1   Pitch perception is adapted to species-specific cochlear filtering

2

3   Kerry MM Walker[1*], Ray Gonzalez[2], Joe Kang[1], Josh H McDermott[2,3¶] & Andrew J

4   King[1¶]

5   ¶Joint senior authorship

6   *Corresponding author (kerry.walker@dpag.ox.ac.uk)

7

8   [1] Department of Physiology, Anatomy & Genetics, University of Oxford

9   [2] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

10   [3] Program in Speech and Hearing Biosciences and Technology, Harvard University

11

12

13

14

15

16   Keywords: pitch, ferret, human, psychophysics, auditory nerve, cochlea, resolved

17   harmonics, hearing

18

19   **Abstract**

20   Pitch perception is critical for recognizing speech, music and animal vocalizations,

21   but its neurobiological basis remains unsettled, in part because of divergent results

22   from different species. We used a combination of behavioural measurements and

23   cochlear modelling to investigate whether species-specific differences exist in the

24   cues used to perceive pitch and whether these can be accounted for by differences in

25   the auditory periphery. Ferrets performed a pitch discrimination task well whenever

26   temporal envelope cues were robust, but not when resolved harmonics only were

27   available. By contrast, human listeners exhibited the opposite pattern of results on an

28   analogous task, consistent with previous studies. Simulated cochlear responses in the

29   two species suggest that the relative salience of the two types of pitch cues can be

30   attributed to differences in cochlear filter bandwidths. Cross-species variation in pitch

31   perception may therefore reflect the constraints of estimating a sound's fundamental

32   frequency given species-specific cochlear tuning.

33

34

Introduction

Many of the sounds in our environment are periodic, and the rate at which such sounds repeat is known as their fundamental frequency, or F0. We perceive the F0 of a sound as its pitch, and this tonal quality is one of the most important features of our listening experience. The way that F0 changes encodes meaning in speech [1] and musical melody [2–4]. The F0 of a person's voice provides a cue to their identity [5–7] and helps us attend to them in a noisy environment [8–10].

The vocal calls of non-human animals are also often periodic, and pitch is believed to help them to identify individuals and interpret communication calls [11,12]. Many mammalian species have been shown to discriminate the F0 of periodic sounds in experimental settings [13–17], and these animal models hold promise for understanding the neural mechanisms that underlie pitch perception. However, pitch acuity can differ markedly across species [16,18], raising the possibility that humans and other mammals may use different neural mechanisms to extract pitch.

The auditory cortex plays a key role in pitch processing, but it remains unclear how cortical neurons carry out the necessary computations to extract the F0 of a sound [19]. Neural correlates of F0 cues [20–22] and pitch judgments [23] have been observed across auditory cortical fields in some species, while a specialized pitch centre has been described in marmoset auditory cortex [24]. There is similar a lack of consensus regarding the neural code for pitch in the human brain [25]. A better understanding of the similarities and differences in pitch processing across species is essential for interpreting neurophysiological results in animals and relating them to human pitch perception.

59   Pitch discrimination in humans is driven by two acoustical cues that result

60 from low-numbered 'resolved' harmonics and high-numbered 'unresolved' harmonics

61 [26], and the relative importance of these cues offers a means to compare pitch

62 mechanisms across species.. In the frequency domain, F0 can be determined from the

63 distribution of harmonics (Fig. 1A, upper panel) [27–29]. In the auditory nerve, the

64 frequency spectrum is represented as a "place code" of activation across the tonotopic

65 map as well as a "time code" of spikes that are phase-locked to the basilar membrane

66 vibrations [30,31]. However, both these representations are limited by the cochlea's

67 frequency resolution [27]. Because cochlear filter bandwidths increase with

68 frequency, only low-numbered harmonics produce discernible peaks of excitation and

69 phase locked spikes at their centre frequency (Fig. 1A, middle panel). Such harmonics

70 are said to be "resolved". By contrast, high-numbered harmonics are not individually

71 resolved, and instead produce beating in time at the F0, conveyed by phase-locking to

72 their envelope [32] (Fig. 1A, bottom panel). For convenience and to be consistent with

73 prior literature, we refer to these unresolved pitch cues as "temporal" cues, cognizant

74 that the representation of resolved harmonics may also derive from a temporal neural
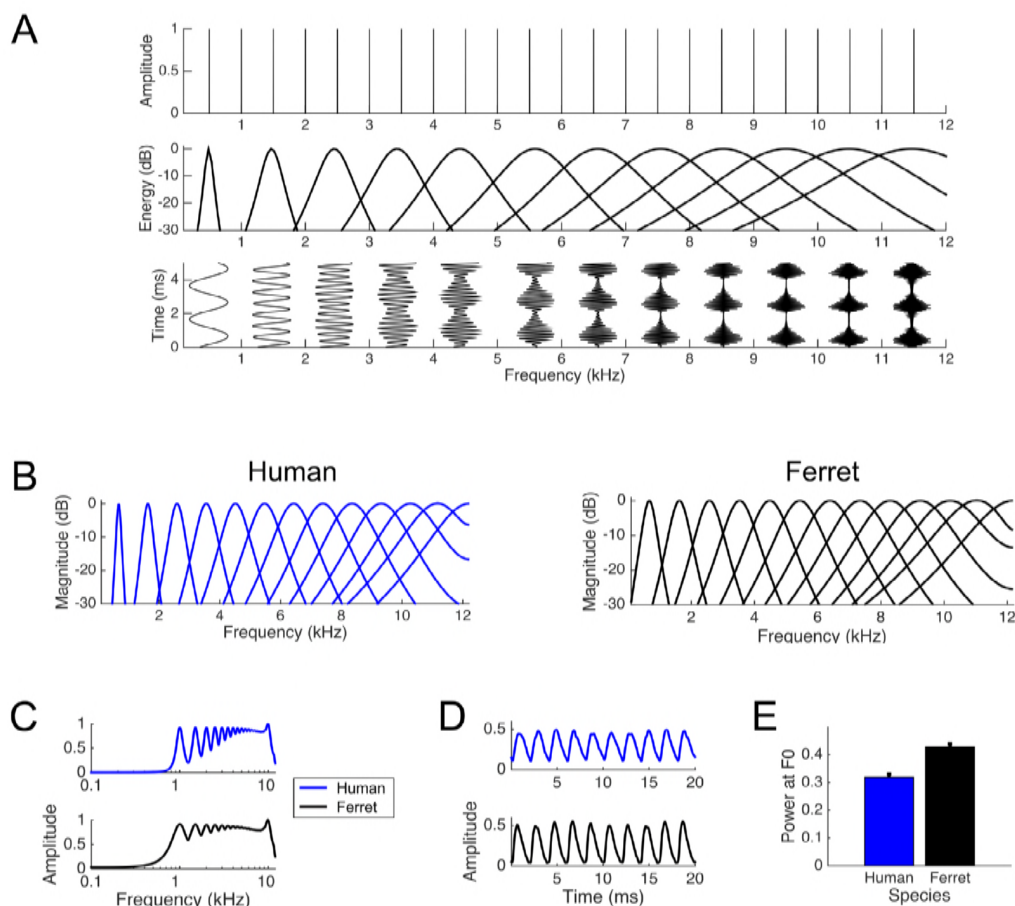
75 code.

Figure 1: Simulated cochlear filters and their responses to a 500 Hz harmonic complex tone. **A.** Illustration of the role of unresolved and resolved harmonics in periodicity encoding. Upper plot: Amplitude spectrum for a tone complex that contains all harmonics of 500 Hz from 0 – 11.5 kHz. This sound will evoke a pitch corresponding to 500 Hz. Middle plot: Cartoon of the cochlear filters centred on every second harmonic of 500 Hz, based on data from Glasberg and Moore [41]. This illustrates that lower harmonics are resolved, while the cochlear filters corresponding to higher order harmonics respond to multiple harmonic components in the tone. Lower plot: The output of each of these cochlear filter is plotted throughout 5 ms of the tone complex. The resolved harmonics phase lock to the frequency of one harmonic, while unresolved harmonics beat at the sum of multiple harmonic components (i.e. 500 Hz), providing an explicit temporal representation of F0. **B-E:** A computational model of the cochlear filter bank was used to simulate representations of complex sounds in the ferret and human auditory nerve. Data are color-coded for the ferret (black) and human (blue). **B.** The frequency tuning of 15 example auditory nerve fibres is shown for the simulated human (left) and ferret (right) cochlea. C-E. Analyses of the responses of human and ferret cochlear filter banks to a 500 Hz tone complex. **C.** The response strengths of each of 500 auditory nerve fibres to a 500-Hz complex tone were averaged across the duration of the sound, and plotted across the full range of centre frequencies. Most harmonics produce clearly resolvable activation peaks across fibres in the human cochlea (upper plot), but only the lower harmonics are resolved in the ferret cochlea (lower plot). **D.** The temporal profile of the output of one simulated auditory nerve fibre with a centre frequency of 5000 Hz is shown for the human (upper plot) and ferret (lower plot) cochlea. **E.** The power at 500 Hz in the output of each frequency filter, averaged across the full duration of the tone complex, calculated for the human (blue) and ferret (black) auditory nerve. For each species, these values were normalized by the maximal F0 power across all channels. The plot shows the mean (+ standard error) normalized power at F0 across all auditory nerve fibres.

76

77

78   Although psychophysical experiments have demonstrated that humans can

79   extract F0 using either resolved harmonics or unresolved harmonics alone [33–35],

80   pitch perception is generally dominated by resolved harmonics [34,36]. Marmosets

81   can also use resolved harmonics to detect F0 changes [37], whereas rodents (i.e.

82   gerbils and chinchillas) rely upon temporal periodicity cues [38–40]. Why resolved

83   harmonics are more important in humans is unknown, but this could relate to the

84   availability of pitch cues following cochlear filtering. The growing evidence that

85   cochlear bandwidths are broader in many other species [41–43] raises the possibility

86   that they might process pitch cues in different ways from humans.

87   The behavioural studies carried out to date are difficult to compare across

88   species. First, pitch in humans is defined as the percept through which sounds are

89   ordered on a scale from low to high [44]. By contrast, animal studies often measure

90   change detection in a go/no-go task, from which it is difficult to determine whether

91   they experience a comparable ordered pitch percept or whether they are responding to

92   a change in the perceived pitch as opposed to some aspect of timbre. A two-

93   alternative forced choice (2AFC) task requiring "low" and "high" judgements

94   analogous to those used in human psychophysical tasks would better enable cross-

95   species comparisons [16], but has yet to be employed to examine the use of resolved

96   and unresolved cues in animals. Second, the spectral range of stimuli was not fully

97   controlled across F0 in previous studies (e.g. [16,37]), making it possible for animals

98   to base their behavioural choices on the lower spectral edge of the sounds, rather than

99   the sound's overall F0. Finally, most animal studies [17,37,40] have not directly

100   compared performance across human and non-human species on an equivalent task,

101   so differences in task demands might therefore account for any apparent species

102   differences. For example, the pitch difference thresholds of ferrets can differ by

103   orders of magnitude between a go/no-go and 2AFC task [45].

104   The present study overcomes these limitations by directly comparing the pitch

105   cues used by humans and ferrets on a common 2AFC pitch classification task. We

106   first use a computational model to simulate the representation of periodic sounds in

107   the inner ear. The simulations generated predictions about the availability of

108   periodicity cues in the auditory nerve of each species. We then tested these

109   predictions by training ferrets and humans to classify the pitch of a harmonic complex

110   tone,. We find differences in their dependence on resolved and unresolved harmonics,

111   which can be accounted by differences in cochlear tuning between ferrets and

112   humans.

113

Results

Simulating the filtering of tones in the ferret and human cochlea

116   Humans are believed to have narrower cochlear filter bandwidths than ferrets and

117   other non-human animals [17,41–43,46–49], and these physiological constraints may

118   predispose them to rely on different acoustical cues to classify the pitch of complex

119   tones. Specifically, individual auditory nerve fibres are believed to respond to a

120   narrower range of frequency in humans than in ferrets, which should result in more

121   resolvable harmonics across the human tonotopic map. On the other hand, if the

122   bandwidth of an auditory nerve fibre is broader, its firing should phase lock more

123   strongly to the beating that results from adjacent harmonics, potentially providing a

124   stronger explicit representation of the temporal periodicity of F0 in ferrets than in

125   humans.

126 To investigate this hypothesis, we modified a standard model of the cochlear

127 filter bank [50] to simulate the representation of tones along the human and ferret

128 basilar membrane. The output of each cochlear filter was half-wave rectified,

129 compressed (by raising the rectified output to the 0.7 power), and lowpass filtered at

130 3kHz to simulate the transformation of basilar membrane motion into spiking output

131 in the auditory nerve. The existing literature guided the design of this model [50–52]

132 and parameters in the model were derived from either human psychophysics [41] or

133 ferret auditory nerve recordings [48].

134 As shown in Figure 1B, the cochlear filters are wider for the ferret auditory

135 nerve than the human. In Figure 1C-E, we compare the human and ferret simulated

136 responses to a 500-Hz missing F0 tone complex that we used as a training sound in

137 our ferret behavioural experiment (described below).

138 When the instantaneous power of the cochlear filters is summed across the

139 duration of the sound and plotted as a function of centre frequency, the individual

140 harmonics of the tone are more clearly resolved in the human cochlea than in the

141 ferret (Fig. 1C). This takes the form of deeper troughs in the activation of nerve fibres

142 whose centre frequencies lie between the harmonic components of the sound. To

143 visualize the temporal representation of the same stimulus, we plotted the output of a

144 single nerve fibre (here, a fibre with a centre frequency of 5 kHz) throughout time

145 (Fig. 1D). In this case, the representation of the 500 Hz F0 is clearer in the ferret – the

146 human cochlea produces weaker temporal modulation because fewer harmonics fall

147 within the fibre's bandwidth.

148 We also examined whether the temporal representation of F0 was enhanced in

149 the ferret cochlea across the full range of frequency filters. A Fourier transform was

150 performed on the output of each fibre throughout a 200ms steady-state portion of the

151   sound. The power of the response at F0 was then expressed as a proportion of the

152   overall power for that fibre. The results of this metric averaged across all fibres in the

153   model are shown in Fig. 1E. The average temporal representation of F0 was enhanced

154   in the ferret compared to the human (Wilcoxon rank sum test; z = 8.286, p = 1.175 x

155   10^{-16}). In fact, this F0 representation metric was higher in the ferret than the human

156   cochlear model across every pair of individually simulated auditory nerve fibres.

157   These simulations suggest that the ferret cochlea provides an enhanced

158   representation of the envelope periodicity of a complex tone, as conveyed by spikes

159   that are phase-locked to the F0 in the auditory nerve. On the other hand, the human

160   auditory nerve provides a better resolved representation of individual harmonics

161   across the tonotopic array. It might thus be expected that these two types of cues

162   would be utilized to different extents by the two species.
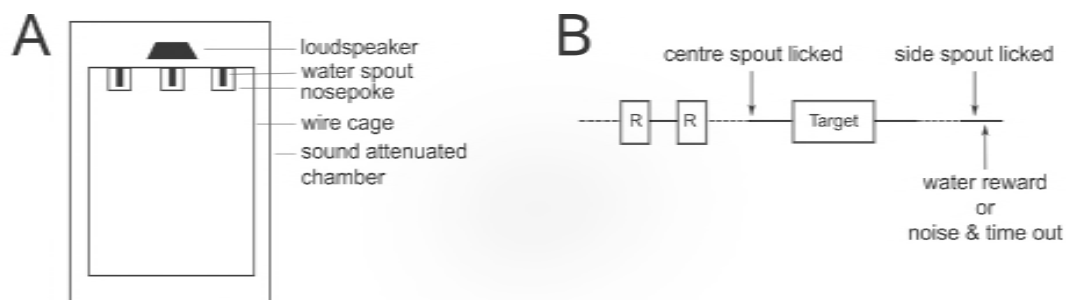
Figure 2: Psychophysical task design. **A.** Schematic of ferret testing apparatus, viewed from above. **B.** Schematic of one trial in the 2-alternative forced choice pitch classification task. The target tone could be lower or higher in F0 than the reference tone (R). Dotted lines indicate time durations that are variable,
163   depending on the animal's behaviour.

164

165   *Behavioural measures of pitch cue use in ferrets*

166   To test the role of different pitch cues in ferret pitch perception, we trained five

167   animals on a two-alternative forced choice (2AFC) task that requires "low" and

168   "high" pitch judgements analogous to those used in human psychophysical tasks (Fig.

169   2A,B). On each trial, a harmonic complex tone was presented at one of two possible

170   fundamental frequencies. Ferrets were given water rewards for responding at the right

171   nose-poke port for a high F0, and at the left port for a low F0. Incorrect responses

172   resulted in a time-out. We began by training four ferrets to classify harmonic complex

173   tones with an F0 of 500 and 1000Hz, with a repeating pure tone presented at 707Hz

174   (the midpoint on a logarithmic scale) for reference before each trial. Two of these

175   animals, along with one naïve ferret, were then trained on the same task using target

176   F0 values of 150 and 450Hz and a 260Hz pure tone reference. In both cases, the

177   harmonics of the low and high stimuli to be discriminated were matched in spectral

178   bandwidth, so that ferrets could not solve the task based on the frequency range of the

179   sound (Fig. 3; left column). Rather, the animals had to discriminate sounds based on

180   some cue to the F0. After completing several pre-training stages to habituate the

181   animals to the apparatus and sound presentation (see Methods), the ferrets learned to

182   perform the pitch classification task within 22 ± 3 (mean $\pm$ standard deviation) days
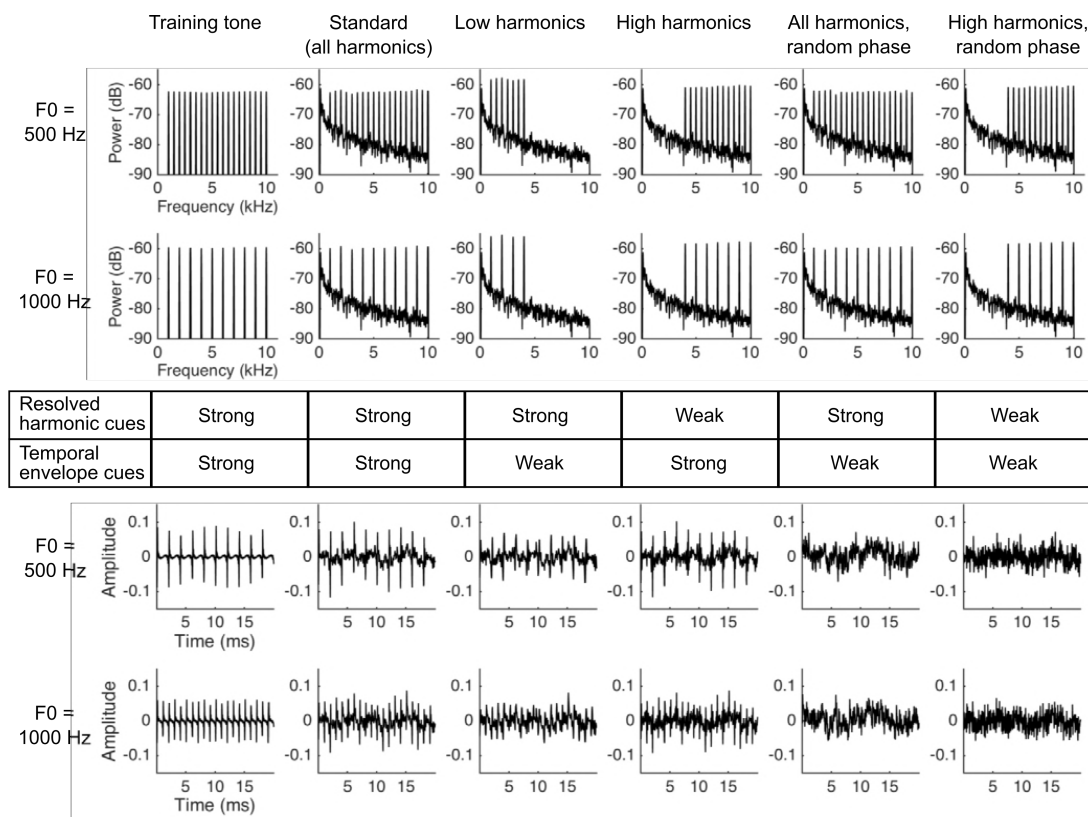
183   of twice daily training.

Figure 3: Stimuli used in the 707 Hz reference pitch classification task. Plots show the training tone (left column), standard stimulus (second column) and 4 probe stimuli (columns 3-6) used in the psychophysical task. Target stimuli either had an F0 of 500 or 1000 Hz, indicated to the left of each row of plots. The top two rows show the power spectra of each target sound, while the bottom 2 rows plot the temporal envelope of the sound throughout 20 ms. The table in the middle of the figure indicates whether resolved harmonic (above) and temporal envelope (below) F0 cues are preserved in each stimulus.

185   Once the ferrets learned to perform this simple 2AFC task, we incorporated

186   "probe trials" into the task in order to determine which acoustical cues they were

187   using to categorize the trained target sounds. Probe trials made up 20% of trials in a

188   given session, and were randomly interleaved with the "standard" trials described

189   above. On probe trials, an untrained stimulus was presented, and the ferret received a

190   water reward regardless of its behavioural choice. This task design discouraged ferrets

191   from learning to use a different strategy to classify the probe sounds.

192   The inner ear is known to produce distortion in response to harmonic tones

193   that can introduce energy at the fundamental frequency to the basilar membrane

194   response, even for missing-fundamental sounds [53]. These distortion products could

195   in principle counter our attempts to match the spectral bandwidths of the sounds,

11

196   since they could cause the lowest frequency present in the ear to differ as a function

197   of F0. To determine if the ferrets relied on such cochlear distortion products to

198   classify tones in our task, we added pink noise to the stimulus on 20% of randomly

199   interleaved probe trials at an intensity that is known to be more than sufficient to

200   mask cochlear distortion products in humans [54,55]. Ferrets performed more poorly

201   on probe trials than on standard trials (paired t-test; t = 4.346, p = 0.005), as expected

202   for an auditory discrimination task performed in noise. However, they continued to

203   perform the pitch classification at 71.85% ± 9.60% correct (mean ± standard

204   deviation) with the noise masker, which is well above chance (1-sample t-test; t =

205   6.025, p = 0.001). This suggests that ferrets did not rely on cochlear distortion

206   products to solve our task.

207   We next moved to the main testing stage of our behavioural experiment,

208   which aimed to determine if ferrets use resolved harmonics, temporal envelope

209   periodicity, or both of these cues to identify the F0 of tones. All tone complexes, both

210   the standard and probe stimuli, were superimposed on a pink noise masker. Our

211   auditory nerve model (above) allowed us to estimate which harmonics in the tone

212   complexes would be resolved in the ferret auditory nerve (Fig. 4A) [56]. This analysis

213   suggests that our standard tones contained both resolved and unresolved harmonics

214   for ferret listeners, as intended. We constructed four types of probe stimuli based on

215   our resolvability estimates: (1) "Low Harmonic" tones containing only harmonics that

216   we expected to be resolved; (2) "High Harmonic" tones containing harmonics

217   presumed to be less well resolved; (3) "All Harmonics Random Phase" probes

218   containing the full set of harmonics present in the standard tone, but whose phases

219   were independently randomized in order to flatten the temporal envelope; and (4)

220   "High Harmonics Random Phase" stimuli with the same randomization of harmonic

221   phases, but containing only presumptively unresolved harmonics. The spectral ranges

222   of these stimuli are given in Figure 4B, and the spectra and audio waveforms

223   (showing the temporal envelope periodicity) of the 500 and 1000 Hz stimuli are

224   illustrated in Figure 3A. Ferrets were again given water rewards irrespective of their

225   behavioural choice on probe trials, in order to avoid reinforcing different pitch

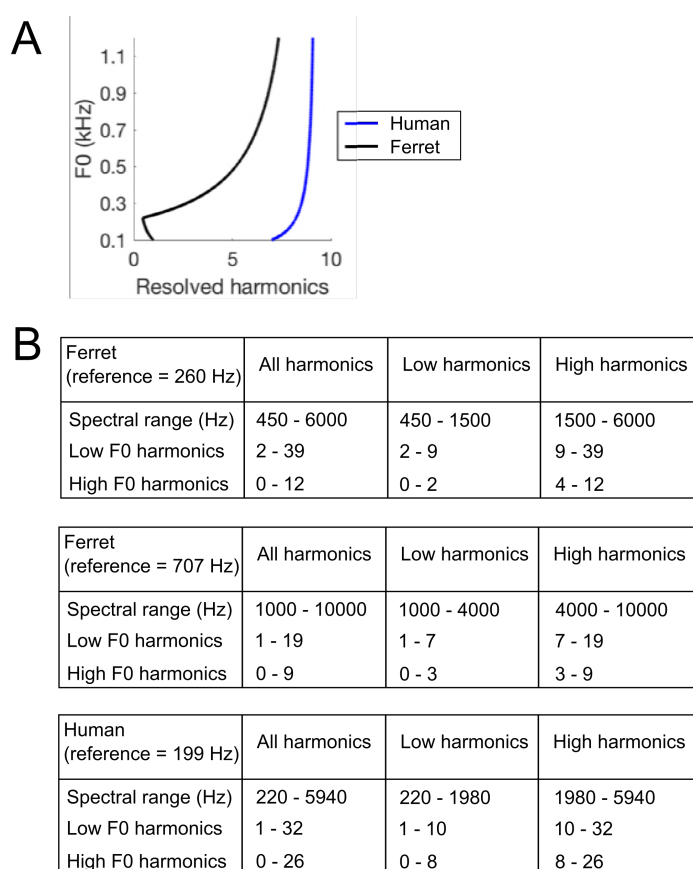226   classification strategies across probe stimuli.

Figure 4: Harmonic content of stimuli. **A.** The number of resolved harmonics was estimated over a range of F0s, for ferret (black line) and human (blue line) cochlea. **B.** The frequency ranges and numbers of harmonic partials included in each stimulus. In this table, the F0 is deemed to be "0", and further harmonics are counted from 1 onwards.

Ferret (reference = 260 Hz)	All harmonics	Low harmonics	High harmonics
Spectral range (Hz)	450 - 6000	450 - 1500	1500 - 6000
Low F0 harmonics	2 - 39	2 - 9	9 - 39
High F0 harmonics	0 - 12	0 - 2	4 - 12

Ferret (reference = 707 Hz)	All harmonics	Low harmonics	High harmonics
Spectral range (Hz)	1000 - 10000	1000 - 4000	4000 - 10000
Low F0 harmonics	1 - 19	1 - 7	7 - 19
High F0 harmonics	0 - 9	0 - 3	3 - 9

Human (reference = 199 Hz)	All harmonics	Low harmonics	High harmonics
Spectral range (Hz)	220 - 5940	220 - 1980	1980 - 5940
Low F0 harmonics	1 - 32	1 - 10	10 - 32
High F0 harmonics	0 - 26	0 - 8	8 - 26

227

228   The performance of ferrets on the standard and probe stimuli is shown in

229   Figure 5A. A repeated-measures 3-way ANOVA indicated that performance varied

230   with stimulus type (i.e., the standard and 4 probe stimuli) (F = 10.540, p = 0.003), but

231   not across subjects (F = 1.060; p = 0.391) or the two reference conditions (i.e., 260

232   and 707 Hz) (F = 0.438, p = 0.576). Scores did not significantly vary across

233   individual ferrets in either the 260 Hz (2-way ANOVA; F = 0.366, p = 0.704) or 707

13

234   Hz condition (2-way ANOVA; F = 2.063, p = 0.158), so data collected from the same

235   animals in these two conditions were treated as independent measurements.
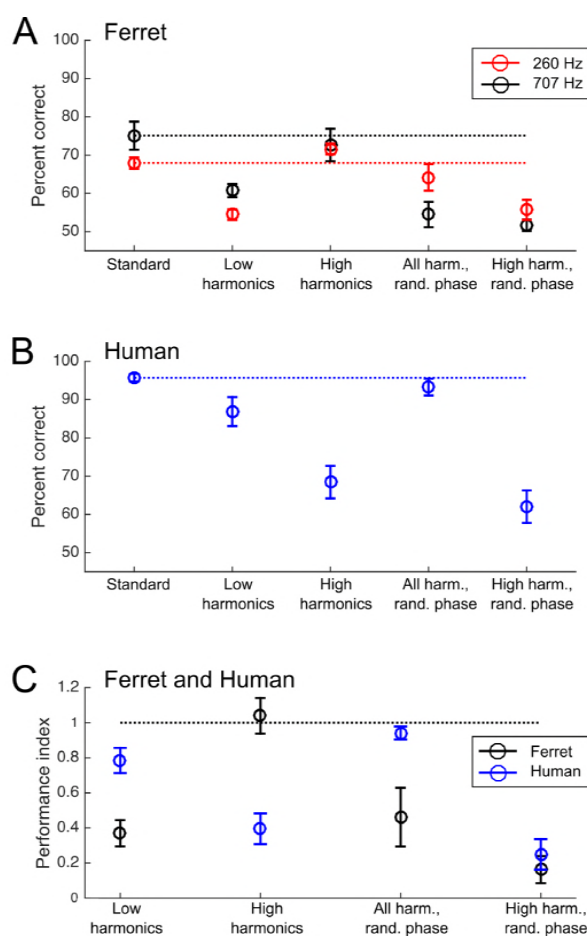

Figure 5: Pitch classification performance of ferrets and humans. **A.** Ferrets' percent correct scores on the pitch classification task are plotted for the standard tone trials (left) and each of 4 probe stimuli (right). The results of testing with the 260 Hz reference (150 and 450 Hz targets; red) and 707 Hz reference (500 and 1000 Hz targets; black) are plotted separately. **B.** Humans' pitch classification performance is plotted, as in (A). **C.** Performance for each of 4 probe stimuli is expressed as the ratio of the percentage correct score and that achieved with the standard training tone stimulus. Data are shown for ferrets (black) and humans (blue). Values of 0 indicate that subjects performed at chance for the probe stimulus, while 1 indicates that they classified the F0 of the probe as accurately as the training stimulus. Error bars show mean ± SEM. Individual data for (A) and (B) are shown in Supp Figure 1.

237   To assess the acoustical cues used by animals to solve the pitch classification

238   task, we compared ferrets' performance on the standard trials with that on each of the

239   four probe trial types (repeated measures 2-way ANOVA, Tukey's HSD test). Ferrets

240   showed impaired performance on probes that contained only low harmonics (p =

241   0.001), but performed as well as on standard trials when only high harmonics were

242   presented (p = 1.000). Their performance was also impaired when we randomized the

243   phases of the high-harmonics (p = 0.002). Phase randomization also impaired

244   performance when the full set of harmonics (both resolved and unresolved) were

245   present (p = 2.173 x 10^{-5}). This pattern of results suggests that ferrets rely more

246   strongly on the temporal envelope periodicity (produced by unresolved harmonics)

14

247   than on resolved harmonics to classify the pitch of tones, unlike what would be

248   expected for human listeners.

249

250   *Comparison of human and ferret pitch classification performance*

251   Humans were trained on a similar pitch classification task to the one described for

252   ferrets in order to best compare the use of pitch cues between these two species.

253   Participants were presented with harmonic complex tones and classified them as high

254   or low. A training phase was used to teach participants the high and low F0s.

255   We tested human listeners using the same types of standard and probe stimuli

256   as in the final stage of ferret testing described above. As the pitch discrimination

257   thresholds of human listeners are known to be superior to those of ferrets [16], we

258   adapted the target F0s (180 and 220 Hz) and harmonic cut-offs for human hearing

259   (Fig. 4). The between-species comparison of interest here is therefore not the

260   difference in absolute scores on the task, but the pattern of performance across probe

261   conditions.

262   Human listeners also showed varied pitch classification performance across

263   the standard and probe stimuli (repeated-measures 2-way ANOVA; $F = 36.999$, $p =$

264   1.443×10^{-15}). However, a different pattern of performance across stimuli was

265   observed for human subjects (Fig. 5B). Tukey's HSD tests indicated that human

266   listeners were significantly impaired when resolved harmonics were removed from

267   the sounds, as demonstrated by impairments in the "High Harmonic" probes with ($p =$

268   9.922×10^{-9}) and without ($p = 1.029 \times 10^{-8}$) randomized phases. Conversely, no

269   impairment was observed when resolved harmonics were available, regardless of

270   whether the phases of stimuli were randomized ("All Harmonics Random Phase"

271   condition; $p = 0.959$) or not ("Low Harmonics" condition; $p = 0.101$). These results

15

272   are all consistent with the wealth of prior work on human pitch perception, but

273   replicate previously reported effects in a task analogous to that used in ferrets.

274   The performance for each probe type relative to performance on the standard

275   stimuli, is directly compared between the two species in Figure 5C. Here, a score of 1

276   indicates that the subject performed equally well for the standard tone and the probe

277   condition, while a score of 0 indicates that the probe condition fully impaired their

278   performance (reducing it to chance levels). This comparison illustrates the differences

279   in acoustical cues underlying ferret and human pitch classifications. As our model

280   simulations predicted, we found that while ferrets were impaired only when temporal

281   envelope cues from unresolved harmonics were disrupted, humans continued to

282   classify the target pitch well in the absence of temporal envelope cues, so long as

283   resolved harmonics were present. This was confirmed statistically as a significant

284   interaction between species and probe type on performance (repeated measures 3-way

285   ANOVA; F = 14.802, p = 3.412 x 10^{-9}). The two species thus appear to

286   predominantly rely on distinct cues to pitch.

287

288   **Discussion**

289   We used a combination of cochlear modelling and behavioural experiments to

290   examine the use of pitch cues in ferrets and human listeners. Our model simulations

291   illustrated how broader cochlear filter widths in ferrets result in fewer resolved

292   harmonics and a more enhanced representation of temporal envelopes than the human

293   cochlea. Based on this result, we predicted that the pitch judgments of ferrets would

294   rely more strongly on temporal envelope cues than that of human listeners. Our

295   behavioural experiments directly compared the use of pitch cues in the two species

296   and found that this is indeed the case. Our results provide the first unambiguous

16

297   dissociation of pitch mechanisms across species, by utilizing the same task across

298   species, and provide an illustration of the potential consequences of species

299   differences in cochlear tuning.

300

301   *Findings in other species*

302   Human listeners have consistently been found to have better pitch discrimination

303   thresholds when stimuli contain resolved harmonics [34–36,57,58]. Moreover,

304   cortical responses to pitched sound in humans are stronger for resolved than

305   unresolved harmonics, mirroring perceptual sensitivity [59,60]. The results of our

306   human experiments are thus fully consistent with this large body of prior work, while

307   enabling comparison with non-human animals. Because most natural sounds contain

308   both low- and high-numbered harmonics, humans may learn to derive pitch primarily

309   from resolved harmonics even when temporal envelope cues are also available, and

310   are thus less equipped to derive pitch from unresolved harmonics alone. This would

311   explain the drop in performance when resolved harmonic cues were removed on

312   probe trials in our experiment.

313   Our cochlear simulations suggest that harmonic resolvability is worse for ferrets

314   than human listeners, so they may conversely learn to rely more on temporal pitch

315   cues when estimating pitch from natural sounds, leading to poorer performance for

316   low harmonic tone complexes. Many non-human mammals are believed to have

317   wider cochlear bandwidths than humans [42,43,61,62], and so we might expect

318   temporal cues to dominate their pitch decisions as we have observed in ferrets. The

319   few studies to directly address F0 cue use in pitch judgments by non-human animals

320   have raised the possibility of species differences in pitch perception, but have relied

321   on go/no-go tasks that differ from standard psychophysical tasks used in humans. For

322   instance, studies in gerbils suggest that they primarily use temporal cues to detect an

323   inharmonic component in a tone complex [38,39]. Chinchillas were similarly shown

324   to detect the onset of a periodic sound following a non-periodic sound using temporal,

325   rather than resolved harmonic, cues [40,63]. While these studies did not explicitly

326   compare the use of resolved and unresolved pitch cues, they are consistent with our

327   findings regarding the importance of temporal cues in non-human species.

328   Marmosets, on the other hand, appear to use the phase of harmonic components to

329   detect changes in the F0 of a repeating tone complex only when resolved harmonics

330   are omitted from the stimulus [17,37]. This suggests that temporal cues are only

331   salient for this species when they occur in unresolved harmonics. Similarly to

332   humans, marmosets were found to detect smaller changes in F0 when harmonics were

333   resolved than when only unresolved harmonics were available [37]. Comparable

334   studies have yet to be carried out in other non-human primates, so it remains unclear

335   whether primates are special in the animal kingdom in their dependence on resolved

336   harmonic cues. We note also that the behavioural task used in previous marmoset

337   experiments [17,37] required animals to detect a change in F0, whereas the task

338   employed in this study required ferrets to label the direction of F0 changes. Ferrets

339   show an order of magnitude difference in pitch acuity on these two tasks [45], raising

340   the possibility that primates might as well.

341   The use of probe trials without feedback in the present experiment allowed us to

342   determine which acoustical cues most strongly influenced listeners' pitch judgements.

343   The ferrets relied predominantly on temporal cues under these conditions, but our

344   results do not preclude the possibility that they could also make pitch judgments

345   based on resolved harmonics if trained to do so. Indeed, although human listeners rely

346   on resolved harmonics under normal listening conditions, we can also extract pitch

18

347 from unresolved harmonics when they are isolated [34,36,57]. Our simulations show

348 that up to 8 harmonics are resolved on the ferret cochlea, depending on the F0 (Fig.

349 4A). Consequently, if specifically trained to do so, one might expect ferrets to be able

350 to derive F0 from these harmonics using the same template matching mechanism

351 proposed for human listeners [27,29]. It is also important to note that the relationship

352 between harmonic resolvability and auditory nerve tuning is not fully understood, and

353 nonlinearities in response to multiple frequency components could cause resolvability

354 to be worse than that inferred from isolated auditory nerve fibre measurements.

355　　Overall, the available evidence fits with the idea that pitch judgments are adapted

356 to the acoustical cues that are available and robust in a particular species, with

357 differences in cochlear tuning thus producing cross-species diversity in pitch

358 perception. A similar principle may be at work in human hearing, since listeners rely

359 on harmonicity for some pitch tasks and spectral changes in others, potentially

360 because of task-dependent differences in the utility of particular cues [7]. The

361 application of normative models of pitch perception will likely provide further insight

362 into the relative importance of these cues.

363

364 *Implications for neurophysiological work*

365　　A better understanding of the similarities and differences in pitch processing

366 across species is essential for relating the results of neurophysiological studies in

367 animals to human pitch perception. The present experiments suggest that ferrets, a

368 common animal model in studies of hearing (e.g. [23,64–67]), can estimate F0 from

369 the temporal envelopes of harmonic complex tones. Our data indicate that ferrets

370 generalize across sounds with different spectral properties (including wideband

371 sounds, sounds in noise, and sounds containing only high harmonics) without relying

19

372   on explicit energy at the F0. In this respect, ferrets appear to have a pitch percept,

373   even though the cues underlying it are apparently weighted differently than in human

374   pitch perception.

375   The existing literature might be taken to suggest that primates are the most

376   appropriate animal models for examining the role of resolved harmonics in human

377   pitch perception, as they appear to be more like humans in their use of this cue

378   [17,37]. On the other hand, our data suggest that ferrets are a powerful animal model

379   for evaluating temporal models of pitch extraction (e.g. [50,68]). Like the ferret,

380   cochlear implant users also have poor spectral resolution at the cochlea, and

381   consequently these devices are severely limited in their ability to represent resolved

382   harmonics. Using species such as the ferret to better understand the neural basis of

383   temporal pitch processing could provide insight into why current implants produce

384   impoverished pitch perception [69], and how they might be improved in the future.

385

386   **Materials and Methods**

387   **EXPERIMENTAL SUBJECTS**

388   *Ferrets (Mustela putorius furo)*

389   Five adult female pigmented ferrets (aged 6 – 24 months) were trained in this study.

390   Power calculations estimated that 5 animals was the minimum appropriate sample

391   size for 1-tailed paired comparisons with alpha = 5%, a medium (0.5) effect size, and

392   beta = 20%. Ferrets were housed in groups of 2-3, with free access to food pellets.

393   Training typically occurred in runs of 5 consecutive days, followed by two days rest.

394   Ferrets could drink water freely from bottles in their home boxes on rest days. On

395   training days, drinking water was received as positive reinforcement on the task, and

396   was supplemented as wet food in the evening to ensure that each ferret received at

397   least 60 ml/kg of water daily. Regular otoscopic and typanometry examinations were

398   carried out to ensure that the animals' ears were clean and healthy, and veterinary

399   checks upon arrival and yearly thereafter confirmed that animals were healthy. The

400   animal procedures were approved by the University of Oxford Committee on Animal

401   Care and Ethical Review and were carried out under license from the UK Home

402   Office, in accordance with the Animals (Scientific Procedures) Act 1986.

403   *Humans*

404   The pitch classification performance of 16 adult humans (9 male, ages 18-53 years;

405   mean age = 25.3 years) was also examined, which provided a 60% beta in the power

406   calculations described for ferrets. All subjects reported having normal hearing. All

407   experimental procedures on humans were approved by the Committee on the Use of

408   Humans as Experimental Subjects at MIT.

409

410   METHOD DETAILS

411   *Cochlear filter simulations*

412   We used a cochlear filter bank previously developed by Patterson et al. [50] and

413   implemented by Slaney [70] to simulate representations of sounds on the basilar

414   membrane. The model simulates the response of the basilar membrane to complex

415   sounds through two processing modules: (a) a set of parallel Gammatone filters, each

416   with a different characteristic frequency and Equivalent Rectangular Bandwidth

417   (ERB), produces a simulation of basilar membrane motion in response to the sound,

418   and (b) a two-dimensional adaptation mechanism as observed in hair cell physiology.

419   In order to compare the representation of harmonic tone complexes in the human and

420   ferret cochlea, we modified this model to use filter constants derived from either

421   psychophysical estimates of human cochlear filters [41], or ferret auditory nerve

422   recordings [48]. Based on these sources, the equivalent rectangular bandwidth of filter

423   i in the human cochlea was calculated as:

424   $ERB_i = f_i / (12.7 * (f_i/1000)^{0.3})$,

425   where f_i is the centre frequency of the filter in Hz.

426   For the ferret cochlea, the equivalent rectangular bandwidth of each filter was

427   estimated using the following linear fit to the data in Sumner and Palmer [48]:

428   $ERB_i = f_i / 8.9047 + 209.6149$.

429   The output of each channel in the above Gammatone filter bank was half-wave

430   rectified and then compressed (to the power of 0.7) to simulate transduction of sound

431   by inner hair cells. Finally, the output was low-pass filtered at 3kHz to reflect the

432   spike rate limit of auditory nerve fibres. This model architecture is similar to that used

433   in previous studies (e.g. [51,52]).

434   *Training apparatus*

435   Ferrets were trained to discriminate sounds in custom-built testing chambers,

436   constructed from a wire mesh cage (44 x 56 x 49 cm) with a solid plastic floor, placed

437   inside a sound-insulated box lined with acoustic foam to attenuate echoes. Three

438   plastic nose poke tubes containing an inner water spout were mounted along one wall

439   of the cage: a central "start spout" and two "response spouts" to the left and right (Fig.

440   2A). Ferrets' nose pokes were detected by breaking an infrared LED beam across the

441   opening of the tube, and water was delivered from the spouts using solenoids. Sound

442   stimuli, including acoustic feedback signals, were presented via a loudspeaker (FRS

443   8; Visaton, Crewe, UK) mounted above the central spout, which had a flat response

444   (±2 dB) from 0.2 – 20 kHz. The behavioural task, data acquisition, and stimulus

445   generation were all automated using a laptop computer running custom Matlab (The

446 Mathworks, Natick, MA, USA) code, and a real-time processor (RP2; Tucker-Davis

447 Technologies, Alachua, FL, USA).

Pre-training

449 Ferrets ran two training sessions daily, and typically completed 94 ± 24 trials per

450 session (mean ± standard deviation). Several pre-training stages were carried out to

451 shape animals' behaviour for our classification task. In the first session, animals

452 received a water reward whenever they nose poked at any of the spouts. Next, they

453 received water rewards only when they alternated between the central and peripheral

454 spouts. The water reward presented from the peripheral response spouts (0.3 - 0.5 ml

455 per trial) was larger than that presented at the central start spout (0.1 - 0.2 ml per

456 trial). The animal was required to remain in the central nose poke for 300 ms to

457 receive a water reward from that spout.

458   Once animals performed this task efficiently, sound stimuli were introduced in

459 the next session. At the start of each trial, a repeating pure tone "reference" (200 ms

460 duration, 200 ms inter-tone interval, 60 dB SPL) was presented to indicate that the

461 central spout could be activated. Nose poking at the central spout resulted in the

462 presentation of a repeating complex tone "target" (200 ms duration, 200 ms inter-tone

463 interval, 70 dB SPL) after a 100 ms delay. The animal was again required to remain at

464 the centre for 300 ms, and early releases now resulted in the presentation of an "error"

465 broadband noise burst (200 ms duration, and 60 dB SPL) and a 3 s timeout before a

466 new trial began. The target tone could take one of two possible F0 values, which

467 corresponded to rewards at one of the two peripheral spouts (right rewards for high F0

468 targets, and left for low F0s). For all training and testing stages, the target tones

469 contained harmonics within the same frequency range, so that animals could not use

470 spectral cut-offs to classify the sounds. The target tone continued to play until the

471   animal responded at the correct peripheral spout, resulting in a water reward. Once the

472   animals could perform this final pretraining task with >70% accuracy across trials,

473   they advanced to pitch classification testing.

474   *Testing stages and stimuli*

475   The complex tone target was presented only once per trial, and incorrect peripheral

476   spout choices resulted in an error noise and a 10 s timeout (Fig. 2B). After such an

477   error, the following trial was an error correction trial, in which the F0 presented was

478   the same as that of the previous trial. These trials were included to discourage ferrets

479   from always responding at the same peripheral spout. If the ferret failed to respond at

480   either peripheral spout for 14 s after target presentation, the trial was restarted.

481   The reference pure tone's frequency was set to halfway between the low and

482   high target F0s on a log scale. We examined ferrets' pitch classification performance

483   using two pairs of complex tone targets in separate experimental blocks: the first with

484   F0s of 500 and 1000 Hz (707 Hz reference), and the second with 150 and 450 Hz

485   targets (260 Hz reference). Four ferrets were trained on the 707 Hz reference. Two of

486   these animals, plus an additional naive animal, were trained on the 260 Hz reference.

487   In each case, testing took place over 3 stages, in which the ferret's task remained the

488   same but a unique set of stimulus parameters was changed (Fig. 3 and 4), as outlined

489   below. Ferrets were allocated to the 260 and 707 Hz reference conditions based on

490   their availability at the time of testing.

491   *Stage 1:* Target sounds were tone complexes, containing all harmonics within

492   a broad frequency range (specified in Fig. 4B). When an animal performed this task

493   >75% correct on 3 consecutive sessions, (32.8 ± 7.1 sessions from the beginning of

494   training; mean $\pm$ standard deviation; n = 4 ferrets), they moved to Stage 2.

495   *Stage 2:* On 80% of trials, the same standard target tones from Stage 1 were

496   presented. The other 20% of trials were "probe trials", in which the ferret was

497   rewarded irrespective of the peripheral spout it chose, without a timeout or error

498   correction trial. Probe trials were randomly interleaved with standard trials. The probe

499   stimuli differed only by the addition of pink noise (0.1-10 kHz) to the target sounds,

500   in order to mask possible cochlear distortion products at F0. The level of the noise

501   masker was set so that the power at the output of a Gammatone filter centred at the F0

502   (with bandwidth matched to ferret auditory nerve measurements in that range [48])

503   was 5dB below the level of the pure tone components of the target. This is

504   conservative because distortion products are expected to be at least 15 dB below the

505   level of the stimulus components [54,55]. When an animal performed this task >75%

506   correct on 3 consecutive sessions, they moved to stage 3.

507   *Stage 3:* The probe stimulus from Stage 2 served as the "Standard" sound on

508   80% of trials, and all stimuli (both the standard and probes) included the pink noise

509   masker described above. Twenty percent of trials were probe trials, as in Stage 2, but

510   this stage contained tones manipulated to vary the available pitch cues. We estimated

511   the resolvability of individual harmonics using ERB measurements available in

512   previously published auditory nerve recordings [48]. For a given F0, the number of

513   resolved harmonics was approximated as the ratio of F0 and the bandwidth of

514   auditory nerve fibres with a characteristic frequency at that F0, as described by Moore

515   and Ohgushi [56], and applied by Osmanski et al. [17]. This measure yielded between

516   1 and 8 resolved harmonics for ferrets, depending on the F0 (Fig. 4A). Four types of

517   probe stimuli were presented: (1) "Low Harmonics", which contained only harmonics

518   presumed to be resolved; (2) "High Harmonics", comprised of harmonics presumed to

519   be unresolved; (3) "All Harmonics Random Phase", which contained the same set of

520   harmonics as the standard, but whose phases were independently randomized in order

521   to reduce temporal envelope cues for pitch; and (4) "High Harmonics Random

522   Phase", which contained the harmonics present in "High Harmonics" stimuli, but with

523   randomized phases. The bandpass cutoffs for the probe stimuli were chosen so that

524   the "Low Harmonic", but not "High Harmonic", probes contained resolved harmonics

525   for ferret listeners. Each probe stimulus was presented on at least 40 trials for each

526   ferret, while the standard was tested on over 1000 trials per ferret.

527   *Human psychophysical task*

528   Human subjects were tested on a pitch classification task that was designed to be as

529   similar as possible to Stage 3 of ferrets' task (see above). Target F0s of 180 and 220

530   Hz were tested on 16 subjects.

531   In the psychophysical task, human listeners were presented with the same

532   classes of stimuli described above for ferrets. The frequency ranges included in the

533   probe stimuli are listed in Fig. 4B. Sounds were presented over headphones

534   (Sennheiser HD280) in a sound attenuated booth (Industrial Acoustics, USA). A

535   repeating reference pure tone (200 ms duration, 200 ms inter-tone interval, 60 dB

536   SPL) was presented at the start of a trial, and the subject initiated the target harmonic

537   tone complex (200 ms duration, 70 dB SPL) presentation with a keypress. Text on a

538   computer monitor then asked the subject whether the sound heard was the low or high

539   pitch, which the subjects answered via another keypress (1 = low, 0 = high). Feedback

540   was given on the monitor after each trial to indicate whether or not the subject had

541   responded correctly. Incorrect responses to the standard stimuli resulted in

542   presentation of a broadband noise burst (200 ms duration, and 60 dB SPL) and a 3 s

543   timeout before the start of the next trial. Error correction trials were not used for

544   human subjects, as they did not have strong response biases. Standard harmonic

26

545   complex tones were presented on 80% of trials, and the 4 probes ("Low Harmonics",

546   "High Harmonics", "All Harmonics Random Phase", and "High Harmonics Random

547   Phase") were presented on 20% of randomly interleaved trials. Feedback for probe

548   trials was always "correct", irrespective of listeners' responses. Humans were given

549   10 practice trials with the standard stimuli before testing, so that they could learn

550   which stimuli were low and high, and how to respond with the keyboard. Each probe

551   stimulus was tested on 40 trials for each subject, while the standard was tested on 680

552   trials per subject.

553

554   QUANTIFICATION AND STATISTICAL ANALYSIS

555   *Psychophysical data analysis*

556   Error correction trials were excluded from all data analysis, as were data from any

557   testing session in which the subject scored less than 60% correct on standard trials. T-

558   tests and ANOVAs with an alpha of 5% were used throughout to assess statistical

559   significance, where the n indicates the number of subjects per group. Error bars in

560   Figures 1 and 5 show mean ± standard errors. Further details of all statistical tests

561   described here are provided as supplementary tables.

562   Because humans produced higher percent correct scores overall than ferrets on

563   the behavioural task, we normalized probe scores against the standard scores when

564   directly comparing performance between species. The score of each species in each

565   probe condition was represented as:

566    $Pnorm_{ai} = (P_{ai} - 50) / (S_a - 50)$,

567   where *Pnorm* is the normalized probe score for species *a* on probe *i*, P_{ai} is the percent

568   correct score for species *a* on probe *i*, and S_a is the percent correct score of species *a*

569   on the standard trials. If the performance of species *a* is unimpaired for a given probe

570   stimulus i relative to the standard stimulus, then $Pnorm_{ai}$ will equal 1. If the listeners

571   are completely unable to discriminate the F0 of the probe, then $Pnorm_{ai} = 0$.

572   The data and custom software developed in this manuscript are available on

573   the Dryad archive.

574

575

584   **Declaration of Interests**

585   The authors and funding bodies have no competing financial interests in the outcomes

586   of this research.

References

1. Ohala, J.J. (1983). Cross-Language Use of Pitch: An Ethological View. Phonetica *40*, 1–18.

2. Dowling, W.J., and Fujitani, D.S. (1971). Contour, interval, and pitch recognition in memory for melodies. J Acoust Soc Am *49*, 524.

3. Krumhansl, C.L. (1990). Cognitive Foundations of Musical Pitch (New York: Oxford University Press).

4. Cousineau, M., Demany, L., and Pressnitzer, D. (2009). What makes a melody: The perceptual singularity of pitch sequences. J Acoust Soc Am *126*, 3179–3187.

5. Gelfer, M.P., and Mikos, V.A. (2005). The relative contributions of speaking fundamental frequency and formant frequencies to gender identification based on isolated vowels. J Voice *19*, 544–554.

6. Latinus, M., and Belin, P. (2011). Human voice perception. Curr Biol *21*, R143–R145.

7. McPherson, M.J., and McDermott, J.H. (2018). Diversity in pitch perception revealed by task dependence. Nat Hum. Beh *2*, 1–15.

8. Darwin, C.J. (2005). Pitch and Auditory Grouping. In Pitch: Neual Coding and Perception, C. J. Plack, R. R. Fay, A. J. Oxenham, and A. N. Popper, eds. (New York: Springer-Verlag), pp. 278–305.

9. Miller, S.E., Schlauch, R.S., and Watson, P.J. (2010). The effects of fundamental frequency contour manipulations on speech intelligibility in background noise. J Acoust Soc Am *128*, 435–443.

10. Popham, S., Boebinger, D., Ellis, D.P.W., Kawahara, H., and McDermott, J.H. (2018). Inharmonic speech reveals the role of harmonicity in the cocktail

612   party problem. Nat Comm *9*, 2122.

613 11. Nelson, D.A. (1989). Song frequency as a cue for recognition of species and

614   individuals in the field sparrow (Spizella pusilla). J Comp Psychol *103*,

615   171–176.

616 12. Koda, H., and Masataka, N. (2002). A pattern of common acoustic

617   modification by human mothers to gain attention of a child and by

618   macaques of others in their group. Psychol Rep *91*, 421–422.

619 13. Heffner, H., and Whitfield, I.C. (1976). Perception of the missing

620   fundamental by cats. J Acoust Soc Am *59*, 915–919.

621 14. Tomlinson, R.W., and Schwarz, D.W. (1988). Perception of the missing

622   fundamental in nonhuman primates. J Acoust Soc Am *84*, 560–565.

623 15. Shofner, W.P., Yost, W.A., and Whitmer, W.M. (2007). Pitch perception in

624   chinchillas (Chinchilla laniger): Stimulus generalization using rippled

625   noise. J Comp Psychol *121*, 428–439.

626 16. Walker, K.M.M., Schnupp, J.W.H., Hart-Schnupp, S.M.B., King, A.J., and

627   Bizley, J.K. (2009). Pitch discrimination by ferrets for simple and complex

628   sounds. J Acoust Soc Am *126*, 1321–1335.

629 17. Osmanski, M.S., Song, X., and Wang, X. (2013). The role of harmonic

630   resolvability in pitch perception in a vocal nonhuman primate, the

631   common marmoset (Callithrix jacchus). J Neurosci *33*, 9161–9168.

632 18. Shofner, W.P. (2005). Comparative Aspects of Pitch Perception. In Pitch:

633   Neual Coding and Perception Springer Handbook of Auditory Research., C.

634   J. Plack, A. J. Oxenham, R. R. Fay, and A. N. Popper, eds. (New York:

635   Springer Science+Business Media, Inc.), pp. 56–98.

636 19. Wang, X., and Walker, K.M.M. (2012). Neural mechanisms for the

637   abstraction and use of pitch information in auditory cortex. J Neurosci *32*,

638   13339–13342.

639 20. Bizley, J.K.K., Walker, K.M.M.M., Silverman, B.W.W., King, A.J.J., and

640   Schnupp, J.W.H.W. (2009). Interdependent encoding of pitch, timbre, and

641   spatial location in auditory cortex. J Neurosci *29*, 2064–2075.

642 21. Bendor, D., and Wang, X. (2010). Neural coding of periodicity in marmoset

643   auditory cortex. J Neurophysiol *103*, 1809–1822.

644 22. Fishman, Y.I., Micheyl, C., and Steinschneider, M. (2013). Neural

645   representation of harmonic complex tones in primary auditory cortex of

646   the awake monkey. J Neurosci *33*, 10312–10323.

647 23. Bizley, J.K., Walker, K.M.M., Nodal, F.R., King, A.J., and Schnupp, J.W.H.

648   (2013). Auditory cortex represents both pitch judgments and the

649   corresponding acoustic cues. Curr Biol *23*, 620–625.

650 24. Bendor, D., and Wang, X. (2005). The neuronal representation of pitch in

651   primate auditory cortex. Nature *436*, 1161–1165.

652 25. Griffiths, T.D., and Hall, D.A. (2012). Mapping pitch representation in

653   neural ensembles with fMRI. J Neurosci *32*, 13343–13347.

654 26. Moore, B.C.J., and Gockel, H.E. (2011). Resolvability of components in

655   complex tones and implications for theories of pitch perception. Hear Res

656   *276*, 88–97.

657 27. Goldstein, J.L. (1974). An optimum processor theory for the central

658   formation of the pitch of complex tones. J Acoust Soc Am *55*, 1061–1069.

659 28. Terhardt, E. (1974). Pitch, consonance, and harmony. J Acoust Soc Am *55*,

660   1061–1069.

661 29. Shamma, S., and Klein, D. (2000). The case of the missing pitch templates:

662   how harmonic templates emerge in the early auditory system. J Acoust Soc

663   Am *107*, 2631–2644.

664   30.   Cariani, P.A., and Delgutte, B. (1996). Neural correlates of the pitch of

665   complex tones. I. Pitch and pitch salience. J Neurophysiol *76*, 1698–1716.

666   31.   Schnupp, J.W.H., Nelken, I., and King, A.J. (2011). Auditory neuroscience :

667   making sense of sound (MIT Press).

668   32.   Schouten, J.F. (1970). The residue revisited. In Frequency Analysis and

669   Periodicity Detection in Hearing, R. Plomp and G. F. Smoorenburg, eds.

670   (Leiden: Sijthoff), pp. 41–58.

671   33.   Houtsma, A.J.M., and Smurzynski, J. (1990). Pitch identification and

672   discrimination for complex tones with many harmonics. J Acoust Soc Am

673   *87*, 304–310.

674   34.   Shackleton, T.M., and Carlyon, R.P. (1994). The role of resolved and

675   unresolved harmonics in pitch perception and frequency modulation

676   discrimination. J Acoust Soc Am *95*, 3529–3540.

677   35.   Bernstein, J.G., and Oxenham, A.J. (2003). Pitch discrimination of diotic and

678   dichotic tone complexes: harmonic resolvability or harmonic number? J

679   Acoust Soc Am *113*, 3323–33234.

680   36.   Ritsma, R.J. (1967). Frequencies dominant in the perception of the pitch of

681   complex sounds. J Acoust Soc Am *42*, 191–198.

682   37.   Song, X., Osmanski, M.S., Guo, Y., and Wang, X. (2016). Complex pitch

683   perception mechanisms are shared by humans and a New World monkey.

684   Proc Natl Acad Sci U S A *113*, 781–786.

685   38.   Klinge, A., and Klump, G.M. (2009). Frequency difference limens of pure

686   tones and harmonics within complex stimuli in Mongolian gerbils and

687 humans. J Acoust Soc Am *125*, 304–314.

688 39. Klinge, A., and Klump, G.M. (2010). Mistuning detection and onset

689 asynchrony in harmonic complexes in Mongolian gerbils. J Acoust Soc Am

690 *128*, 280–290.

691 40. Shofner, W.P., and Chaney, M. (2013). Processing pitch in a nonhuman

692 mammal (Chinchilla laniger). J Comp Psychol *127*, 142–153.

693 41. Glasberg, B.R., and Moore, B.C.J. (1990). Derivation of auditory filter

694 shapes from notched-noise data. Hear Res *47*, 103–138.

695 42. Shera, C.A., Guinan, J.J., and Oxenham, A.J. (2002). Revised estimates of

696 human cochlear tuning from otoacoustic and behavioral measurements.

697 Proc Natl Acad Sci U S A *99*, 3318–3323.

698 43. Joris, P.X., Bergevin, C., Kalluri, R., Mc Laughlin, M., Michelet, P., van der

699 Heijden, M., and Shera, C.A. (2011). Frequency selectivity in Old-World

700 monkeys corroborates sharp cochlear tuning in humans. Proc Natl Acad

701 Sci U S A *108*, 17516–17520.

702 44. Anon (1994). American National Standard Acoustical Terminology

703 (American National Standards Inst.).

704 45. Walker, K.M.M., Bizley, J.K., King, A.J., and Schnupp, J.W.H. (2011). Cortical

705 encoding of pitch: Recent results and open questions. Hear Res *271*, 74–87.

706 46. Pickles, J.O., and Comis, S.D. (1976). Auditory-nerve-fiber bandwidths and

707 critical bandwidths in the cat. J Acoust Soc Am *60*, 1151–1156.

708 47. Kittel, M., Wagner, E., and Klump, G.M. (2002). An estimate of the auditory-

709 filter bandwidth in the Mongolian gerbil. Hear Res *164*, 69–76.

710 48. Sumner, C.J., and Palmer, A.R. (2012). Auditory nerve fibre responses in
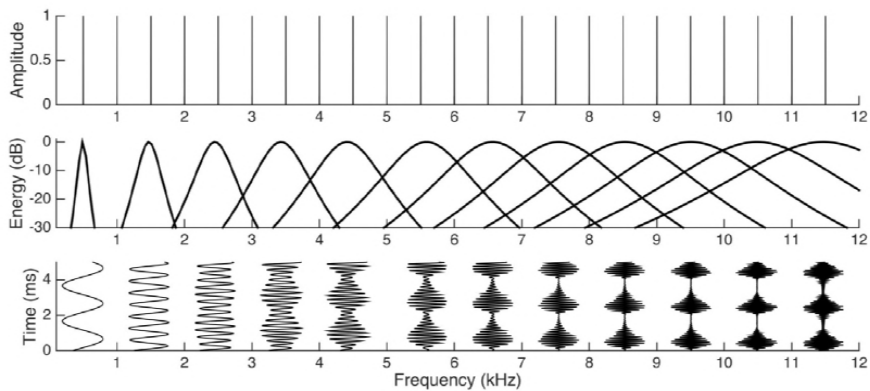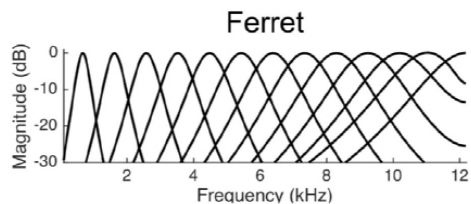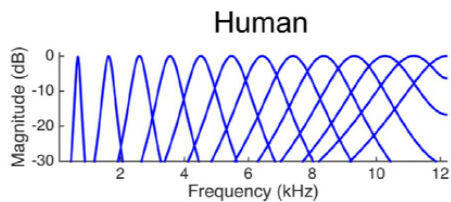
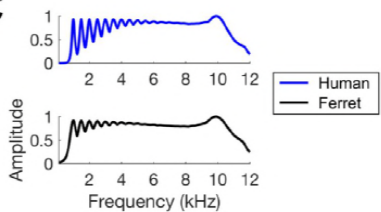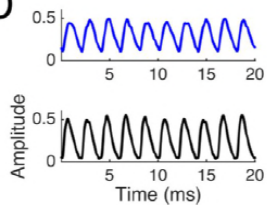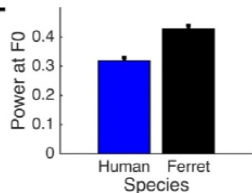711 the ferret. Eur J Neurosci *36*, 2428–2439.

712   49.   Alves-pinto, A., Sollini, J., Wells, T., Sumnerqjf, C.J., Alves-pinto, A., Sollini, J.,

713   Wells, T., and Sumner, C.J. (2016). Behavioural estimates of auditory filter

714   widths in ferrets using notched-noise maskers. J Acoust Soc Am *139*, EL19.

715   50.   Patterson, R., Robinson, K., Holdworth, J., McKeown, D., Zhang, C., and

716   Allerhand, M. (1992). Complex sounds and auditory images. In Auditory

717   Physiology and Perception, Y. Cazals, K. Horner, and L. Demany, eds.

718   (Elsevier Ltd), pp. 429–446.

719   51.   Karajalainen, M. (1996). A binaural auditory model for sound quality

720   measurements and spatial hearing studies. In IEEE International

721   Conference on Acoustics, Speech, and Signal Processing (Atlanta, GA, USA:

722   IEEE), pp. 985–988.

723   52.   Roman, N., Wang, D., and Brown, G.J. (2003). Speech segregation based on

724   sound localization. J Acoust Soc Am *114*, 2236–2252.

725   53.   Robles, L., Ruggero, M.A., and Rich, N.C. (1991). Two-tone distortion in the

726   basilar membrane of the cochlea. Nature *349*, 413–414.

727   54.   Pressnitzer, D., and Patterson, R.D. (2001). Distortion products and the

728   perceived pitch of harmonic complex tones. In Physiological and

729   Psychophysical Bases of Auditory Function, D. Breebaart, A. Houtsma, A.

730   Kohlrausch, V. Prijs, and R. Schoonhoven, eds. (Maastrict: Shaker BV), pp.

731   97–104.

732   55.   Norman-Haignere, S., and McDermott, J.H. (2016). Distortion products in

733   auditory fMRI research: Measurements and solutions. Neuroimage *129*,

734   401–413.

735   56.   Moore, B.C.J., and Ohgushi, K. (1993). Audibility of partials in inharmonic

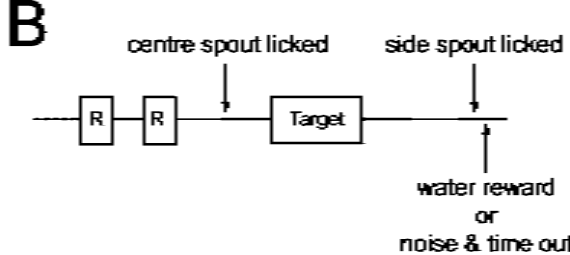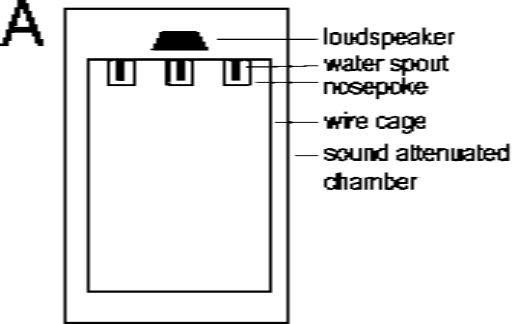736   complex tones. J Acoust Soc Am *93*, 452–461.

737   57.   Moore, B.C.J., Peters, R.W., and Glasberg, B.R. (1985). Thresholds for the
738   detection of inharmonicity in complex tones. J Acoust Soc Am *77*, 1861–
739   1867.

740   58.   Kaernbach, C., and Bering, C. (2001). Exploring the temporal mechanism
741   involved in the pitch of unresolved harmonics. J Acoust Soc Am *110*, 1039–
742   1048.

743   59.   Penagos, H., Melcher, J.R., and Oxenham, A.J. (2004). A neural
744   representation of pitch salience in nonprimary human auditory cortex
745   revealed with functional magnetic resonance imaging. J Neurosci *24*,
746   6810–6815.

747   60.   Norman-Haignere, S., Kanwisher, N., and McDermott, J.H. (2013). Cortical
748   pitch regions in humans respond primarily to resolved harmonics and are
749   located in specific tonotopic regions of anterior auditory cortex. J Neurosci
750   *33*, 19451–19469.

751   61.   Liberman, M.C. (1978). Auditory-nerve response from cats raised in a low-
752   noise chamber. J Acoust Soc Am *63*, 442–455.

753   62.   Temchin, A.N., Rich, N.C., and Ruggero, M.A. (2008). Threshold tuning
754   curves of chinchilla auditory-nerve fibers. I. Dependence on characteristic
755   frequency and relation to the magnitudes of cochlear vibrations. J
756   Neurophysiol *100*, 2889–2898.

757   63.   Shofner, W.P. (2002). Perception of the periodicity strength of complex
758   sounds by the chinchilla. Hear Res *173*, 69–81.

759   64.   Fritz, J., Shamma, S., Elhilali, M., and Klein, D. (2003). Rapid task-related
760   plasticity of spectrotemporal receptive fields in primary auditory cortex.
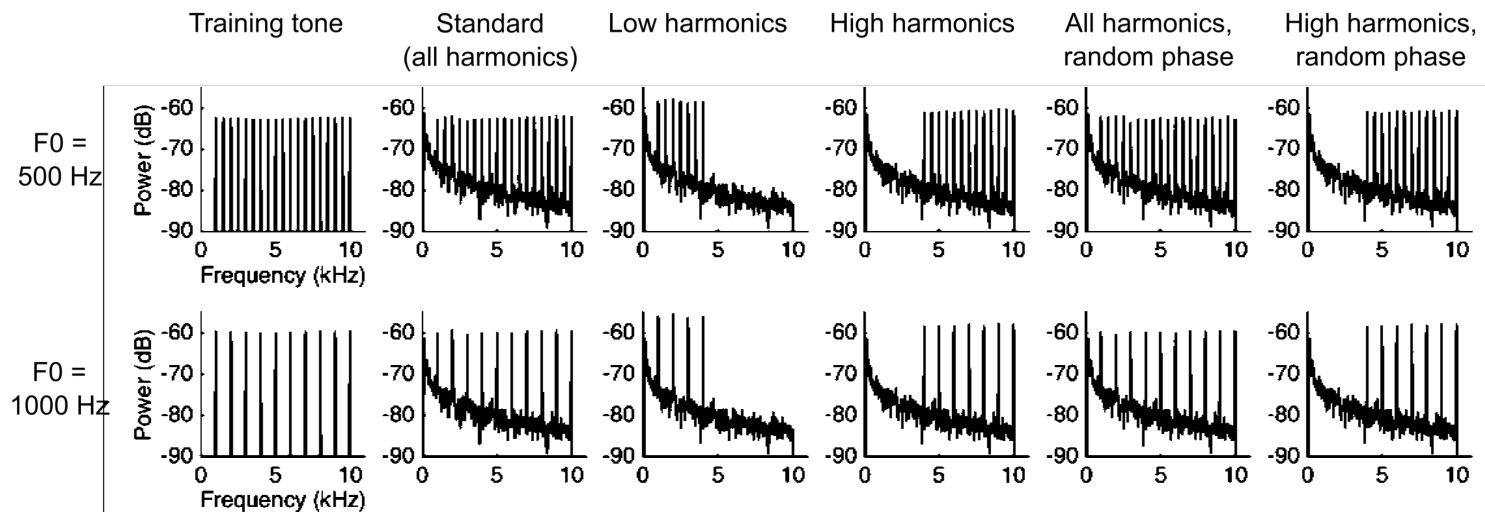761   Nat Neurosci *6*, 1216–1223.

762   65.   Walker, K.M.M., Bizley, J.K., King, A.J., and Schnupp, J.W.H. (2011).

763   Multiplexed and robust representations of sound features in auditory

764   cortex. J Neurosci *31*, 14565–14576.

765   66.   Atilgan, H., Town, S.M., Wood, K.C., Jones, G.P., Maddox, R.K., Lee, A.K.C., and

766   Bizley, J.K. (2018). Integration of visual information in auditory cortex

767   promotes auditory scene analysis through multisensory binding. Neuron

768   *97*, 640–655.

769   67.   Schwartz, Z.P., and David, S. V (2018). Focal suppression of distractor

770   sounds by selective attention in auditory cortex. Cereb Cortex *28*, 323–

771   339.

772   68.   Meddis, R., and O'Mard, L. (1997). A unitary model of pitch perception. J

773   Acoust Soc Am *102*, 1811–1820.

774   69.   Gfeller, K., Turner, C., Oleson, J., Zhang, X., Gantz, B., Froman, R., and

775   Olszewski, C. (2007). Accuracy of cochlear implant recipients on pitch

776   perception, melody recognition, and speech reception in noise. Ear Hear

777   *28*, 412–423.

778   70.   Slaney, M. (1993). An efficient implementation of the Patterson-

779   Holdsworth auditory filter bank.
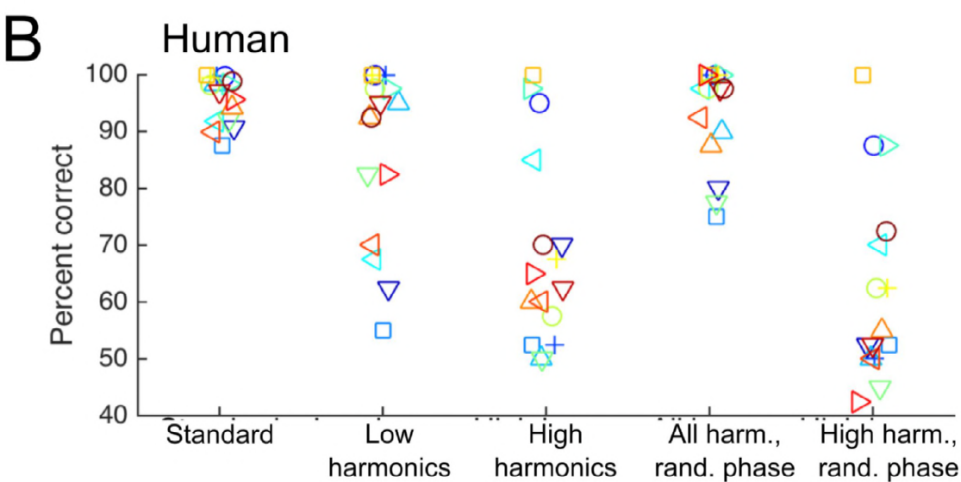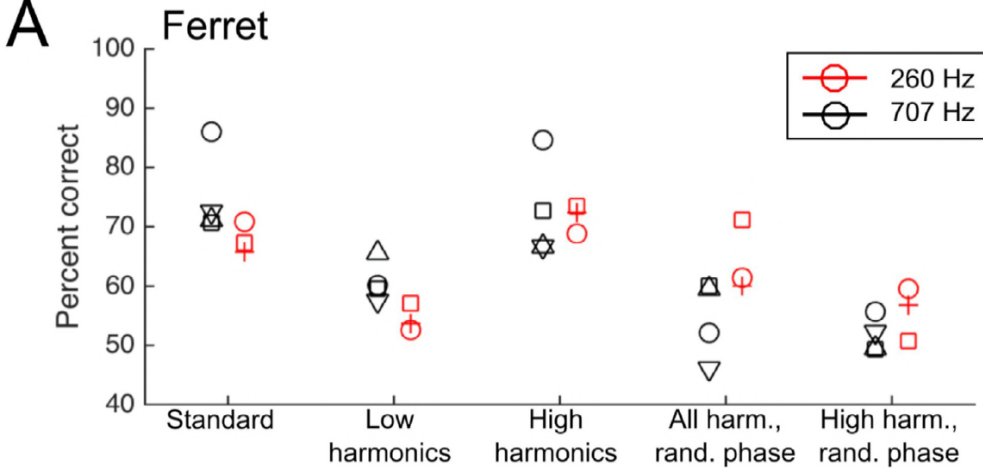
Supplemental Figure 1: Pitch classification performance of ferrets and humans, as shown in Figure 5A and B. Here, data are plotted for individual subjects. **A.** Ferrets' percent correct scores on the pitch classification task are plotted for the standard tone trials (left) and each of 4 probe stimuli (right). The results of testing with the 260 Hz reference (150 and 450 Hz targets; red) and 707 Hz reference (500 and 1000 Hz targets; black) are plotted separately. Symbol shapes represent individual ferrets. **B.** Humans' pitch classification performance, as plotted in (A) above. Data are randomly jittered on the x-axis to fascilitate visualization of individual points. Each symbol and colour combination indicates an individual subject.

A

Amplitude / Energy (dB) / Time (ms) vs Frequency (kHz)

B

Human   Ferret

Magnitude (dB) vs Frequency (kHz)

C

Amplitude vs Frequency (kHz)

— Human
— Ferret

D

Amplitude vs Time (ms)

E

Power at F0 vs Species

Human   Ferret

A

- loudspeaker
- water spout
- nosepoke
- wire cage
- sound attenuated chamber

B

centre spout licked   side spout licked

R — R — | — Target — |

water reward
or
noise & time out

	Training tone	Standard (all harmonics)	Low harmonics	High harmonics	All harmonics, random phase	High harmonics, random phase
Resolved harmonic cues	Strong	Strong	Strong	Weak	Strong	Weak
Temporal envelope cues	Strong	Strong	Weak	Strong	Weak	Weak
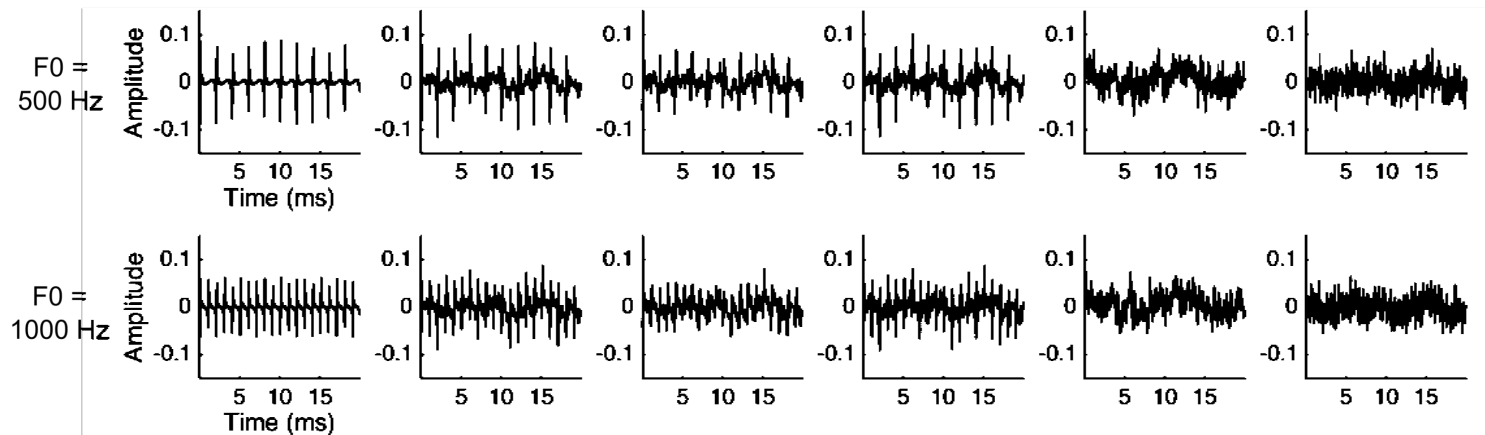
A
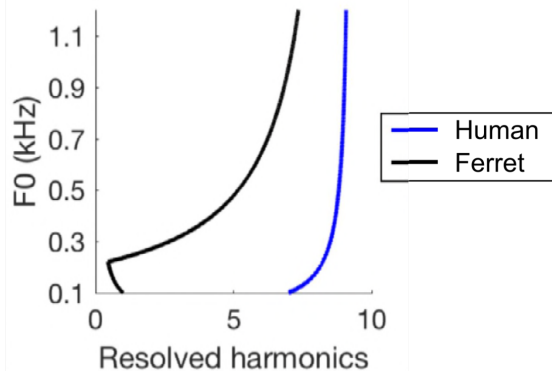
B

Ferret (reference = 260 Hz)	All harmonics	Low harmonics	High harmonics
Spectral range (Hz)	450 - 6000	450 - 1500	1500 - 6000
Low F0 harmonics	3 - 40	3 - 10	10 - 40
High F0 harmonics	1 - 13	1 - 3	5 - 13

Ferret (reference = 707 Hz)	All harmonics	Low harmonics	High harmonics
Spectral range (Hz)	1000 - 10000	1000 - 4000	4000 - 10000
Low F0 harmonics	2 - 20	2 - 8	8 - 20
High F0 harmonics	1 - 10	1 - 4	4 - 10

Human (reference = 199 Hz)	All harmonics	Low harmonics	High harmonics
Spectral range (Hz)	220 - 5940	220 - 1980	1980 - 5940
Low F0 harmonics	2 - 33	2 - 11	11 - 33
High F0 harmonics	1 - 27	1 - 9	9 - 27