferret experiments, with forward masking and a fixed target level (see Supplementary Methods). The estimated Q_{ERB} at 4 kHz obtained using these ferret-based procedures with humans is similar to that found in earlier human studies (22), and is more than a factor of 2 sharper than the behavioral estimates from ferrets (p<0.001; Supplementary Fig. S3).

Discussion

Disparate methods for measuring cochlea tuning were employed in a single animal model. Both psychophysical and otoacoustic methods provided reliable and quantitatively accurate estimates of cochlear frequency selectivity. These direct and indirect measures combined with new human behavioral data, collected using the same methods, provide strong support for the claim that frequency resolution is sharper in humans than in common laboratory mammals (summarized in Supplementary Fig. 3)

We attribute the close correspondence in tuning measures in large part to the refined methods employed in this study and their application within a single species. However, some modest discrepancies remain that are important to address. Tuning estimates obtained here using simultaneous masking are broader than those from ANF and forward-masked methods, consistent with studies in humans (22) and macaque (28, 30). However, other published data suggest either a closer correspondence of simultaneous masking and auditory nerve tuning (31) or even little difference compared to humans (32). Our data also fail to reveal the expected difference in frequency selectivity depending on whether the signal or masker were varied to determine thresholds (19, 25, 26). These inconsistencies may point to species differences other than tuning bandwidth, such as differences in the nature and extent of cochlear nonlinearities or cognition (33). However, the sizes of any differences are not large in comparison with the variability of the data (for example, of individual nerve fibers or of individual animals). A comprehensive assessment in non-human mammals of the effects of iso-level (fixed-masker) vs. iso-response (fixed-signal) measurements, forward vs. simultaneous masking, and overall sound level, with larger numbers of measurements, is required to resolve these issues.

The agreement of the three tuning measures provides compelling evidence that the limits of perceptual frequency resolution (as measured in our paradigm) are determined primarily in the cochlea, in contrast to previous suggestions (16). This conclusion therefore warrants a fresh evaluation of spectral decomposition in the central auditory system. In some cases, this

agreement could obviate the need to postulate additional neural sharpening mechanisms, located between the cochlea and the cortex, to explain previously presumed discrepancies between sharp cortical tuning found in humans and the broad cochlear tuning found in laboratory animals (16) or from earlier estimates in humans using simultaneous masking (6). The tuning bandwidths estimated in human cortical neurons (~1/12 octave) are in fact remarkably similar to the estimates of human cochlear tuning that we have validated here (~1/13 octave, 11), indicating that further central processing may not be necessary to account for narrow cortical tuning. Our results also provide new data to inform a classical debate in auditory neuroscience on whether the auditory system extracts spectral information from sounds in the form of a rate-place code or a code based on spike timing information, or a combination of the two (34). Proposals involving timing codes have been partly motivated by the poor rate-place coding found in animal studies (13, 14). Indeed, ferret cochlear bandwidths are barely sufficient to resolve adjacent formants (e.g., in the 2-3-kHz region the 2nd and 3rd formants can be around 1/3 octave apart (35), close to the bandwidth of ferret auditory filters in this region). According to the narrower human bandwidths validated here, however, rate-place coding schemes would have considerably more success at representing the formant peaks of human speech in the human auditory system than in other species.

Although we have confirmed sharp human cochlear tuning using low-intensity sounds similar to those used to measure auditory-nerve tuning curves in other species, tuning is known to change with sound intensity, becoming broader at high intensities. Behavioral measures in humans have also revealed broader tuning at high sound intensities (36), in line with expectations. In addition, the saturation of firing rate in the auditory nerve at higher intensities also leads to effectively broader tuning and poorer resolution in the majority of auditory nerve fibers at sound levels where human speech recognition remains robust (13). It is possible that tuning under more complex acoustic conditions is sharpened by central auditory processing, beyond what can be explained by firing rate in the auditory nerve, especially at high levels. Such sharpening might occur through mechanisms involving stimulus-driven spike timing, or

phase locking, and lateral inhibition based on the rapid phase transitions produced by the basilar-membrane traveling wave (37). The extent to which putative sharpening mechanisms are required to explain behavioral performance at high sound intensities remains to be explored in light our new understanding of human cochlear tuning at low intensities.

It is tempting to relate sharp human cochlear tuning to our ability to perceive the subtleties of speech (particularly those involving prosody and pitch) in complex backgrounds, and thus our ability to solve the 'cocktail party problem' (38). However, there is evidence for intermediate cochlear tuning in non-human primates (28), and one study reported cortical tuning in a non-human primate that approached that observed in humans (39). In addition, studies of otoacoustic emissions in another large mammal—the tiger—have also suggested that tuning may approach that found in humans (40). These findings imply that the physical size of the cochlea and its associated tonotopic map play a more important role than any human-specific evolution of cochlear tuning (41). Even though sharp cochlear tuning may not be a sufficient condition for the emergence of speech as an effective communication mode (42), it may nevertheless have played an important and perhaps necessary role in its development. Given the complexity of this and the other issues discussed, the development of cochlear models that produce realistic sharp tuning and the non-linear characteristics which impart dependence on stimulus paradigms, will provide an important step towards evaluating such claims and consolidating our understanding of frequency selectivity, the cochlea and their relation to perception.

Experimental Methods

Full details of experimental methods are given in the SI Appendix. Briefly, we trained ferrets to detect (43) or to lateralize (44) brief tones or narrowband noise, in the presence of masking noise, in a positive reinforcement procedure. Using these behavioral methods in ferrets, we measured perceptual thresholds using different variants of notched noise maskers (6, 22). We also made measurements using similar stimulus paradigms in humans. We also recorded, in

lightly anaesthetized ferrets, the otoacoustic emissions elicited by pure tone stimuli, using the stimulus frequency otoacoustic emissions (SFOAE) method (45). Estimates of frequency selectivity derived from these data were compared with previous recordings from the auditory nerve of anaesthetized ferrets (27). In the human studies, all participants provided written informed consent prior to participating, and all procedures were approved by the Institutional Review Board of the University of Minnesota.

Acknowledgements

We thank technicians at Nottingham for assistance with collecting behavioral data, and Angie Killoran for formatting assistance. Experimental work in ferrets was supported by MRC intramural funding (MC_UU_00010/1 and U135097127). Behavioral experiments in humans (AJO, HAK) were supported by NIH grant R01DC012262. CAS was supported by NIH grant R01 DC003687. CB was supported by the Natural Sciences and Engineering Research Council of Canada (RGPIN-430761-2013). We also thank Brian Moore and an additional reviewer for thorough and constructive comments.

Author Contributions

CJS, AJO, CAS, and ARP conceived the study and wrote the paper with CB and JS. TTW and JS conducted the behavioral experiments and analyzed these data with support from CJS and AJO. CB, ARP and CJS conducted the OAE experiments and CB and CAS analyzed these data. ARP and CJS conducted the auditory nerve experiments. HAK conducted the human behavioral experiments.

Author information

The authors declare no competing financial interests. Correspondence or requests for material should be addressed to CJS, AJO, or CAS.

Bibliography (45)

1. Merzenich MM, Roth GL, Anderson RA, Knight PL, & Colwell SA (1977) Some basic features of the organization of the central auditory nervous system. *Psychophysics and Physiology of Hearing* eds Evans EF & Wilson JP (Academic Press), pp 485-496.
2. Malmierca MS*, et al.* (2008) A discontinuous tonotopic organization in the inferior colliculus of the rat. *J Neurosci* 28(18):4767-4776.
3. Langers DR & van Dijk P (2012) Mapping the tonotopic organization in human auditory cortex with minimally salient acoustic stimulation. *Cereb Cortex* 22(9):2024-2038.
4. Shera CA (2015) The spiral staircase: tonotopic microstructure and cochlear tuning. *J Neurosci* 35(11):4683-4690.
5. Oxenham AJ (2018) How We Hear: The Perception and Neural Coding of Sound. *Annual review of psychology* 69:27-50.
6. Glasberg BR & Moore BCJ (1990) Derivation of auditory filter shapes from notched-noise data. *Hearing research* 47(1-2):103-138.
7. Zwicker E (1961) Subdivision of the Audible Frequency Range into Critical Bands *J Acoust Soc Am* 33:248.
8. Rhode WS (1971) Observations of the vibration of the basilar membrane in squirrel monkeys using the Mossbauer technique. *J. Acoust. Soc. Am.* 49(4):Suppl 2:1218+.
9. Liberman MC (1978) Auditory-nerve response from cats raised in a low-noise chamber. *J. Acoust. Soc. Am.* 63(2):442-455.
10. Harrison RV, Aran JM, & Erre JP (1981) AP Tuning Curves from Normal and Pathological Human and Guinea-Pig Cochleas. *J. Acoust. Soc. Am.* 69(5):1374-1385.
11. Shera CA, Guinan JJ, Jr., & Oxenham AJ (2002) Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. *Proceedings of the National Academy of Sciences of the United States of America* 99(5):3318-3323.
12. Bergevin C, Manley GA, & Koppl C (2015) Salient features of otoacoustic emissions are common across tetrapod groups and suggest shared properties of generation mechanisms. *Proceedings of the National Academy of Sciences of the United States of America* 112(11):3362-3367.
13. Sachs MB & Young ED (1979) Encoding of steady-state vowels in the auditory nerve: representation in terms of discharge rate. *J Acoust Soc Am* 66(2):470-479.
14. Delgutte B (1984) Speech coding in the auditory nerve: II. Processing schemes for vowel-like sounds. *J. Acoust. Soc. Am.* 75(3):879-886.
15. Baer T & Moore BCJ (1994) Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech. *J Acoust Soc Am* 95(4):2277-2280.
16. Bitterman Y, Mukamel R, Malach R, Fried I, & Nelken I (2008) Ultra-fine frequency tuning revealed in single neurons of human auditory cortex. *Nature* 451(7175):197-201.
17. Ruggero MA & Temchin AN (2005) Unexceptional sharpness of frequency tuning in the human cochlea. *Proc. Natl. Acad. Sci. USA.* 102(51):18614-18619.
18. Shera CA, Guinan JJ, Jr., & Oxenham AJ (2010) Otoacoustic estimation of cochlear tuning: validation in the chinchilla. *J. Assoc. Res. Otolaryngol.* 11(3):343-365.
19. Lopez-Poveda EA & Eustaquio-Martin A (2013) On the controversy about the sharpness of human cochlear tuning. *J. Assoc. Res. Otolaryngol.* 14(5):673-686.
20. Siegel JH*, et al.* (2005) Delays of stimulus-frequency otoacoustic emissions and cochlear vibrations contradict the theory of coherent reflection filtering. *J Acoust Soc Am* 118(4):2434-2443.
21. Charaziak KK & Siegel JH (2015) Tuning of SFOAEs Evoked by Low-Frequency Tones Is Not Compatible with Localized Emission Generation. *J Assoc Res Otolaryngol* 16(3):317-329.
22. Oxenham AJ & Shera CA (2003) Estimates of human cochlear tuning at low levels using forward and simultaneous masking. *J Assoc Res Otolaryngol* 4(4):541-554.

23. Moore BCJ & Glasberg BR (1981) Auditory filter shapes derived in simultaneous and forward masking. *J Acoust Soc Am* 70(4):1003-1014.
24. Houtgast T (1974) Lateral suppression in hearing. PhD Thesis (University of Amsterdam, Amsterdam).
25. Glasberg BR & Moore BCJ (2000) Frequency selectivity as a function of level and frequency measured with uniformly exciting notched noise. *J Acoust Soc Am* 108(5 Pt 1):2318-2328.
26. Rosen S, Baker RJ, & Darling A (1998) Auditory filter nonlinearity at 2 kHz in normal hearing listeners. *J Acoust Soc Am* 103(5 Pt 1):2539-2550.
27. Sumner CJ & Palmer AR (2012) Auditory nerve fibre responses in the ferret. *Eur. J. Neurosci.* 36(4):2428-2439.
28. Joris PX, Bergevin C, Kalluri R, McLoughlin M, Michelet P, van der Hiejden M & Shera CA (2011) Frequency selectivity in Old-World monkeys corroborates sharp cochlear tuning in humans. *Proc. Natl. Acad. Sci. USA.* 108(42):17516-17520.
29. Patterson RD, Nimmo-Smith I, Weber DL, & Milroy R (1982) The deterioration of hearing with age: frequency selectivity, the critical ratio, the audiogram, and speech threshold. *J Acoust Soc Am* 72(6):1788-1803.
30. Burton JA, Dylla ME, & Ramachandran R (2018) Frequency selectivity in macaque monkeys measured using a notched-noise method. *Hearing research* 357:73-80.
31. Evans EF (2001) Latest comparisons between physiological and behavioural frequency selectivity. *Physiological and Psychological Bases of Auditory Function*, eds Breebaart DJ, Houtsma AJM, Kohlrausch AG, Prijs VF, & Schoonhoven R (Shaker Publishing, Mierlo, the Netherlands. Maastricht), pp pp. 382–387.
32. May BJ, Kimar S, & Prosen CA (2006) Auditory filter shapes of CBA/CaJ mice: behavioral assessments. *J Acoust Soc Am* 120(1):321-330.
33. Yost WA & Shofner WP (2009) Critical bands and critical ratios in animal psychoacoustics: an example using chinchilla data. *J Acoust Soc Am* 125(1):315-323.
34. Young ED & Sachs MB (1979) Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers. *J. Acoust. Soc. Am.* 66(5):1381-1403.
35. Catford JC (1988) *A Practical Introduction to Phonetics* (Oxford University Press, Oxford).
36. Nelson DA & Freyman RL (1984) Broadened forward-masked tuning curves from intense masking tones: delay-time and probe-level manipulations. *J Acoust Soc Am* 75(5):1570-1577.
37. Shamma SA (1985) Speech processing in the auditory system. I: The representation of speech sounds in the responses of the auditory nerve. *J Acoust Soc Am* 78(5):1612-1621.
38. Cherry EC (1953) Some experiments on the recognition of speech, with one and with two ears. *J. Acoust. Soc. Am.* 25:975-979.
39. Bartlett EL, Sadagopan S, & Wang X (2011) Fine frequency tuning in monkey auditory cortex and thalamus. *J. Neurophysiol.* 106(2):849-859.
40. Bergevin C, Walsh EJ, McGee J, & Shera CA (2012) Probing cochlear tuning and tonotopy in the tiger using otoacoustic emissions. *Journal of comparative physiology. A, Neuroethology, sensory, neural, and behavioral physiology* 198(8):617-624.
41. Shera CA & Charaziak KK (2018) Cochlear Frequency Tuning and Otoacoustic Emissions. *Cold Spring Harbor perspectives in medicine*.
42. Fitch WT (2000) The evolution of speech: a comparative review. *Trends. Cogn. Sci.* 4(7):258-267.
43. Alves-Pinto A, Sollini J, & Sumner CJ (2012) Signal detection in animal psychoacoustics: Analysis and simulation of sensory and decision-related influences. *Neurosci.* 220:215-227.
44. Sollini J, Alves-Pinto A, & Sumner CJ (2016) Relating approach-to-target and detection tasks in animal psychoacoustics. *Behav. Neurosci.* 130(4):393-405.
45. Shera CA & Bergevin C (2012) Obtaining reliable phase-gradient delays from otoacoustic emission data. *J Acoust Soc Am* 132(2):927-943.

Figure legends

Figure 1 | Three different ways of estimating cochlear tuning used in ferrets. Auditory Nerve Fibers (ANFs): Threshold levels (grey line) for a response are fit with a filter model (brown line), from which the equivalent rectangular bandwidth (ERB; dashed grey line) is calculated. **Otoacoustic Emissions (OAEs):** The mean phase gradient of OAEs (red line) is used to estimate filter sharpness, Q_{ERB} (= f/ERB), using the approximate species invariance of the tuning ratio. **Psychophysical Masking (PSY):** The behavioral detection of a pure tone in the presence of two bands of noise, separated by varying spectral distances. ERB (blue dashed line) is estimated by fitting a filter model (brown) to the detection thresholds.

Figure 2 | Three measures of frequency selectivity agree. a. Filter sharpness from psychophysical forward masking (PSY-F) agrees closely with auditory nerve fiber (ANF) and otoacoustic emission (OAE) measurements. Tuning in individual nerve fibers (grey points), psychophysical forward masking (blue points) and a loess trend and its bootstrapped 95% CI for the otoacoustic emissions measurements. Dashed lines indicate bootstrapped 95% CIs for the perceptual data. **b.** Forward masking (PSY-F; blue points, n=8) yields a better match to auditory nerve tuning than simultaneous masking (PSY-S; magenta points, n=22). In **b** auditory nerve data are shown as the area within the loess (see Supplementary Methods) trend 95% CI.

Figure 3 | Comparing different measures of frequency resolution in the ferret, independently of the effect of signal frequency. a. The different tuning measurements as a fraction of the mean auditory-nerve fiber tuning at a given frequency. Dashed red lines show excluded OAE outliers (see text). b. Statistical comparison of the different measures of tuning. Horizontal bars show the mean of each measure as a fraction of auditory nerve tuning, and also as effect size (relative to ANF tuning). Asterisks next to data points indicate significant differences compared to auditory nerve tuning. * $p<.05$; ** $p<.01$; *** $p<.001$.

Supplementary Information for:

Mammalian behavior and physiology converge to confirm sharper cochlear tuning in humans

Christian J. Sumner, Toby T. Wells, Christopher Bergevin, Joseph Sollini, Heather A. Kreft, Alan R. Palmer, Andrew J. Oxenham, Christopher A. Shera

Corresponding author: Christian J. Sumner

Email: christian.sumner2@nottingham.ac.uk

This PDF file includes:

Supplementary text

Figs. S1 to S3

References for SI reference citations

Supplementary Information - Experimental methods

All procedures with ferrets were carried out under license from the UK Home Office, in accordance with the Animals (Scientific Procedures) Act 1986.

Behavioral methods: Ferrets

Several variants of the notched-noise masking paradigm (1) were used to measure perceptual frequency selectivity in ferrets. Tasks measured the ability of ferrets to detect tones (or narrowband noise) presented in a noise which had the possibility to mask the signal (render it inaudible). The noise masker was constructed of two bands of noise, separated by a 'notch' of a specified and symmetrical frequency range around the target tone (Fig. 1 PSY; Supplementary Fig. S1b). The notch width is defined as the distance from the signal frequency (f_s) to the inner edge of each noise band (f_n) divided by f_s, i.e. $\Delta f = (f_s - f_n)/f_s$. The method assumes that detectability of the tone is determined by the signal-to-noise ratio at the output of a linear filter through which the subject is listening. The masker notch width determines how much of the masking noise is passed by this filter, and therefore affects the signal-to-noise ratio at which behavioral performance equals a criterion detection threshold. Systematically varying the notch width provides a profile of the spectral sensitivity to the noise, which is a function of the auditory filter shape.

Ferrets were trained in positive reinforcement procedures to respond to tonal stimuli for water reward, following standard protocols and welfare monitoring as detailed previously (2). Behavior took place in a sound attenuated booth, with walls lined with mineral wool, within which was a caged arena containing a number of water spouts and loudspeakers around the perimeter. Ferrets were trained on a 1-interval 2-alternative forced choice 'left/right' lateralization task. During a trial, target tones or narrowband noise were presented from a loudspeaker to the left (-90°) or right (+90°; equal probability). The ferret responded at the water spout co-located with the source loudspeaker for a reward. Lateralization produces thresholds which are closely matched to a detection task, but is easier to train (3).

Detection thresholds of the signals in notched-noise maskers were measured in several different ways, which encompassed a number of different stimulus paradigms used in humans. Nine ferrets were required to lateralize a train of ten 25-ms (including 12.5-ms raised-cosine on/off ramps) pulses of 1/32nd octave wide narrowband noise, centered on the signal frequency. Each pulse was separated by 175-ms of silence, making a 2-s signal in total. The masker consisted of 150-ms (including 5-ms raised-cosine on/off ramps) pulses of notched-noise, separated by 50-ms of silence (Supplementary Fig. S1c); the same (i.e. correlated)

noise samples were presented simultaneously from two loudspeakers which were placed close to the signal loudspeakers. Note that having correlated noise presented from two locations and a target from one location could potentially create binaural cues especially at low frequencies (< 1 kHz). However, given the brief duration of the signal, the potential binaural cues were unlikely to have affected performance. The bandwidth of the masker bands below and above the signal frequency were each $0.25f_c$. The masker ran continuously, with a fixed notch width (centred on the signal in linear frequency with a notched width, Δf, of 0.0-0.4) through a behavioral session. The relative timing of the signal and masker pulses was controlled so that signals were either simultaneous with a masker pulse (100-ms onset asymmetry) or immediately following (150-ms onset asymmetry). This allowed filter measurements to be made using either simultaneous (PSY-S) or forward masking (PSY-F; Supplementary Fig. S1c). Psychometric functions were generated, individually for each ferret and filter measurement, either by fixing the masker at a low sensation level for which a masker with $\Delta f = 0$ raised thresholds compared with silence by >20dB (noise spectrum levels of either 0 or 20 dB SPL/Hz) and varying the signal level from trial-to-trial using the method of constant stimuli (this was done for 10-kHz signals only), or by fixing the signal at a low level near to the top of the psychometric function in silence (in the range 0-40 dB SPL), and varying the masker level from trial to trial. This allowed us to measure filter widths using either the fixed signal level method, or the fixed masker level method (4, 5).

In addition, four ferrets were required to lateralize a 500-ms tone (including 25-ms cosine-squared on-off ramps) in a continuous notched-noise masker (thus PSY-S), which was fixed at a low sensation level, as above (with noise spectrum levels of 0, 10, or 20 dB SPL/Hz). The notch width of the noise was varied across sessions in log-frequency around the signal with notches varying from 0- to 1-octave in 0.1 or 0.2 octave steps. The level of the tone was varied according to the method of constant stimuli. Data from four additional ferrets in a previous study were also included in our analysis (2). Stimuli were single 500-ms pure tones in a continuous notched-noise masker as above. However, they were trained in a 'yes/no' detection task (for details of training and testing see 2).

All datasets were analyzed in the same way. An unbiased measure of performance, $P(c)_{max}$, was calculated using standard signal detection theory methods (6). Psychometric functions were fitted with a logistic function [$P(c)_{max}(level, \alpha, m,s) = 0.5 + (0.5 + \alpha)/(1 + \exp(-(level - m)/s)$] from which threshold levels were measured with a 75% correct criterion. For a given filter calculation s and α were fixed, which constrained all psychometric functions to be parallel. For the forward-masking, fixed-masker-level measurements, thresholds were adjusted using

a separate measurement of the growth of forward masking, following Glasberg and Moore (4). The level of the signal at threshold was estimated as a function of masker level (spanning ±20 dB SPL/Hz in 10-dB intervals) for the zero-notch condition, with a temporal arrangement of masker and signal identical to that used in the notched-noise measurements. From these thresholds the slope of the growth of masking (dB/dB) was estimated by fitting a straight line on a dB scale (0.43 dB/dB at 10 kHz).

Each set of threshold-versus-notch-width functions, for each ferret, was fitted to an analytical linear filter-bank model, which takes into account the possibility of off-frequency listening, closely following Oxenham and Shera (5). We used a symmetrical, 2-parameter rounded exponential function, *roex(pr)*. This was originally proposed by Patterson et al. (7) for fitting to human psychophysical data, and applied to guinea-pig auditory-nerve tuning data by Evans et al. (8). It is expressed as:

$$I(g) = k(1-r)(1+pg)\exp(-pg) + r, \qquad \text{eqn 1}$$

Where g is the normalized frequency deviation from the filter's center frequency, $(f-f_c)/f_c$, p defines the slope (and hence the bandwidth) of the filter, and r is the function floor. A constant, k, which defines the ratio between the signal and the noise power within the roex filter, is adjusted to produce the best predictions in a least-squares sense, but does not affect the tuning of the filter. The ERB was calculated as the bandwidth of a rectangular filter that would pass the same power as the fitted *roex* function, and a corresponding measure of sharpness, Q_{ERB}, was calculated as CF/ERB. The confidence intervals (CIs) were calculated via a bootstrapping technique.

A total of 45 auditory filter estimates were made (conservatively estimated to have taken 60 ferret-months of data collection, not including training, method development, and piloting). We rejected a proportion of measurements (n=14), and did not include these in our comparisons with other data, if the bootstrapped CIs were larger than the mean auditory nerve ERB at that frequency (and thus could not be compared meaningfully) or if k fell outside of 0.953 x interquartile range (equivalent to 95th-percentile in a normal distribution; 9): large values indicate poor thresholds indicative of poor behavior; low values are physically impossible and indicative of a fitting failure. Behavioral ERBs were analyzed using a general linear model [$\log_{10}$(ERB) vs. $\log_{10}$(Signal frequency) x forward/simultaneous masking x fixed signal/masker x task; where task varied in both the number of targets and whether the responses was signal lateralization or signal presence]. This showed a significant relationship between signal frequency and ERB (p<<0.0001) and between forward and simultaneous masking (p<0.05),

but no differences between three different task methodologies (p=0.67). Therefore we grouped these data together for comparison with the OAE and AN derived ERBs and considered only the difference between forward (n=8) and simultaneous masking (n=22).

OAE methods

Measurements were made in 21 ferrets, lightly anesthetized via a single dose of metatomidine (0.1mg/kg) and ketamine (5mg/kg). Body temperatures were maintained by a heating blanket. Following recordings, the animals were given a reversal agent (antipamezole, 0.25mg/kg) and returned to their home cages once recovered.

OAE data were collected using stimulus paradigms identical to those previously used with humans. An Etymotic ER-10C containing the stimulus transducers and the measurement microphone was coupled to the meatus. Stimulus-frequency (SF)OAEs were evoked from 1–11 kHz using swept tones (10) and extracted using the interleaved suppression paradigm (11). Briefly, after recording the response to a single swept "probe" tone (which contains both the stimulus at 40 dB SPL and the SFOAE at the stimulus frequency), the recording is repeated with the addition of a swept "suppressor" tone (15 dB higher in level and 40 Hz higher in frequency). Because the additional tone suppresses the SFOAE from the cochlea, vector (frequency-domain) subtraction of the suppressed response from the probe-alone recording reveals the emission evoked by the probe tone. Frequency-delay functions were computed from the slope of the unwrapped SFOAE phase (e.g., Fig. 1 OAE) using a peak-picking algorithm (12).

SFOAE frequency-delay functions (Supplementary Fig. S2a) have been related to the sharpness (Q_{ERB}) of cochlear frequency tuning using the "tuning ratio" (11). The tuning ratio—defined as the ratio of the Q_{ERB} trend obtained from auditory-nerve fibers and the SFOAE delay trend (expressed in periods of the OAE frequency)—has been found to be similar across numerous species of mammals (approximate species invariance of the tuning ratio). Here, we applied previous estimates of the tuning ratio—averaged across cats, guinea pigs, and chinchillas (13) after compensating for differences in the apical-basal transition frequency—to derive estimates of the tuning trend from the SFOAE delays recorded in ferrets (Supplementary Fig. S2b). The apical-basal transition frequency in ferret was estimated from the bend in the frequency-delay function located using power-law fits to the data and was approximated as 3.5 kHz (i.e., similar to that in cats). Quarter-octave variations in the transition frequency had only minor effects on the results reported here. For the statistical analysis, the tuning ratio was applied to the individual SFOAE delay data points (rather than simply to their

trend). Extreme outliers (262/1670) were excluded (shown in Fig. 3 and Supplementary Fig. S2) if $\log_{10}(Q_{ERB})$ exceeded 2.5 standard deviations from the mean.

Auditory-nerve recordings

Auditory-nerve recordings were made from anesthetized adult ferrets, as reported in detail in Sumner and Palmer (14). Recordings were carried out in a sound-attenuated booth, by inserting 2.7 M KCl filled micropipettes (50-150 MΩ) into the auditory nerve on the left hand side. Signals were generated digitally (16 bit, 100 kHz sample rate), amplified and filtered (300-2000 Hz) and delivered monaurally via a closed-field system (flat ±10 dB from 100 to 35000 Hz). Spike times were recorded using TDT System II (Tucker-Davis Technologies, Alachua, FL, USA) hardware.

For each recording a single nerve fiber frequency response area was recorded by presenting pseudorandom sequences of pure tones (50 ms duration, 2 ms rise-fall time, presented every 200 ms) varying frequency (from -3 to +1 octaves around CF in 1/8 octave steps) and sound level (from below CF threshold to +40 dB or more in 1-5 dB steps). The number of spikes elicited by each condition was calculated (Fig. 1 ANF).

Frequency-threshold-tuning curves (FTCs; grey solid line in Fig. 1 ANF) were calculated from the frequency response areas (14) with a threshold criterion equivalent to 1 spike per tone presentation. The ERB of each fiber was calculated by fitting a rounded exponential function, *roex(2pr)*, to the FTC (brown line in Fig. 1 ANF), after inverting the FTC in sound-level to create a pseudo-filter response function. The upper and lower slopes were allowed to vary independently:

$$I(g) = k\,(1+p_U\,|g|)\,\exp(-p_U|g|) + r, \quad \text{for } g \le 0 \text{ and} \qquad \text{eqn 2}$$

$$I(g) = k\,(1+p_L\,|g|)\exp(-p_L|g|) + r, \quad \text{for } g > 0$$

where I is the threshold sound intensity at stimulus frequency f, $g = (CF-f)/CF$, k is the threshold sound intensity at CF ($g=0$) and r is the function floor.

Choice of analytical filter model

The two parameter *roex(pr)* filter model (7; eqn 1) used to estimate psychophysical bandwidths was chosen after comparing different filter models from the family of *rounded-exponential* functions which have previously been used to fit psychophysical data (5).

We compared eqn 1 with a simpler *one* parameter model, *roex(p)*, which omits the noise floor parameter, *r*.

$$W(g) = k(1 + pg)\exp(-pg), \qquad \text{eqn 3}$$

In a subset of our psychophysical data in ferrets (9 separate fixed masker-level, forward-masked filter estimates at 4 different frequencies in 6 different ferrets), we measured thresholds when notched noise maskers were asymmetric (upper and lower band edges of noise bands having different Δfs of 0.2/0.4 or 0.4/0.2). These data allowed us to evaluate the *roex(2pr)* (eqn 2) and the 3 parameter model, *roex(pwt)*, which had been previously shown to yield good fits to human psychophysical data (5):

$$W(g) = (1 - w)(1 + pg)\exp(-pg) + w(1 + pg/t)\exp(-pg/t), \quad \text{for } g>0 \qquad \text{eqn 4}$$
$$W(g) = (1 + pg)\exp(-pg) \quad \text{for } g\leq 0$$

where w determines the tip-to-tail ratio on the low frequency side, and t determines the difference in slope of the tip and the tail on the low-frequency side; the steepness of the tip of the filter is determined by p on both sides. The symmetrical models (eqns 1 and 3) were fitted both with the symmetrical set of notch widths alone, and also with the inclusion of the asymmetrical notch conditions.

All models were subjected to a bootstrap estimation of confidence intervals of the quality of fit (RMS error in dB), where trials were resampled (500 iterations) with replacement and the fitting process was rerun. The *roex(pr)*, fitted to the symmetrical conditions only, yielded the overall lowest RMS error for both the bootstrap tests. A further advantage was that it could be applied to all the ferret psychophysical data. For further confidence in our conclusions, we also repeated our statistical comparisons of the behavioral data with the simpler *roex(p)* model which was fitted only to symmetrical notch conditions, and the best fitting asymmetric model, the *roex(2pr)*, fitted to the asymmetrical notch conditions. These alternative models did not change the results of any statistical tests (Supplementary Fig. S3). Thus, our conclusions are not contingent on the choice of filter model.

Supplemental behavioral data in humans

Human auditory filter shapes have been estimated in previous studies using signal detection methods (5, 15). However, to provide a direct comparison to the behavioral estimates in ferrets, new data were collected in humans using the same signal location discrimination task as was used in ferrets, with the same spectral notch widths, the same correlated noise emanating from two directions, and the same filter fitting procedure. Nine listeners were tested (3 males, 6 females; ages 19 to 35). All had audiometrically normal hearing (thresholds<20 dB hearing level at octave frequencies between 250 and 8000 Hz).

The target tone was a 4000-Hz tone (total duration of 10-ms, gated with 5-ms raised-cosine ramps). The notched-noise masker consisted of two spectral bands of Gaussian noise, each with a bandwidth of 1000 Hz ($0.25f_s$, as with the ferrets) and a total duration of 200-ms, including 5-ms raised-cosine onset and offset ramps. The gap between the offset of the masker and the onset of the target tone was 2 ms.

The absolute threshold for the target tone was measured first for each listener. The target tone was passed through non-individualized head-related transfer functions (HRTFs) to simulate a tone originating from 30 degrees to the left or right of the listener. On each trial the target tone was presented from either the left or right (with equal probability) and the listener had to indicate via button press whether the sound emanated from the left or right. An adaptive tracking procedure was used that tracks the 79.4% correct point on the psychometric function. The target level began at a clearly audible level of 40 dB SPL and was adaptively varied in steps of 8 dB for the first two reversals, 4 dB for the next four reversals, and 2 dB for the final six reversals. The run was terminated after 12 reversals, and the threshold was defined as the mean target level at the last six reversals. Each subject completed six runs and the thresholds were averaged for that subject.

In the forward-masking experiment, the target tone level was fixed at 12 dB above the threshold for each subject, for all spectral notches. In each notch condition, the spectrum level of the masker bands was varied adaptively to determine the masker level necessary to just mask the target tone. The masker bands were presented from both 30 degrees to the left and 30 degrees to the right of the listener (i.e., the same positions from which the target could be presented). The noises presented from each side were identical, as in the ferret experiments, and were individually filtered using HRTFs to simulate free-field stimulation. The notch conditions tested were Δf = 0/0, 0.1/0.1, 0.2/0.2, 0.3/0.3, 0.4/0.4, 0.1/0.3, 0.3/0.1, 0.2/0.4 and 0.4/0.2. The masker levels at threshold were then used to derive auditory filter shapes, ERB and Q_{ERB} for each subject, based on the *roex(2pr)* function (eqn 2), as was also possible with a subset of ferret data, for which exactly the same set of notch widths had been measured. The mean ERB of 290 Hz (s.d. 63 Hz), corresponding to a Q_{ERB} of 13.8, is similar to the values (ERB of 271 Hz and Q_{ERB} of 14.7) derived in earlier human studies using a monaural detection (rather than localization) task (5, 15), and is narrower than the mean estimates at 4 kHz in ferrets by a factor of ~3 (see Supplementary Fig. S3).

Statistical models

The data were fitted with a linear model of $\log_{10}Q_{ERB}$ with $\log_{10}$CF and dataset (ANF/OAE/PSY-S/PSY-F) as predictors. The very different nature of these measurements meant that sample sizes differed greatly between datasets (PSY n=30; ANF n=91; OAE n=1408), and variance was also not identical (Fig. 3). Therefore the statistical significance of differences between groups was performed using a (two-sided) *sandwich-test* (16), which is robust to differences in sample size, variance, and normality. The results of these tests are shown in Fig. 3 and Supplementary Figure S3. This model was also separately applied to the cases where behavioral estimates were fitted with the *roex(p)* and *roex(2pr)* models. In the case of the *roex(2pr)* model, the results from the human behavior were also included (Supplementary Fig. 3).

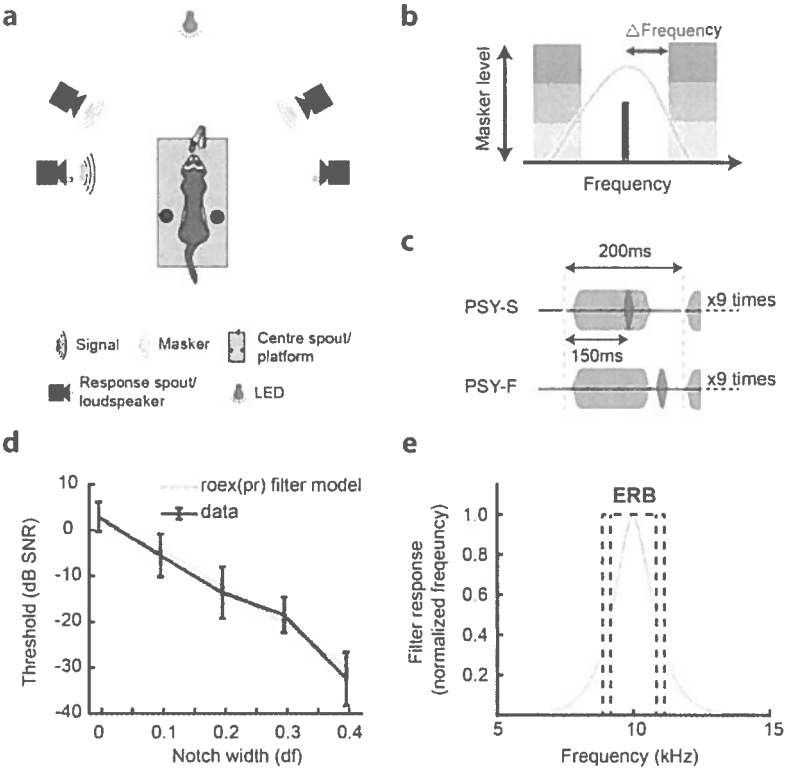
Data and code availability

The data and code used to analyze the data are available on request.

Supplementary references

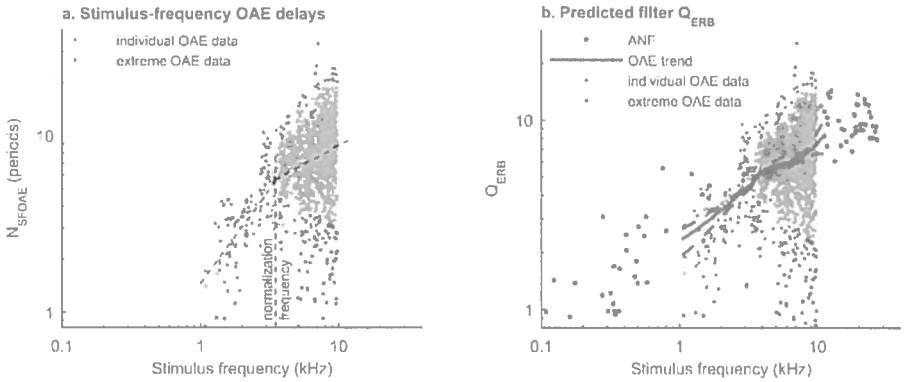
1. Patterson RD (1976) Auditory filter shapes derived with noise stimuli. *J Acoust Soc Am* 59(3):640-654.
2. Alves-Pinto A, Sollini J, & Sumner CJ (2012) Signal detection in animal psychoacoustics: Analysis and simulation of sensory and decision-related influences. *Neurosci.* 220:215-227.
3. Sollini J, Alves-Pinto A, & Sumner CJ (2016) Relating approach-to-target and detection tasks in animal psychoacoustics. *Behav. Neurosci.* 130(4):393-405.
4. Glasberg BR & Moore BCJ (1982) Auditory filter shapes in forward masking as a function of level. *J Acoust Soc Am* 71(4):946-949.
5. Oxenham AJ & Shera CA (2003) Estimates of human cochlear tuning at low levels using forward and simultaneous masking. *J Assoc Res Otolaryngol* 4(4):541-554.
6. Macmillan NA & Creelman CD (2005) *Detection Theory: A User's Guide* (Lawrence Erlbaum Associates) 2nd Ed.
7. Patterson RD, Nimmo-Smith I, Weber DL, & Milroy R (1982) The deterioration of hearing with age: frequency selectivity, the critical ratio, the audiogram, and speech threshold. *J Acoust Soc Am* 72(6):1788-1803.
8. Evans EF (2001) Latest comparisons between physiological and behavioural frequency selectivity. *Physiological and Psychological Bases of Auditory Function*, eds Breebaart DJ, Houtsma AJM, Kohlrausch AG, Prijs VF, & Schoonhoven R (Shaker Publishing, Mierlo, the Netherlands. Maastricht), pp pp. 382–387.
9. Tukey J (1977) *Exploratory data analysis* (Addison-Wesley Publications Co, Reading, PA).
10. Kalluri R & Shera CA (2013) Measuring stimulus-frequency otoacoustic emissions using swept tones. *J. Acoust. Soc. Am.* 134(1):356-368.
11. Shera CA, Guinan JJ, Jr., & Oxenham AJ (2010) Otoacoustic estimation of cochlear tuning: validation in the chinchilla. *J. Assoc. Res. Otolaryngol.* 11(3):343-365.
12. Shera CA & Bergevin C (2012) Obtaining reliable phase-gradient delays from otoacoustic emission data. *J Acoust Soc Am* 132(2):927-943.
13. Joris PX, Bergevin C, Kalluri R, McLoughlin M, Michelet P, van der Heijden M & Shera CA (2011) Frequency selectivity in Old-World monkeys corroborates sharp cochlear tuning in humans. *Proc. Natl. Acad. Sci. USA.* 108(42):17516-17520.
14. Sumner CJ & Palmer AR (2012) Auditory nerve fibre responses in the ferret. *Eur. J. Neurosci.* 36(4):2428-2439.
15. Shera CA, Guinan JJ, Jr., & Oxenham AJ (2002) Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. *Proceedings of the National Academy of Sciences of the United States of America* 99(5):3318-3323.
16. Hothorn T, Bretz F, & Westfall P (2008) Simultaneous inference in general parametric models. *Biom. J.* 50(3):346-363.
17. Cleveland WS & Devlin S (1988) Locally-Weighted Regression: An Approach to Regression Analysis by Local Fitting. *Journal of the American Statistical Association* 83(403):596-610.

Supplementary Information – Figures

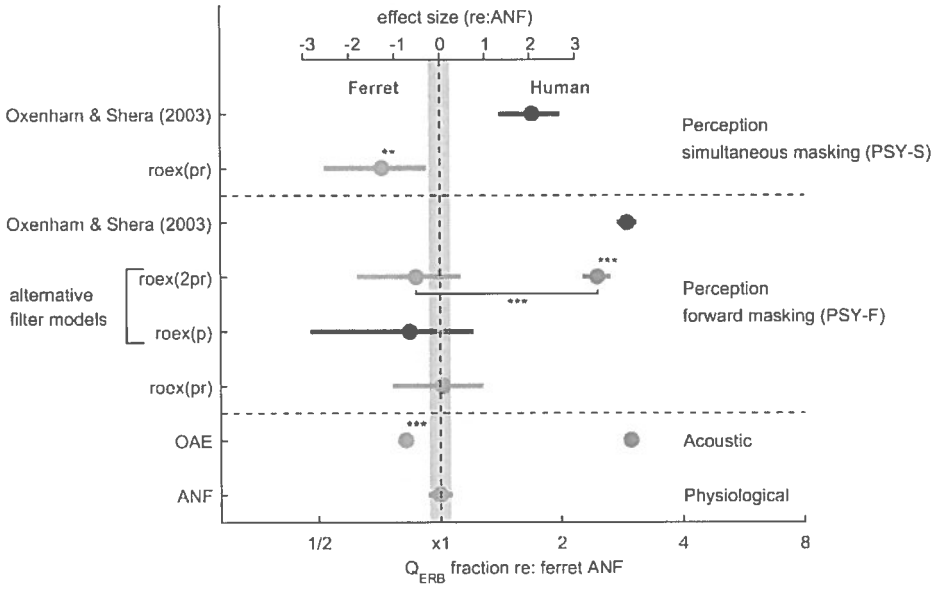


Supplementary Figure S1 | Exemplar behavioral methods. a. The lateralization task. Animals initiate a trial by licking a spout on the central platform. They respond to the left or right spout to indicate the spatial location of the signal. **b. Notched-noise masking in the frequency domain**: The behavioral detection of a pure tone (or narrowband noise) in the presence of two bands of noise, separated from the signal in the frequency domain by varying spectral distances (Δf). Here the level of the masker is shown as varying, to find the threshold for detection of the tone. This threshold depends on how much noise energy passes through the perceptual filter used to detect the tone. **c. Forward masking in the time domain**. The masker is a continuous train of bursts of notched-noise. Each trial measures the lateralization of a train of 10 signal pulses, each of which is simultaneous (PSY-S) or immediately follows (PSY-F) a noise-burst. **d. Thresholds for a single filter measurement, and fits by the *roex* filter model.** Thresholds measured at each notch width (Δf) shift to lower SNRs (i.e., higher masker spectrum levels) with increasing Δf. Q_{ERB} is estimated by fitting a filter model to the detection thresholds, as for ANF data. In this case,

the target stimulus is a narrowband noise (1/32nd octave wide), centered at 10 kHz and presented at 7 dB SPL in a forward-masking paradigm, as in **b**. The threshold signal-to-noise ratio (SNR) is the difference between the signal level and the masker spectrum level at threshold. The shaded area and error bars show 95% CI for the data and fit. **e. Derived filter frequency response.** The filter shape is the best-fitting filter to the data shown in panel d. The shaded area shows the 95% CI for the filter shape. The ERB is the width of a rectangular filter which passes the same total energy and has equal peak transmission to the filter of interest. Dashed lines show the 95% CI limits for the ERB.

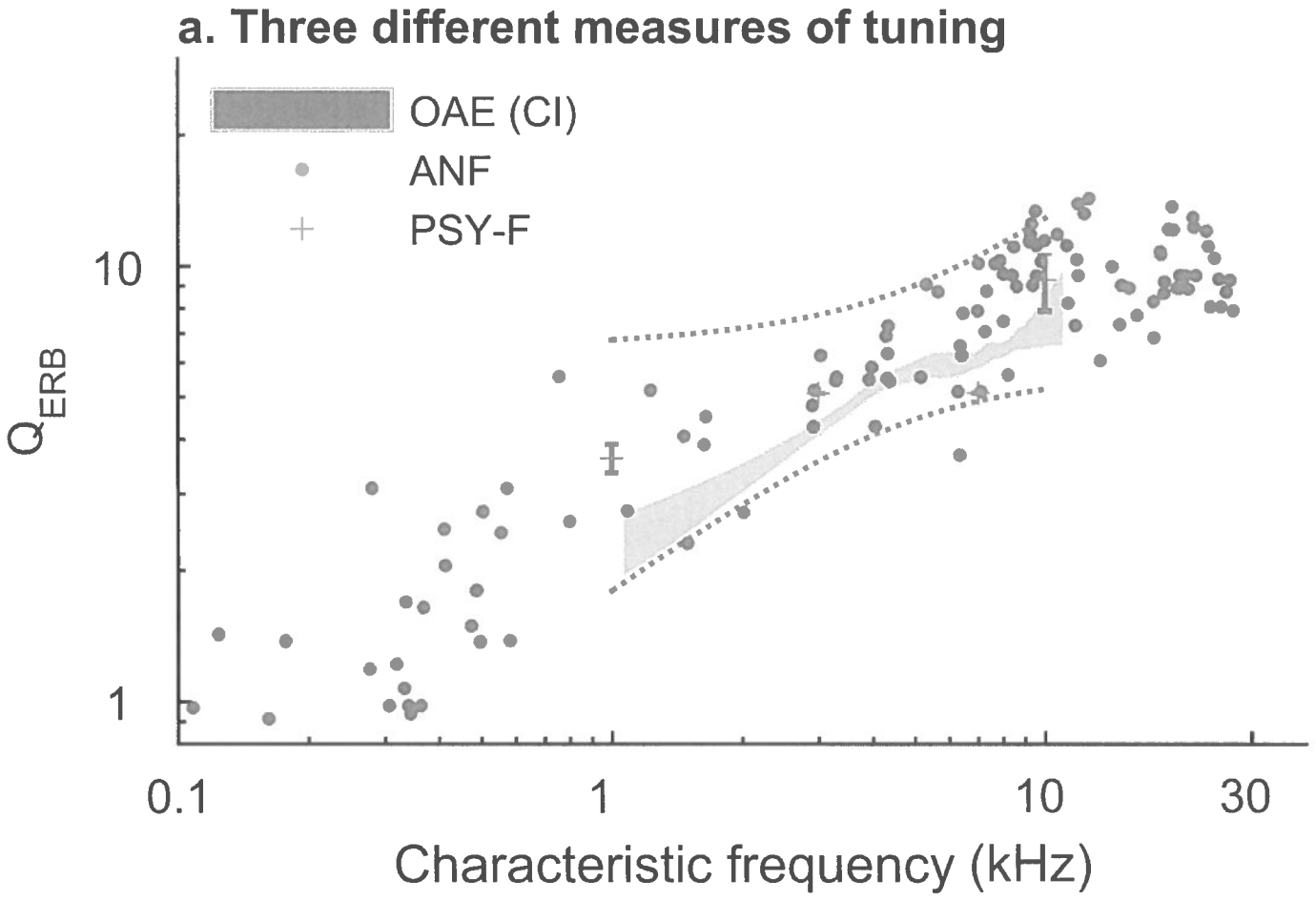


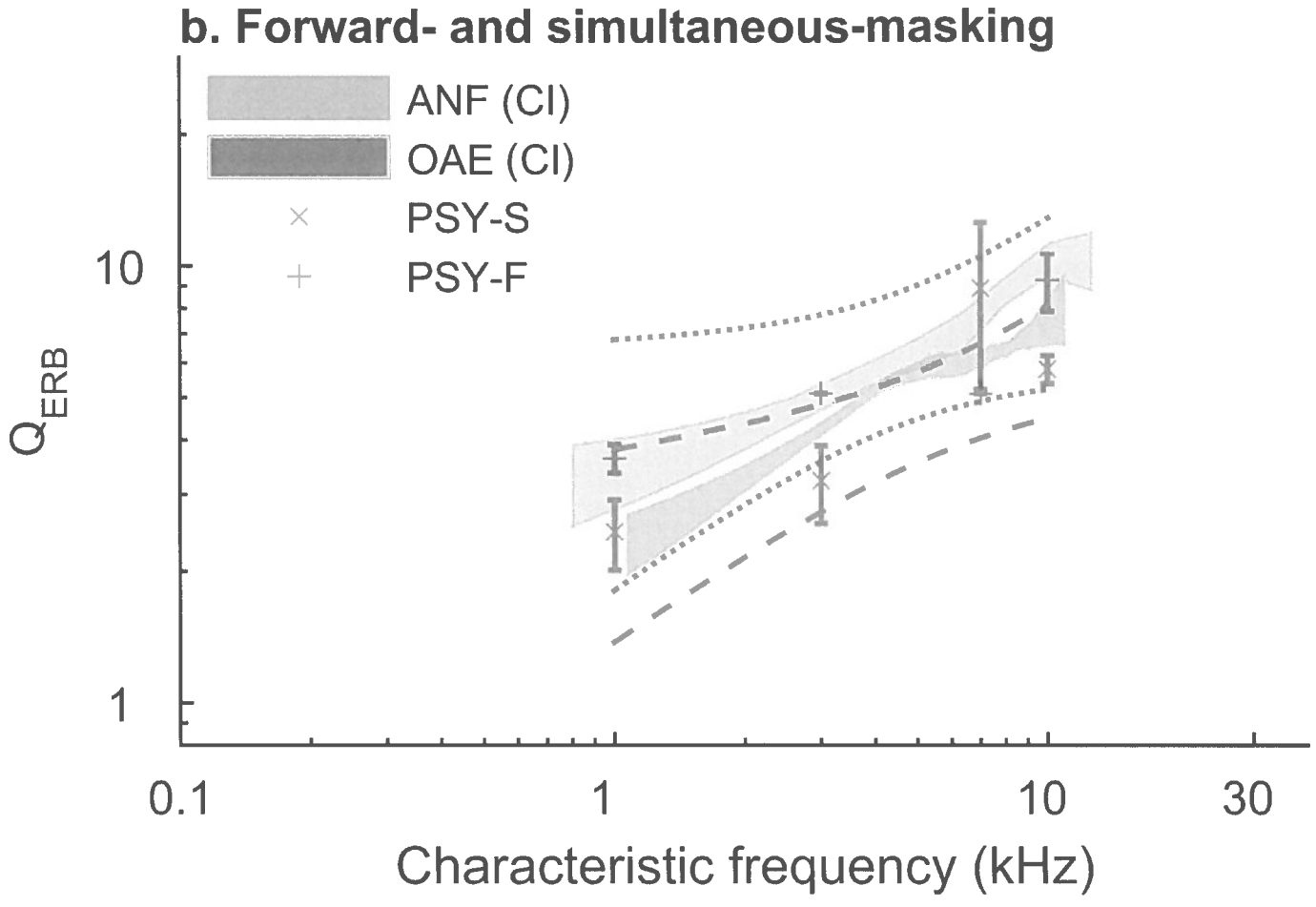
Supplementary Figure S2 | Otoacoustic emission estimates of filter sharpness. Estimates of cochlear tuning obtained from SFOAEs (light red points). **a**. OAE delays expressed as the number of periods at the stimulus frequency. Dashed black lines show a broken stick function used to estimate the apical-basal transition frequency (see methods). **b**. Q_{ERB} values predicted using the delays plotted in panel **a**. Solid and dashed lines show the loess trend-line and its 95% CIs respectively (the loess trend-line is a non-parametric, locally-weighted regression method for fitting a smooth line to scatterplot data; 17). Grey points show auditory-nerve fiber tuning for reference. The SFOAE delay data (and thus the corresponding estimates of Q_{ERB} obtained via multiplication by tuning ratio) show considerable scatter about the trend due to mechanisms of SFOAE generation thought to be unrelated to variations in cochlear tuning (i.e., the micromechanical irregularities that reflect the traveling wave). The most extreme outliers are plotted using dark red dots (see statistical methods).

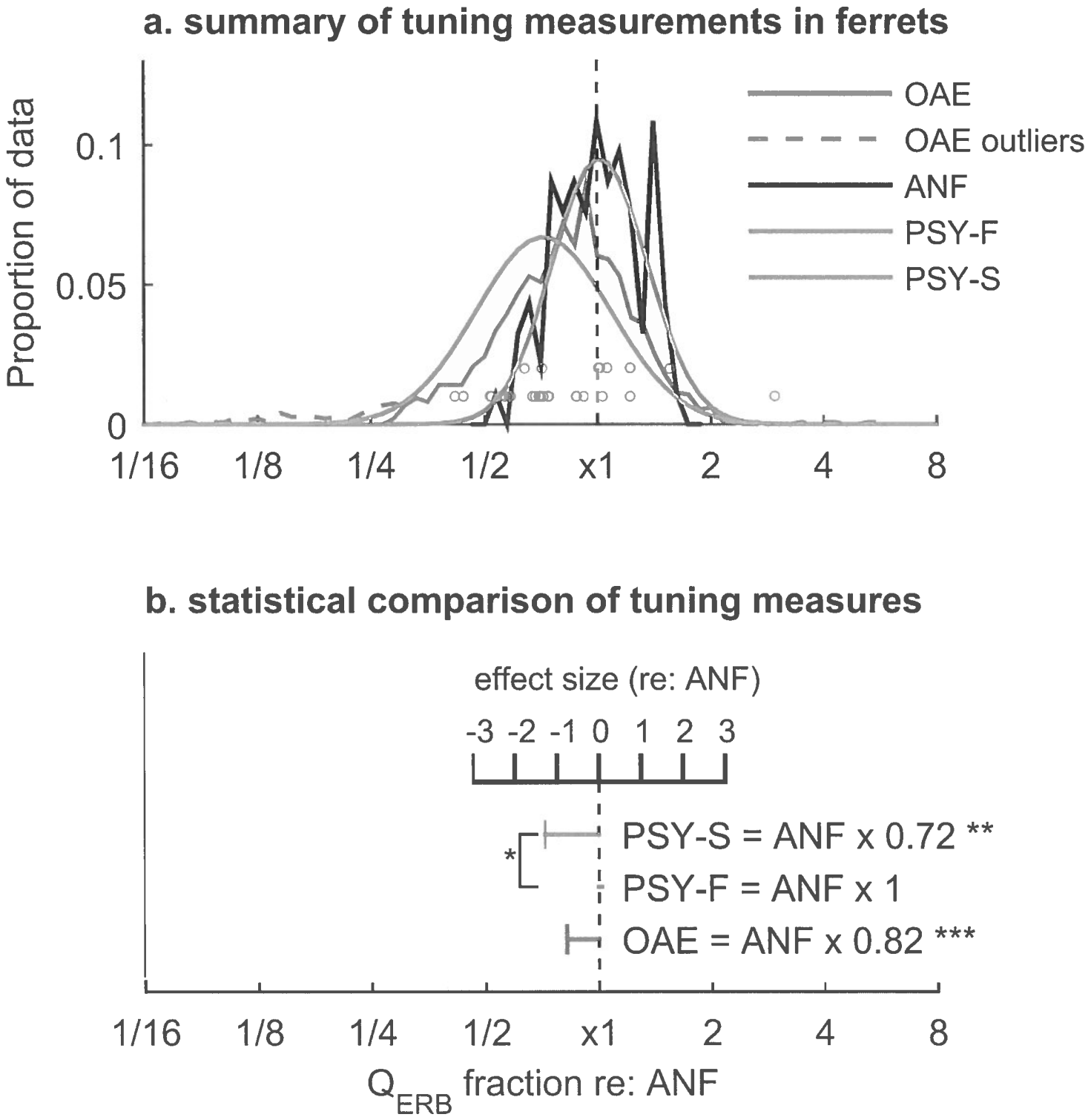


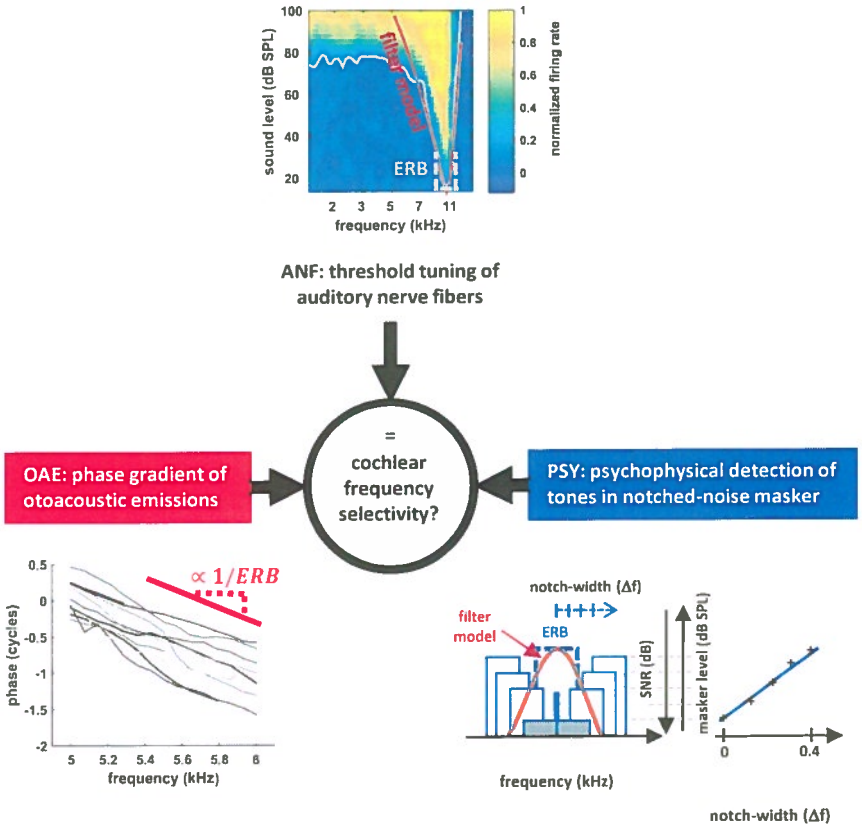
Supplementary Figure S3 | Estimates of frequency selectivity in ferrets are consistent across all measurements, and qualitatively poorer than available estimates in humans. This figure summarizes the linear model analysis of ferret data (blue points) shown in the results, expressing the mean of the residuals and 95% CIs for each group either as a fraction of the auditory-nerve filter sharpness (lower scale) or the effect size (upper axis), which is calculated as these values normalized to the standard deviation of the residuals for the auditory-nerve data. The dark grey vertical shading indicates 95% CI for the auditory-nerve fiber tuning, and the pale grey vertical shading indicates the s.d. (corresponding to an effect size of 1). In addition to the psychophysical data fitted to the *roex(pr)* model (see methods) described in the results, we show estimates of tuning derived from perceptual forward masking using two alternative analytical filter models: a single-parameter symmetrical *roex(p)* model and a three-parameter asymmetric *roex(2pr)* model. Although these were not chosen for the main analysis, due to their inferior fitting to ferret data, both alternative models produce similar tuning estimates and neither model affects our conclusions concerning the similarity of estimated bandwidths between different methods or the difference in tuning between ferrets and humans. Previously collected estimates of tuning from human (red points), based on

otoacoustic emissions (OAE) and perceptual masking (Oxenham and Shera 2003), suggest that human tuning is qualitatively sharper than in ferrets. Data at 4 kHz, collected using the same lateralization method employed in ferrets and fitted using the *roex(2pr)* model, for direct comparison of perceptual estimates, are also consistent with this conclusion ($p<0.001$). Asterisks next to data points indicate where significant differences exist compared to auditory-nerve tuning, * $p<0.05$; ** $p<0.01$; *** $p<0.001$. The additional asterisks accompanying the square bracket refer to the comparison between human and ferret perceptual bandwidths, fitted to the *roex(2pr)* model.


a. Three different measures of tuning
OAE (CI)
ANF
PSY-F
10
1
Q_{ERB}
0.1
1
10
30
Characteristic frequency (kHz)


b. Forward- and simultaneous-masking
ANF (CI)
OAE (CI)
PSY-S
PSY-F
10
1
Q_{ERB}
0.1
1
10
30
Characteristic frequency (kHz)


a. summary of tuning measurements in ferrets
Proportion of data
0.1
0.05
0
OAE
OAE outliers
ANF
PSY-F
PSY-S
1/16
1/8
1/4
1/2
x1
2
4
8
b. statistical comparison of tuning measures
effect size (re: ANF)
-3 -2 -1 0 1 2 3
*
PSY-S = ANF x 0.72 **
PSY-F = ANF x 1
OAE = ANF x 0.82 ***
1/16
1/8
1/4
1/2
x1
2
4
8
Q ERB fraction re: ANF


sound level (dB SPL)
100
80
60
40
20
2 3 5 7 11
frequency (kHz)
filter model
ERB
normalized firing rate
1
0.8
0.6
0.4
0.2
0
ANF: threshold tuning of auditory nerve fibers
=
cochlear frequency selectivity?
OAE: phase gradient of otoacoustic emissions
PSY: psychophysical detection of tones in notched-noise masker
phase (cycles)
0.5
0
-0.5
-1
-1.5
-2
5 5.2 5.4 5.6 5.8 6
frequency (kHz)
$\propto 1/ERB$
notch-width (Δf)
filter model
ERB
SNR (dB)
frequency (kHz)
masker level (dB SPL)
0
0.4
notch-width (Δf)