

# 1 Beat-Relevant Signals in Auditory Cortical 2 Responses to Musical Excerpts

3

4

5 **Vani G. Rajendran<sup>1,2</sup>, Nicol S. Harper<sup>1</sup>, Jan W. H. Schnupp<sup>2</sup>**

6

7

8

9

10

11 <sup>1</sup> Auditory Neuroscience Group, University of Oxford, Dept. of Physiology, Anatomy, and  
12 Genetics, Oxford, UK.

13

14 <sup>2</sup> City University of Hong Kong, Dept. of Biomedical Science. 31 To Yuen Street, 2/F 1B-202,  
15 Kowloon Tong, Hong Kong.

16

17

18

19

20

21

22

23

24 **Corresponding author:**

25 Jan W. H. Schnupp,

26 City University of Hong Kong, Dept. of Biomedical Science.

27 31 To Yuen Street, 2/F 1B-202, Kowloon Tong, Hong Kong

28 [wschnupp@cityu.edu.hk](mailto:wschnupp@cityu.edu.hk)

## 29 **Abstract**

30 Musical beat perception is widely regarded as a high-level ability involving widespread  
31 coordination across brain areas, but how low-level auditory processing must necessarily shape  
32 these dynamics, and therefore perception, remains unexplored. Previous cross-species work  
33 suggested that beat perception in simple rhythmic noise bursts is shaped by neural transients in  
34 the ascending sensory pathway. Here, we found that low-level processes even substantially  
35 explain the emergence of beat in real music. Firing rates in the rat auditory cortex in response to  
36 twenty musical excerpts were on average higher on the beat than off the beat tapped by human  
37 listeners. This “neural emphasis” distinguished the perceived beat from alternative  
38 interpretations, was predictive of the degree of consensus across listeners, and was accounted for  
39 by a spectrotemporal receptive field model. These findings indicate that low-level auditory  
40 processing may have a stronger influence on the location and clarity of the beat in music than  
41 previously thought.

42

43 **Keywords:** beat, rhythm, music, electrophysiology, psychophysics, temporal processing,  
44 perception, sensorimotor synchronization

45

## 46 **Introduction**

47 The perception of a steady pulse or beat in music is a curious phenomenon that arises from the  
48 interaction between rhythmic sounds and the way our brain processes them. There are two things  
49 that make musical beat perception particularly intriguing. Firstly, no mammalian species apart  
50 from humans consistently show spontaneous motor entrainment to the beat in music (e.g. tapping  
51 a foot, nodding the head, moving the body)<sup>1-4</sup>. Secondly, despite beat being a subjective percept  
52 rather than an acoustic feature of music, individual listeners tend to overwhelmingly agree on  
53 where the beat is. Some of this consistency might be due to certain “top-down” constraints such  
54 as cultural and cognitive priors<sup>5-7</sup>. However, apart from theory<sup>8,9</sup>, relatively little is known about  
55 the neurophysiological dynamics that cause the feeling of musical beat to emerge in the first  
56 place.

57  
58 A key piece of information currently lacking is which aspects of the neural representation of  
59 music might be important for the induction of beat. Previous cross-species work revealed that  
60 firing rates as early as the auditory midbrain are significantly higher on the beat than off the beat  
61 in simple rhythms constructed from identical broadband noise bursts<sup>10</sup>. If large firing rate  
62 transients resulting from low-level auditory processing are indeed necessary for the induction of  
63 beat, then this insight could shed light on the dynamics of the entrainment of cortical oscillations  
64 to beat<sup>11-16</sup>, the role played by the motor system<sup>8,17-26</sup>, and why different species differ so much in  
65 their beat perception and synchronization capacity<sup>27</sup>.

66  
67 Importantly, if a consequence of auditory processing is to create points of neural emphasis that  
68 predispose beats being felt there, then we should observe this not just for simple rhythmic  
69 “laboratory sounds,” but also for real music. Twenty musical excerpts<sup>28</sup>, which were diverse in  
70 tempo and musical genre, were played to three anesthetized rats while recording extracellularly  
71 from auditory cortex. In line with previous findings, population firing rates were higher on the  
72 beat than off the beat, and large on-beat to off-beat firing rate ratios were a distinguishing feature  
73 of the consensus beat interpretation across human listeners. Comparison with the output of an  
74 auditory nerve model revealed that small effects may already be present at the auditory periphery  
75 but are amplified substantially in cortical responses. Musical excerpts that evoked a larger

76 cortical on-beat emphasis also showed a stronger consensus in tapping behavior across listeners.  
77 Finally, these results could be accounted for by the spectrotemporal receptive field properties of  
78 recorded units. These findings add to growing evidence that beat perception is not entirely  
79 culturally determined, but is also heavily constrained by low-level auditory processing common  
80 to mammals.

81

## 82 **Results**

83 Neural activity from a total of 98 single and multi-units were analyzed in response to 12 repeats  
84 of the first 10 seconds of 20 musical excerpts taken from the MIREX 2006 dataset online, which  
85 included beat annotations made by 40 human listeners<sup>28</sup>. In all songs, listeners reported a steady  
86 beat well within the first 10 s. The most common tapping pattern for each excerpt was taken to  
87 be that excerpt's "consensus" beat interpretation (see *Methods*), and consensus tapping rates  
88 ranged from 0.7 Hz to 3.7 Hz (42 to 222 beats per minute, corresponding to beat periods of 1.42  
89 down to 0.27 s). The analyses that follow investigate correspondences between firing rates in the  
90 rat auditory cortex around the consensus beat as reported by human listeners.

91

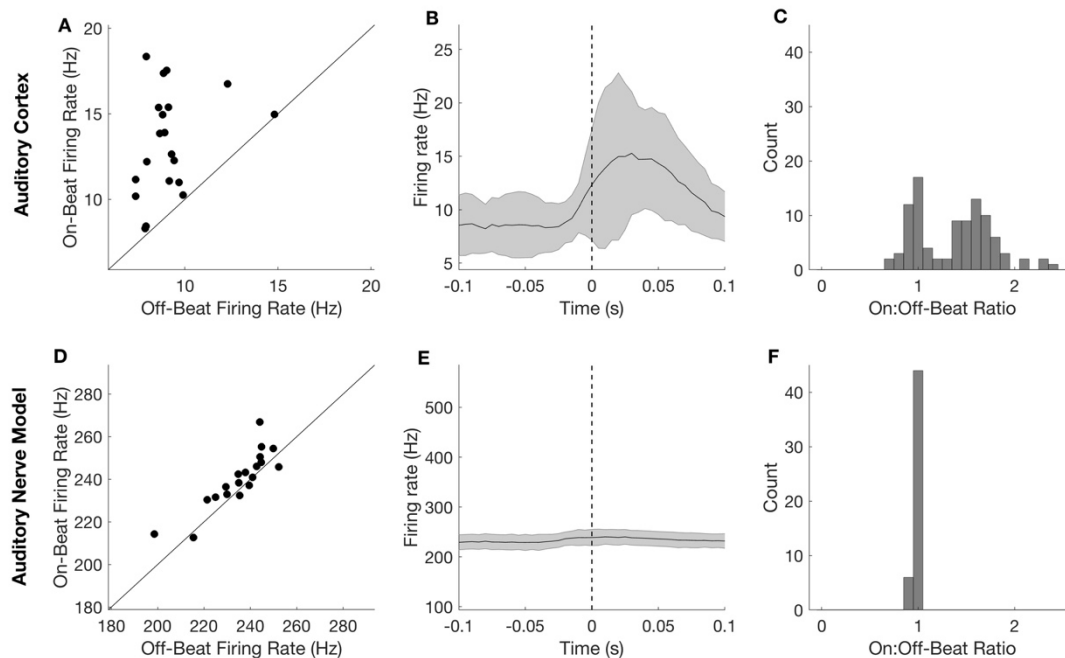
### 92 **Auditory cortical firing rates are higher on the beat than off the beat**

93 For each song, the 100 ms time window following each consensus tap was defined as on-beat,  
94 and all time excluding these on-beat windows was defined as off-beat (the results are not  
95 sensitively dependent on this precise definition, see *Methods*). Fig 1A shows the average on-beat  
96 population firing rate plotted against the off-beat population firing rate for each of the 20 tested  
97 musical excerpts. On-beat firing rates were significantly larger than off-beat firing rates ( $p < 10^{-4}$ ,  
98 Wilcoxon paired signed-rank test,  $N=20$  songs), an observation that is consistent with previous  
99 work examining gerbil midbrain responses to simple rhythmic patterns<sup>10</sup>. The beat-triggered  
100 average population firing rate in the 200 ms window around consensus beats (averaged across all  
101 beats in all excerpts) provides a more detailed picture of population neural activity around the  
102 beat (Fig 1B). The distribution of on:off-beat ratios (OORs; average on-beat firing rate divided  
103 by average off-beat firing rate) for each recorded unit ( $N=98$ ) is shown in Fig 1C. An  $OOR > 1$   
104 indicates that firing rates were higher on the beat than off the beat. Most units show an  $OOR > 1$ ,

105 and the bimodal distribution suggests that there may exist distinct sub-populations in the  
106 recorded data, one with OORs centered around 1 and the other with OORs around 1.5.

107

108 For comparison, an auditory nerve model<sup>29</sup> was used to predict firing rates at the auditory nerve  
109 for 50 logarithmically spaced frequency channels between 150 Hz and 24 kHz. Fig 1D shows  
110 predictions of on-beat versus off-beat population activity at the auditory nerve. Notably, the  
111 auditory nerve model would also predict higher average population firing rates on the beat than  
112 off the beat ( $p < 0.005$ , Wilcoxon paired signed-rank test,  $N = 20$  songs). Fig 1E-1F show beat-  
113 triggered averages and OORs for auditory nerve model fibers. OORs based on the auditory nerve  
114 model, though significantly larger than one, are much smaller than cortical OORs ( $p < 10^{-4}$ ,  
115 Wilcoxon paired signed-rank test,  $N = 20$  songs).



116

117 **Fig 1. Consensus beat-triggered neural activity and on:off-beat ratios (OORs) in the auditory cortex and**  
118 **auditory nerve. (A)** Mean on-beat versus off-beat population firing rate in auditory cortical neurons. Each dot is  
119 one musical excerpt. On-beat firing rates are significantly higher than off-beat firing rates ( $p < 10^{-4}$ , Wilcoxon paired  
120 signed-rank test,  $N = 20$  songs) **(B)** Population “beat-triggered” average firing rate in the auditory cortex in a 200 ms  
121 window around the consensus beat times  $\pm$  standard deviation across the 20 musical excerpts. **(C)** Histogram of  
122 on:off-beat firing rate ratios (OORs) for each recorded unit ( $N = 98$ ), where “on-beat” is the average firing rate  
123 during the 100 ms post-tap window, and “off-beat” is the average firing rate over the entire song excluding on-beat  
124 windows. **(D)** Same as A, but for population activity based on an auditory nerve model with 50 log-spaced  
125 frequency channels between 150 Hz and 24 kHz. Predicted firing rates at the auditory nerve were significantly  
126 higher on the beat than off the beat ( $p < 0.005$ , Wilcoxon paired signed-rank test,  $N = 20$  songs) **(E)** Same as B, but for  
127 population activity based on the auditory nerve model. **(F)** Same as C, but for auditory nerve model fibers ( $N = 50$ ).

128 **A large neural emphasis is a distinguishing feature of the consensus beat**

129 While we have shown that firing rates are higher on the beat than off the beat, this on its own  
130 does not imply that large OORs are necessarily relevant to beat perception. From a purely signal  
131 processing perspective, a musical excerpt could theoretically be perceived as having any  
132 combination of tempo and time signature, and if most of these possible alternative beat  
133 interpretations were associated with more or less equally large OORs, then large OORs would be  
134 of little value as physiological markers of musical beat. Therefore, if a large OOR is relevant for  
135 the induction of beat, we hypothesized that it should be large for the consensus beat relative to  
136 plausible alternatives.

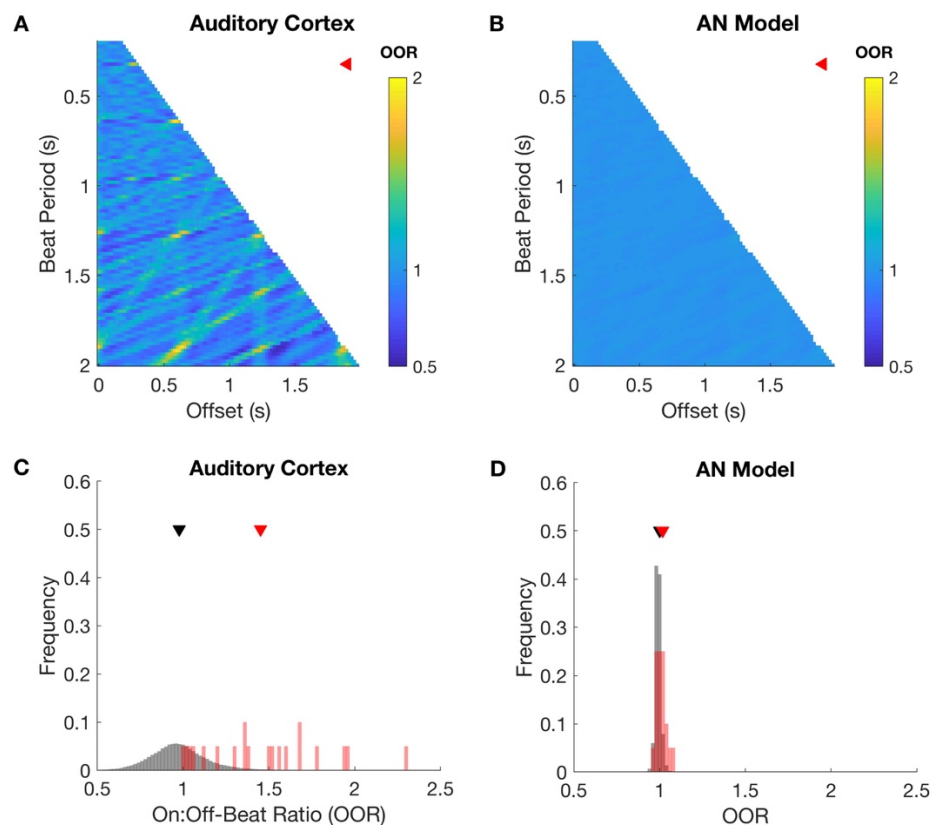
137

138 To test this, we computed hypothetical OORs for the full range of plausible beat period and  
139 phase combinations. For each song, possible beat periods (representing the different rates at  
140 which a listener might tap) were allowed to range from 0.2 s to 2 s (5 Hz down to 0.5 Hz)  
141 sampled in 20 ms steps. Likewise, for each beat period, the phase offset was allowed to range  
142 from 0 up to the full beat period sampled in 20 ms steps to capture the fact that two listeners  
143 tapping at the same rate may nonetheless exhibit different interpretations of the beat if their taps,  
144 rather than being synchronous, have a constant offset between them. The OOR was then  
145 computed for each of these beat interval and beat offset combinations, resulting in 4,995 possible  
146 OOR values for each musical excerpt. The heatmaps in Fig 2A and 2B show the computed set of  
147 plausible OOR values calculated from cortical and auditory nerve model firing rates,  
148 respectively, for an example musical excerpt, with possible beat periods on the y-axis and  
149 possible starting phase offsets on the x-axis (see Supplementary Figs S1-S2 for heatmaps of all  
150 musical excerpts). The histograms in Fig 2C and 2D pool together hypothetical (in gray) and  
151 consensus (in red) OOR values from all musical excerpts (histograms for individual excerpts in  
152 Supplementary Fig S3-S4).

153

154 If the OOR is a distinguishing feature of the perceived beat, we would expect it to rank above the  
155 50<sup>th</sup> percentile of the underlying distribution of hypothetically plausible OORs for a given  
156 musical excerpt. As hypothesized, the consensus OORs rank significantly larger than the 50<sup>th</sup>  
157 percentile, both in the auditory cortex ( $p < 10^{-4}$ , Wilcoxon signed-rank test,  $N = 20$  songs), and in  
158 the auditory nerve model ( $p < 0.005$ ). However, the percentiles were significantly larger in the

159 auditory cortex than in the auditory nerve model ( $p < 0.005$ , Wilcoxon paired signed-rank test,  $N$   
160 = 20 songs). Notably, 14 out of the 20 musical excerpts tested had consensus OORs above the  
161 95<sup>th</sup> percentile in the auditory cortex, in contrast to only 7 out of 20 based on the auditory nerve  
162 model. Additionally, fewer hypothetical beat interpretations resulted in large OORs in the  
163 auditory cortex, as evidenced by the higher skewness, or longer right tails, of the OOR  
164 distributions in the auditory cortex compared to those based on the auditory nerve model ( $p < 10^{-4}$ ,  
165 Wilcoxon paired signed-rank test,  $N = 20$  songs). Together, these results suggest that a large  
166 OOR is a feature that distinguishes the consensus beat from most other possible beat structures,  
167 and that two important consequences of auditory processing might be an amplification of small  
168 differences in OOR already present at the auditory periphery, and a further restriction of the  
169 candidate beat interpretations that would result in large OORs.



170

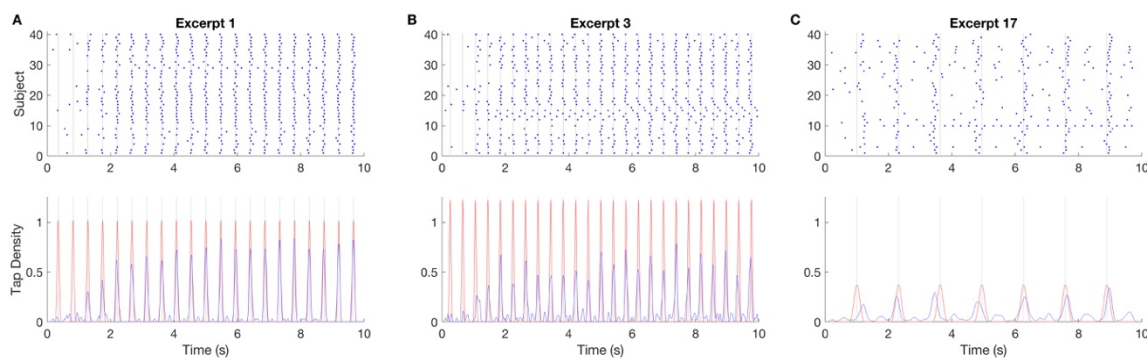
171 **Fig 2. How does the consensus beat compare with other possible beat structures?** (A) Heatmap depicting  
172 cortical on:off-beat ratios for plausible beat period (y-axis) and beat phase offset (x-axis) combinations between 200  
173 ms and 2 s (or tap rates of 5 Hz down to 0.5 Hz) for one example musical excerpt. Color indicates the OOR value.  
174 (B) Same as A, but for population activity based on the auditory nerve model. (C) Histogram pooled across musical  
175 excerpts of all OOR values (gray), and consensus OOR values (red) in the auditory cortex. (D) Same as C, but based  
176 on OOR values from the auditory nerve model.

177

## 178 **The stronger the on-beat neural emphasis, the stronger the tapping consensus**

179 It is clear from Fig 2C (and Supplementary Fig S3) that consensus OORs are consistently among  
180 the largest possible OORs across our set of musical excerpts, but they are not always the largest.  
181 However, it is not uncommon for the beat in a given piece of music to be perceived in different  
182 ways. More often than not, listeners will exhibit a variety of tapping patterns, for example with  
183 some tapping twice as fast or half as fast as others, or 180 degrees out of phase with others.  
184 Additionally, if the beat is not very salient, there will be uncertainty about when exactly a beat  
185 occurs and therefore an increased variance in observed inter-tap-intervals. In such cases, and  
186 indeed in the dataset we use, listeners display a range of perceived beat interpretations, and what  
187 we have termed the consensus beat is merely the beat interpretation that happens to be favored  
188 by a (sometimes narrow) majority of listeners. This variability is illustrated in Fig 3, where for  
189 some excerpts tapping behavior was consistent across a large majority of listeners (e.g. Fig 3A),  
190 and for others tapping behavior was more variable, indicating a less salient or more ambiguous  
191 beat percept (e.g. Fig 3B-3C; see Supplementary Fig S5 and S6 for tapping behavior for all  
192 excerpts).

193



194

195 **Fig 3. A glimpse into the variability across human listeners tapping to the beat in music. (A) Top:** Raster plot  
196 of tap times for the 40 human annotators across the 10 s excerpt of an example song. Each row is one subject, and  
197 location along the x-axis represents when the subject tapped during the 10 s musical excerpt. Consensus beat times  
198 are marked by gray vertical lines (see *Methods*). Note that most subjects' taps line up in time with each other and  
199 with consensus beat for this example excerpt. **Bottom:** Tap density estimates based on tap times pooled across  
200 subjects, binned with 2 ms bins, and smoothed with a Gaussian kernel with a standard deviation of 5% of the  
201 consensus beat period (blue). Shown in red is a smoothed tap density estimate of the “ideal” tap histogram (with  
202 realistic motor error) that would have been obtained if all subjects had tapped on every consensus beat (see  
203 *Methods*). The correlation between real and idealized density is high for this excerpt ( $r=0.88$ ), indicating a strong  
204 tapping consensus. **(B)** Same as A, but for a musical excerpt with multiple minority beat interpretations and  
205 therefore a lower correlation coefficient ( $r=0.78$ ). **(C)** Same as B, but where the tapping consensus is even weaker  
206 ( $r=0.59$ ). See Supplementary Figs S5 and S6 for all musical excerpts.

207



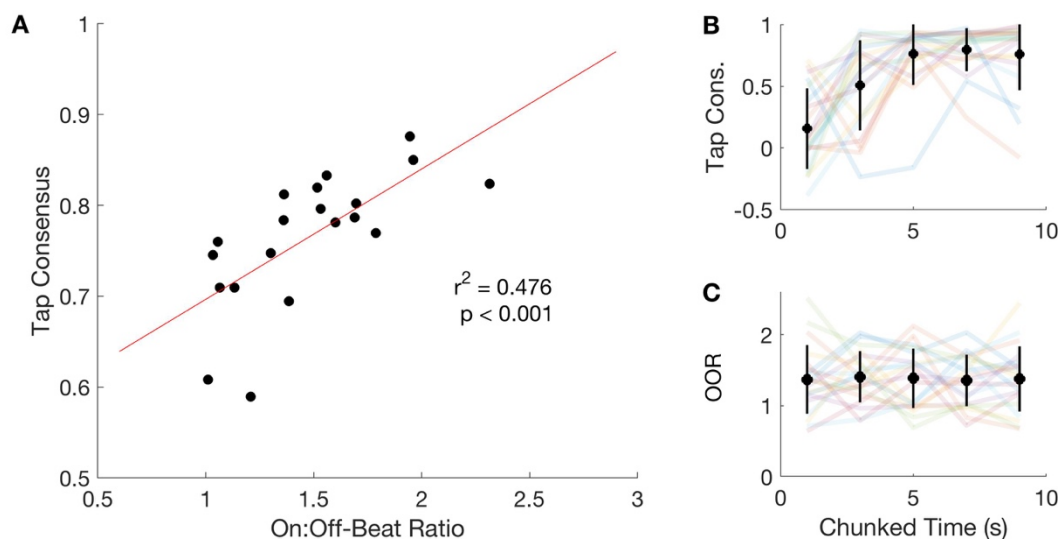
208 However, if we hypothesize that a large OOR predisposes a listener to hear a particular beat  
209 interpretation, then we would predict that the excerpts that evoke the largest OORs in cortical  
210 responses should also be the ones that evoke the clearest, most unambiguous beat percept across  
211 listeners. Can the variability in tapping behavior be explained by the size of OORs in the  
212 auditory cortex?

213

214 To answer this question, we quantified the strength of the tapping consensus for each song by  
215 calculating the correlation coefficient between the smoothed histogram of observed tap times and  
216 the smoothed histogram of the “ideal” case in which all 40 listeners would have tapped on each  
217 consensus beat within a realistic degree of sensory or motor error (see *Methods*). Examples of  
218 observed (blue) and idealized (red) tap density estimates are shown the lower panels of Fig 3.

219

220 Consistent with our hypothesis, the size of the consensus OOR evoked in the auditory cortex by  
221 a musical excerpt correlated significantly with the strength of the tapping consensus across  
222 listeners (Fig 4A;  $p < 0.001$ , Pearson correlation,  $N = 20$  songs). Neither OOR ( $p = 0.48$ ) nor  
223 consensus strength ( $p = 0.44$ ) varied with the consensus tempo of musical excerpts (Pearson  
224 correlation,  $N = 20$  songs). Fig 4B and 4C show how OOR and consensus strength, respectively,  
225 develop over the course of the 10 s duration of the musical excerpts. Data were split into five 2-s  
226 chunks, and OORs and correlation coefficients were calculated based on the data in each chunk.  
227 Tapping consensus strength, which is low initially, is nearly at ceiling from about 4 s into the  
228 excerpts, indicating that listeners only needed a few seconds to find the beat. OORs, on the other  
229 hand, did not change systematically over time, suggesting that the correspondences observed in  
230 this study between neural activity and behavior are unlikely to be due to cortical entrainment or  
231 buildup in neural responses.



232

233 **Fig 4. The stronger the on-beat neural emphasis, the stronger the tapping consensus.** (A) Each dot is one  
234 musical excerpt. There is a strong correlation between auditory cortical OOR (x-axis) and the tapping consensus  
235 across listeners, quantified as described in Fig 3 (y-axis;  $p < 0.001$ , Pearson correlation,  $N = 20$  songs). (B) Tapping  
236 consensus, calculated for each sequential 2 s segment of musical excerpts. Colored lines are individual songs. In  
237 black is the mean across songs for each time chunk  $\pm$  standard deviation. (C) Same as panel B but for OOR values.

238

### 239 Spectrotemporal receptive field based models explain nearly 90% of the variance in OOR

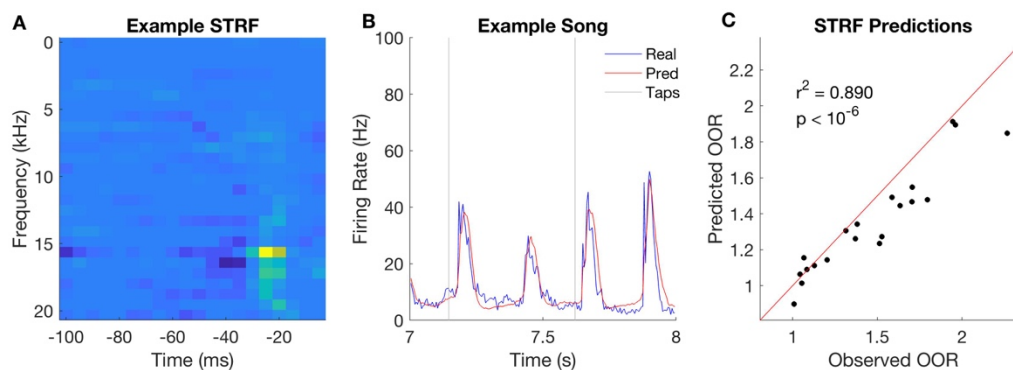
240 The beat-related processing observed in the rat auditory cortex may be due to beat-specific  
241 processes, or, as we hypothesized might be more likely, due to the spectrotemporal tuning  
242 properties of recorded units. If this were the case, neural responses predicted using a standard  
243 linear-nonlinear (LN) model fitted to each unit should largely reproduce observed OORs. To test  
244 this, we first estimated each unit's spectrotemporal receptive field (STRF), or the linear model  
245 that describes the frequency and timing properties of incoming sounds that would either excite or  
246 inhibit a neuron. Next, we estimated the unit's static sigmoid output nonlinearity to arrive at a  
247 fitted LN model for each unit (see *Methods*). The LN model was fitted 20 times for each unit,  
248 each time using that unit's responses to 19 of the musical excerpts while setting aside one  
249 excerpt as a test song. This ensured that predicted neural responses for a test song were true  
250 predictions since the model was not trained on the test excerpt. In this manner, firing rate  
251 predictions were generated for each unit and each musical excerpt, and these were then analyzed  
252 to arrive at predicted OOR values.

253

254 An STRF from an example unit is shown in Fig 5A, with frequency on the y-axis and stimulus  
255 history on the x-axis. This unit shows a preference for frequencies at and above 16 kHz, and is

256 excited if sounds in that frequency range were heard 25 ms ago but inhibited if they occurred 40  
257 ms ago. A short excerpt from a test song is shown in Fig 5B, where it can be seen that LN model  
258 predictions are in good agreement with observed firing rates. Fig 5C shows consensus OOR  
259 values for each musical excerpt based either on observed (x-axis) or predicted (y-axis) firing  
260 rates. The LN model slightly underestimates OORs ( $p < 0.001$ , Wilcoxon paired signed-rank test,  
261  $N = 20$  songs), suggesting that there is some nonlinear process that slightly increases OOR  
262 beyond processes captured by a standard LN model. However, despite this minor difference, the  
263 LN model successfully accounts for 89% of the variance in OOR values for the tested musical  
264 excerpts ( $p < 10^{-6}$ , Pearson correlation,  $N = 20$  songs). Predictions made using the linear STRF  
265 alone (without the static nonlinearity) accounted for 61% of the variance in OOR ( $p < 0.01$ ,  
266 Pearson correlation,  $N = 20$  songs).

267



268  
269  
270 **Fig 5. Cortical firing rate predictions based on fitted linear-nonlinear (LN) models incorporating**  
271 **spectrotemporal receptive fields (STRFs).** (A) STRF from an example unit, with frequency on the y-axis and time  
272 on the x-axis and color representing the coefficients. This unit shows a classic pattern of excitation and inhibition in  
273 a relatively narrow frequency range. Convolution of this filter with the spectrogram of a sound stimulus, and then  
274 applying a static nonlinearity, would result in the LN model's prediction of this unit's firing rate over time. (B)  
275 Measured (blue) and LN model predictions (red) of the population firing rate for a 1 s segment of an example  
276 musical excerpt. Gray vertical lines mark consensus tap times in this segment. (C) Observed (x-axis) versus  
277 predicted (y-axis) consensus on:off-beat ratios for each song. LN models account for 89% of the variance in OOR.  
278

279

## 280 **Discussion**

281 The aim of this study was to explore how firing rate transients in the auditory cortical  
282 representation of music might set the stage for the perception of musical beat. Our results, based  
283 on the twenty musical excerpts that were diverse in tempo and genre, revealed that population  
284 firing rates were on average higher on the beat than off the beat, and that large on:off-beat ratios  
285 (OORs) were a distinguishing feature of the beat interpretations most commonly tapped by  
286 human listeners. While small differences between on-beat and off-beat responses were already  
287 present in auditory nerve model responses, these differences were substantially amplified in  
288 auditory cortical responses. Furthermore, musical excerpts that evoked larger OORs in the  
289 auditory cortex also showed stronger tapping consensus among listeners. Finally, the  
290 spectrotemporal receptive field (STRF) properties of cortical units were able to account for the  
291 magnitude of the OOR each musical excerpt would induce. Together, these findings suggest that  
292 large OORs in the auditory cortex, which arise due to the spectrotemporal tuning properties of  
293 neurons, may be key to establishing the location and clarity of the perceived beat.

294  
295 It is worth noting is the extent to which the physiology corresponded to tapping behavior and the  
296 extent to which standard LN STRF models could capture the physiology for real musical  
297 excerpts. These observations strongly suggest that the related low-level mechanisms of neuronal  
298 adaptation<sup>10</sup>, amplitude modulation tuning<sup>30</sup>, and STRFs play a formative role musical in beat  
299 perception. This is not inconsistent with the theory that the induction of the beat percept is the  
300 result of an interaction between “bottom-up” sensory processes and “top-down” cognitive ones<sup>31</sup>.  
301 Our data suggest that beat perception may really begin weakly at the ear, with neural activity  
302 showing stronger correspondences to behavior as information ascends through the brainstem and  
303 primary cortical structures of the ascending auditory pathway<sup>32,33</sup>. Since these parts of the  
304 ascending auditory system are often highly conserved across mammalian species<sup>34-37</sup>, cross-  
305 species investigations may be a promising way to understand the neural signals and dynamics  
306 that underlie beat induction, which to date remain mysterious.

307  
308 Though our results indicate that beat perception is strongly influenced by basic physiological  
309 mechanisms and therefore only partly culturally determined, they do not imply that “bottom-up”

310 processes could possibly explain everything. For example, some well-studied constraints on-beat  
311 perception include the tendency to perceive a beat within a frequency range of roughly 0.5–4  
312 Hz<sup>38</sup> with a special preference for 2 Hz<sup>39</sup>, and an overall preference for binary (e.g. 2, 4) meters  
313 over ternary (e.g. 3, 6) or other complex meters<sup>38,40</sup>. These constraints are likely driven by top-  
314 down influences or may result from auditory-motor interactions<sup>8,17-25</sup> and are unlikely to be  
315 explained by bottom-up sensory processing alone. Furthermore, the perceived beat and its neural  
316 signatures can be modulated at will by top-down attention or mental imagery of beat  
317 structure<sup>12,41-43</sup>. Bringing these ideas together, we propose that the perception of beat relies on the  
318 application of learned and implicit rhythmic priors<sup>6,7</sup> onto an ascending sensory  
319 representation<sup>10,30</sup> with a bias towards configurations that maximize the difference between  
320 neural activity on and off the beat.

321  
322 That we see as much correspondence as we do between the representation in auditory cortex and  
323 beat perception could be an indication that neural activity in the auditory cortex is a key interface  
324 between the sensory and motor and/or cognitive processes involved in beat perception. Probing  
325 the cortico-basal ganglia-thalamo-cortical loop<sup>44</sup> may be a promising avenue for future  
326 investigations. Projections from auditory cortical fields to the basal ganglia have been well-  
327 characterized<sup>45</sup>, and the basal ganglia in humans have been repeatedly implicated in beat  
328 perception<sup>22,43,46,47</sup> as well as other auditory cognitive abilities<sup>48</sup>. We speculate that large firing  
329 rate transients in the auditory cortex, observed in this study to co-occur with the perceived beat,  
330 could set into motion the dynamics of this loop and thereby enable the possible entrainment of  
331 cortical oscillations to the beat<sup>21,42,49-51</sup>. We suggest caution, however, as there is currently some  
332 debate around what constitutes neural entrainment to auditory rhythms<sup>52-54</sup>, and whether  
333 frequency-domain representations of rhythms and brain signals necessarily reflect beat  
334 perception<sup>55</sup>.

335  
336 The extent of the correspondence observed in this study between auditory cortical activity in rats  
337 and human beat perception also invites the intriguing question of whether rodents too can  
338 perceive musical beat. Preliminary evidence suggests that rats can be trained to discriminate  
339 isochronous rhythms from non-isochronous ones<sup>56</sup>. Mice too appear capable of performing a  
340 synchronization-continuation task, and in that study, primary auditory cortex was implicated as

341 being necessary for the generation of anticipatory motor actions<sup>57</sup>. These studies at minimum  
342 suggest that rodents have the capacity to perceive temporal structure and execute motor actions  
343 timed to an external isochronous rhythm. Future behavioral studies are needed to explore the  
344 limits of sensorimotor synchronization in rodents.

345

346 At the other end of the spectrum are humans, whose ability to synchronize with an external  
347 rhythm, whether it is to a metronome or to the beat in music, is spontaneous<sup>1</sup>, highly  
348 anticipatory<sup>58</sup>, innate<sup>59</sup>, and often involuntary<sup>2,60,61</sup>. The gradual audiomotor evolution hypothesis  
349 posits that the ability to entrain movements to musical beat relies on strong coupling between the  
350 auditory and motor systems, and that the neurophysiology and behavioral capacity to do so  
351 evolved gradually<sup>20</sup>. This hypothesis is supported by evidence that nonhuman primates, like  
352 humans, are capable of producing tempo-flexible anticipatory movements in time with a  
353 metronome<sup>62,63</sup> and can detect rhythmic groupings, but cannot detect or synchronize to a musical  
354 beat<sup>4</sup>. The dissociation between perceiving auditory rhythms and perceiving musical beat may  
355 relate to findings that distinct networks underpin “duration-based” and “beat-based” temporal  
356 predictions<sup>64-66</sup>. It is important for future studies in the area of beat perception to be clear about  
357 precisely what is being perceived, since there is demonstrable nonequivalence between the  
358 detection of a pulse in isochronous rhythms, a pulse in real music, and beat in the context of the  
359 different levels of nested hierarchical structure present in music, the latter of which has arguably  
360 not yet been demonstrated in any nonhuman species<sup>67</sup>.

361

362 This leads to the question of why beat perception exists in the first place. Some clues might be  
363 found in parallels that beat perception has with other abilities, particularly with the human  
364 capacity for language<sup>68-70</sup>. Another possibility is that beat may provide a way to quickly assess  
365 locomotion speed from the sound of a complex gait. Though this speculation has not yet been  
366 tested directly, gait studies have shown that humans are able to assess a number of attributes of a  
367 walker based only on their walking sounds, including gender, posture, and emotional state<sup>71,72</sup>.

368

369 However, at the heart of these complex abilities are neural circuits that are very old and also  
370 underlie more general auditory cognitive abilities<sup>73</sup> such as perception of time<sup>74</sup> and prediction of  
371 future sensory inputs<sup>75</sup>. Therefore, a unified perspective that would bring all of this together is

372 that the information processing performed by the auditory system up to primary auditory cortex  
373 is largely consistent across most mammals, but the complexity of the operations the organism  
374 ecologically needs to perform with this information may be the determinant for what is “top-  
375 down.” Our data suggest that strong firing rate transients in the neural representation of real  
376 music may shape where the beat is felt, and while an on-beat neural emphasis is certainly not the  
377 whole story, it is a lead worth exploring further. Ultimately, this work underscores the  
378 importance of low-level auditory processing in creating a representation of sound where certain  
379 features are emphasized based on temporal context, a representation on which other high-level  
380 processes rely to give rise to complex perception.  
381

## 382 **Methods**

383

### 384 *Stimuli*

385

386 The 20 songs tested were the training dataset for the MIREX 2006 beat tracking algorithm  
387 competition<sup>76</sup>. Each song had beat annotations collected from 40 human listeners<sup>28</sup>. Only the first  
388 10 s of songs and beat annotations were used in this study.

389

### 390 *Surgical Protocol*

391

392 All procedures were approved and licensed by the UK home office in accordance with governing  
393 legislation (ASPA 1986). Three female Lister Hooded rats weighing approximately 250 grams  
394 were anesthetized with an intraperitoneal injection of 0.05 ml domitor and 0.1 ml ketamine. To  
395 maintain anesthesia, a saline solution containing 16 ug/kg/h domitor, 4 mg/kg/h ketamine, and  
396 0.5 mg/kg/h torbugesic were infused continuously during recording at a rate of 1 ml/h. A  
397 craniotomy was performed 4.7 mm caudal to bregma and extending 3.5 mm lateral from the  
398 midline on the right hand side.

399

400 Recordings were made using a 64 channel silicon probe (Neuronexus Technologies, Ann Arbor,  
401 MI, USA) with 175  $\mu\text{m}^2$  recording sites arranged in a square grid pattern at 0.2 mm intervals  
402 along eight shanks with eight channels per shank. The probe was inserted into the auditory cortex  
403 in a medio-lateral orientation wherever possible.

404

405 The 20 songs were played in randomized order for a total of 12 repeats, with 3 seconds of silence  
406 separating each song from the next. Stimuli were presented binaurally through headphones at 80  
407 dB SPL. Sounds were presented with a sampling rate of 48828.125 Hz, and data were acquired at  
408 a sampling rate of 24414.0625 Hz using a TDT system 3 recording setup (Tucker Davis  
409 Technologies).

410

411



412 ***Data Analysis***

413

414 ***Tapping Analysis***

415 To calculate consensus tap times, the histogram of tap times, pooled across the 40 subjects and  
416 then binned using 2 ms bins, was smoothed using a Gaussian kernel with a width (standard  
417 deviation) of 40 ms. This width was chosen because visual inspection of tap histograms showed  
418 the standard deviation around taps to be approximately 40 ms, so a Gaussian kernel with that  
419 width would approximate a “matched filter.” The precise width of the smoothing kernel was not  
420 critical to our results as long as it roughly matched the spread in the data. A peak-finder  
421 (*findpeaks.m*, built-in Matlab function) was then used to identify peaks that were larger than 40%  
422 of the maximum value in the smoothed histogram. The consensus inter-tap-interval (ITI) for a  
423 song was taken to be the mean interval between successive peaks, after the exclusion of intervals  
424 larger than 1.5 times the median inter-peak-interval (which would happen if the peak-finder  
425 missed a peak). The consensus phase was determined by finding the offset that optimally aligned  
426 a temporal grid with consensus ITI spacing with the peaks found by the peak-finder. Consensus  
427 tap times can be described by a consensus ITI (beat period) and consensus offset (beat phase)  
428 combination for each song.

429

430 On-beat neural activity was defined as the average population firing rate in the 100 ms following  
431 consensus tap times, and off-beat neural activity was the average population firing rate during all  
432 time excluding these on-beat windows. The justification for this definition is that (i) the true  
433 perceived beat location is almost certainly just after a listener taps, given the well documented  
434 tendency of listeners to anticipate the beat with their movements by several tens of milliseconds  
435 (negative beat asynchrony)<sup>61</sup>, (ii) defining off-beat activity as all neural activity that is not on the  
436 beat is consistent with previous work<sup>10</sup>, and (iii) an interval of 100 ms is less than one half a beat  
437 cycle for the fastest beat period observed in these data of 273 ms. The precise choice of time  
438 window is not critical, and this was confirmed by running all analyses using on-beat windows  
439 that ranged between 40 ms and 120 ms in 10 ms increments. The results were entirely consistent  
440 with those presented here for a time window of 100 ms, and if anything, slightly stronger when  
441 shorter time windows were used.

442

443 To compute the strength of the consensus, an “ideal tap histogram” was constructed by assuming  
444 all 40 listeners tapped precisely at each consensus beat time as determined by the excerpt’s  
445 consensus ITI and phase. A realistic degree of motor error was added by convolving this with a  
446 Gaussian kernel whose width was 5% of the beat period. The same 5% Gaussian kernel was then  
447 used for kernel density estimation on the two signals: the raw pooled histogram of (measured)  
448 tap times that already contained motor error, and the idealized tap histogram with motor error  
449 added. The choice of temporal filter value was guided by the magnitude of errors reported in  
450 studies of human sensorimotor synchronization<sup>77-79</sup>, but other kernel widths close to 5% also  
451 produce consistent results. The correlation coefficient between real and idealized tap density  
452 estimates for a given musical excerpt was taken as a measure of the strength of the tapping  
453 consensus, where a large value would indicate a high degree of similarity between real and  
454 “ideal” tapping behavior. Estimation of the real and idealized tap densities is also possible using  
455 a constant width (e.g. 40 ms) Gaussian kernel rather than a proportional one. However, while  
456 doing so would lead to the to the same main result shown in Fig 4, this measure of tapping  
457 consensus strength would have the undesirably effect of also correlating with song tempo since,  
458 as mentioned above, it is well-established that the magnitude of sensorimotor synchronization  
459 errors scale with interval duration.

460

#### 461 *Electrophysiology Data Preprocessing*

462 Offline spike sorting and clustering was done on the raw data using an automated expectation-  
463 maximization algorithm (Spikedetekt/Klustakwik)<sup>80</sup>, and clusters were manually sorted using  
464 Klustaviewa (Cortical Processing Lab, University College London). Firing rates over time for  
465 multi-units were calculated by binning spike times into 5 ms bins, which resulted in peri-  
466 stimulus time histograms (PSTHs) at an effective sampling rate of 200 Hz.

467

468 To determine whether spikes were reliably stimulus-driven, a noise power to signal power cutoff  
469 of 40 was chosen<sup>81</sup>. Song 1 was arbitrarily chosen to be the stimulus for which the  
470 repeatability of responses was measured. Units that failed to show a noise power to signal power  
471 ratio less than 40 based on the 12 repeats were excluded from further analysis, leaving a total of  
472 98 multi-units. All subsequent analyses were performed using custom-written Matlab code.

473

#### 474 *Fitting the LN Model*

475 The relevant scripts used at all stages of this process are available on Github<sup>82</sup>. First, music  
476 stimuli were transformed into a simple approximation of the activity pattern received by the  
477 auditory pathway by calculating the log-scaled spectrogram ('cochleagram')<sup>82-84</sup>. For each sound,  
478 the power spectrogram was taken using 10 ms Hanning windows, overlapping by 5 ms. The  
479 power across neighboring Fourier frequency components was then aggregated using overlapping  
480 triangular windows comprising 27 frequency channels with center frequencies ranging from 50  
481 Hz to 20,319 Hz (1/3 octave spacing). Next, the log was taken of the power in each time-  
482 frequency bin, and finally any values below a low threshold were set to that threshold. These  
483 calculations were performed using code adapted from `melbank.m`  
484 (<http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html>). The STRF model was trained to  
485 predict the firing rate at time  $t$  from a snippet of the cochleagram extending 100 ms (20 time  
486 bins) back in time from time  $t$ . The linear weights describing the firing rate of each neuron were  
487 estimated by regressing, with elastic net regularization, each neuron's firing rate at each time  
488 point against the 100 ms cochleagram snippet directly preceding it. Regularization strength was  
489 set by using a randomly chosen 10% of time bins from the cross-validation set as a validation set,  
490 and then by choosing the regularization parameters that led to the fit on the validation set with  
491 the lowest mean squared error. A sigmoidal nonlinearity<sup>85</sup> was then fitted to map from the linear  
492 activation to the predicted PSTH such that it minimized the error between the predicted PSTH  
493 and the observed PSTH. LN model predictions of a unit's PSTH to a test song were made by first  
494 convolving the cochleagram of the test song with the linear STRF and then applying the  
495 nonlinearity. Each unit's LN model was calculated 20 times, each time setting a different song  
496 aside as the test set. This was done so that PSTH predictions for any musical excerpt were true  
497 predictions since that excerpt was not included in the training set for the model.

498

## 499 **References**

- 500 1. Van Dyck, E. *et al.* Spontaneous Entrainment of Running Cadence to Music Tempo.  
501 *Sports Medicine - Open* **1**, 15–14 (2015).
- 502 2. Repp, B. H. & Su, Y.-H. Sensorimotor synchronization: A review of recent research  
503 (2006–2012). *Psychon Bull Rev* **20**, 403–452 (2013).
- 504 3. Schachner, A., Brady, T. F., Pepperberg, I. M. & Hauser, M. D. Spontaneous Motor  
505 Entrainment to Music in Multiple Vocal Mimicking Species. *Current Biology* **19**, 831–  
506 836 (2009).
- 507 4. Honing, H., Merchant, H., Háden, G. P., Prado, L. & Bartolo, R. Rhesus Monkeys  
508 (Macaca mulatta) Detect Rhythmic Groups in Music, but Not the Beat. *PLoS ONE* **7**,  
509 e51369 (2012).
- 510 5. London, J. Cognitive Constraints on Metric Systems: Some Observations and Hypotheses.  
511 *Music Perception: An Interdisciplinary Journal* **19**, 529–550 (2002).
- 512 6. Drake, C. & Ben El Heni, J. Synchronizing with music: intercultural differences. *Ann. N.*  
513 *Y. Acad. Sci.* **999**, 429–437 (2003).
- 514 7. Jacoby, N. & McDermott, J. H. Integer Ratio Priors on Musical Rhythm Revealed Cross-  
515 culturally by Iterated Reproduction. *Current Biology* **27**, 359–370 (2017).
- 516 8. Large, E. W., Herrera, J. A. & Velasco, M. J. Neural Networks for Beat Perception in  
517 Musical Rhythm. *Frontiers in Systems Neuroscience* **9**, 583 (2015).
- 518 9. Todd, N. P. M. & Lee, C. S. The sensory-motor theory of rhythm and beat induction 20  
519 years on: a new synthesis and future perspectives. *Front Hum Neurosci* **9**, 357 (2015).
- 520 10. Rajendran, V. G., Harper, N. S., Garcia-Lazaro, J. A., Lesica, N. A. & Schnupp, J. W. H.  
521 Midbrain adaptation may set the stage for the perception of musical beat. *Proc. Biol. Sci.*  
522 **284**, 20171455 (2017).
- 523 11. Fujioka, T., Trainor, L. J., Large, E. W. & Ross, B. Beta and Gamma Rhythms in Human  
524 Auditory Cortex during Musical Beat Processing. *Ann. N. Y. Acad. Sci.* **1169**, 89–92  
525 (2009).
- 526 12. Iversen, J. R., Repp, B. H. & Patel, A. D. Top-Down Control of Rhythm Perception  
527 Modulates Early Auditory Responses. *Ann. N. Y. Acad. Sci.* **1169**, 58–73 (2009).
- 528 13. Fujioka, T., Ross, B. & Trainor, L. J. Beta-Band Oscillations Represent Auditory Beat and

- 529 Its Metrical Hierarchy in Perception and Imagery. *J Neurosci* **35**, 15187–15198 (2015).
- 530 14. Doelling, K. B. & Poeppel, D. Cortical entrainment to music and its modulation by  
531 expertise. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E6233–42 (2015).
- 532 15. Snyder, J. S. & Large, E. W. Gamma-band activity reflects the metric structure of  
533 rhythmic tone sequences. *Cognitive Brain Research* **24**, 117–126 (2005).
- 534 16. Zanto, T. P., Snyder, J. S. & Large, E. W. Neural correlates of rhythmic expectancy.  
535 *Advances in Cognitive Psychology* **2**, 221–231 (2006).
- 536 17. Schroeder, C. E., Wilson, D. A., Radman, T., Scharfman, H. & Lakatos, P. Dynamics of  
537 Active Sensing and perceptual selection. *Current Opinion in Neurobiology* **20**, 172–176  
538 (2010).
- 539 18. Patel, A. D. & Iversen, J. R. The evolutionary neuroscience of musical beat perception:  
540 the Action Simulation for Auditory Prediction (ASAP) hypothesis. *Frontiers in Systems*  
541 *Neuroscience* (2014). doi:10.3389/fnsys.2014.00057/abstract
- 542 19. Morillon, B., Hackett, T. A., Kajikawa, Y. & Schroeder, C. E. ScienceDirect Predictive  
543 motor control of sensory dynamics in auditory active sensing. *Current Opinion in*  
544 *Neurobiology* **31**, 230–238 (2015).
- 545 20. Merchant, H. & Honing, H. Are non-human primates capable of rhythmic entrainment?  
546 Evidence for the gradual audiomotor evolution hypothesis. *Front Neurosci* **7**, (2014).
- 547 21. Tal, I. *et al.* Neural Entrainment to the Beat: The ‘Missing-Pulse’ Phenomenon. *J.*  
548 *Neurosci.* **37**, 6331–6341 (2017).
- 549 22. Grahn, J. A. The Role of the Basal Ganglia in Beat Perception. *Ann. N. Y. Acad. Sci.* **1169**,  
550 35–45 (2009).
- 551 23. MacDougall, H. G. & Moore, S. T. Marching to the beat of the same drummer: the  
552 spontaneous tempo of human locomotion. *J. Appl. Physiol.* **99**, 1164–1173 (2005).
- 553 24. Maes, P.-J., Leman, M., Palmer, C. & Wanderley, M. M. Action-based effects on music  
554 perception. *Front Psychol* **4**, 1008 (2014).
- 555 25. Zatorre, R. J., Chen, J. L. & Penhune, V. B. When the brain plays music:  
556 auditory|motor interactions in music perception and production. *Nat Rev Neurosci*  
557 **8**, 547–558 (2007).
- 558 26. Geiser, E., Notter, M. & Gabrieli, J. D. E. A corticostriatal neural system enhances  
559 auditory perception through temporal context processing. *J. Neurosci.* **32**, 6177–6182

- 560 (2012).
- 561 27. Patel, A. D. The Evolutionary Biology of Musical Rhythm: Was Darwin Wrong? *PLoS*  
562 *Biol* **12**, e1001821–6 (2014).
- 563 28. McKinney, M. F. & Moelants, D. Ambiguity in Tempo Perception: What Draws Listeners  
564 to Different Metrical Levels? *Music Perception: An Interdisciplinary Journal* **24**, 155–166  
565 (2006).
- 566 29. Zilany, M. S. A., Bruce, I. C. & Carney, L. H. Updated parameters and expanded  
567 simulation options for a model of the auditory periphery. *J. Acoust. Soc. Am.* **135**, 283–  
568 286 (2014).
- 569 30. Zuk, N. J., Carney, L. H. & Lalor, E. C. Preferred Tempo and Low-Audio-Frequency Bias  
570 Emerge From Simulated Sub-cortical Processing of Sounds With a Musical Beat. *Front*  
571 *Neurosci* **12**, 349 (2018).
- 572 31. Honing, H., Bouwer, F. L. & Háden, G. P. in *Neurobiology of Interval Timing* **829**, 305–  
573 323 (Springer New York, 2014).
- 574 32. Smith, P. H. & Spirou, G. A. in *Integrative Functions in the Mammalian Auditory*  
575 *Pathway* **15**, 6–71 (Springer, New York, NY, 2002).
- 576 33. Malmierca, M. S. in *Encyclopedia of Computational Neuroscience* 155–186 (Springer,  
577 New York, NY, 2015). doi:10.1007/978-1-4614-6675-8\_286
- 578 34. Dean, I., Harper, N. S. & McAlpine, D. Neural population coding of sound level adapts to  
579 stimulus statistics. *Nat Neurosci* **8**, 1684–1689 (2005).
- 580 35. Dean, I., Robinson, B. L., Harper, N. S. & McAlpine, D. Rapid Neural Adaptation to  
581 Sound Level Statistics. *J Neurosci* **28**, 6430–6438 (2008).
- 582 36. Ingham, N. J. & McAlpine, D. Spike-Frequency Adaptation in the Inferior Colliculus. *J*  
583 *Neurophysiol* **91**, 632–645 (2004).
- 584 37. Bulkin, D. A. & Groh, J. M. Systematic mapping of the monkey inferior colliculus reveals  
585 enhanced low frequency sound representation. *J Neurophysiol* **105**, 1785–1797 (2011).
- 586 38. London, J. *Hearing in Time*. (Oxford University Press, 2012).  
587 doi:10.1093/acprof:oso/9780199744374.001.0001
- 588 39. van Noorden, L. & Moelants, D. Resonance in the Perception of Musical Pulse. *Journal of*  
589 *New Music Research* **28**, 43–66 (1999).
- 590 40. Large, E. W. in *Psychology of Time* (2008).

- 591 41. Large, E. W. & Jones, M. R. The dynamics of attending: How people track time-varying  
592 events. *Psychol Rev* **106**, 119–159 (1999).
- 593 42. Nozaradan, S., Peretz, I., Missal, M. & Mouraux, A. Tagging the Neuronal Entrainment to  
594 Beat and Meter. *J Neurosci* **31**, 10234–10240 (2011).
- 595 43. Chapin, H. L. *et al.* Neural Responses to Complex Auditory Rhythms: The Role of  
596 Attending. *Front Psychol* **1**, (2010).
- 597 44. Parent, A. & Hazrati, L.-N. Functional anatomy of the basal ganglia. I. The cortico-basal  
598 ganglia-thalamo-cortical loop. *Brain Research Reviews* **20**, 91–127 (1995).
- 599 45. Reale, R. A. & Imig, T. J. Auditory cortical field projections to the basal ganglia of the  
600 cat. *Neuroscience* **8**, 67–86 (1983).
- 601 46. Grahn, J. A. & Rowe, J. B. Finding and feeling the musical beat: striatal dissociations  
602 between detection and prediction of regularity. *Cereb. Cortex* **23**, 913–921 (2013).
- 603 47. Lewis, P. A., Wing, A. M., Pope, P. A., Praamstra, P. & Miall, R. C. Brain activity  
604 correlates differentially with increasing temporal complexity of rhythms during  
605 initialisation, synchronisation, and continuation phases of paced finger tapping.  
606 *Neuropsychologia* **42**, 1301–1312 (2004).
- 607 48. Kotz, S. A., Schwartze, M. & Schmidt-Kassow, M. Non-motor basal ganglia functions: A  
608 review and proposal for a model of sensory predictability in auditory language perception.  
609 *Cortex* **45**, 982–990 (2009).
- 610 49. Stupacher, J., Wood, G. & Witte, M. Neural Entrainment to Polyrhythms: A Comparison  
611 of Musicians and Non-musicians. *Front Neurosci* **11**, 1664 (2017).
- 612 50. Lehmann, A., Arias, D. J. & Schönwiesner, M. Tracing the neural basis of auditory  
613 entrainment. *Neuroscience* **337**, 306–314 (2016).
- 614 51. Zoefel, B., Oever, Ten, S. & Sack, A. T. The Involvement of Endogenous Neural  
615 Oscillations in the Processing of Rhythmic Input: More Than a Regular Repetition of  
616 Evoked Neural Responses. *Front Neurosci* **12**, 95 (2018).
- 617 52. Obleser, J., Henry, M. J. & Lakatos, P. What do we talk about when we talk about  
618 rhythm? *PLoS Biol* **15**, e2002794 (2017).
- 619 53. Breska, A. & Deouell, L. Y. Dance to the rhythm, cautiously: Isolating unique indicators  
620 of oscillatory entrainment. *PLoS Biol* **15**, e2003534 (2017).
- 621 54. Zhou, H., Melloni, L., Poeppel, D. & Ding, N. Interpretations of Frequency Domain

- 622 Analyses of Neural Entrainment: Periodicity, Fundamental Frequency, and Harmonics.  
623 *Front Hum Neurosci* **10**, 509–8 (2016).
- 624 55. Henry, M. J., Herrmann, B. & Grahn, J. A. What can we learn about beat perception by  
625 comparing brain signals and stimulus envelopes? *PLoS ONE* **12**, e0172454–17 (2017).
- 626 56. Celma-Miralles, A. & Toro, J. M. Beat Perception in a Non-Vocal Learner: Rats Can  
627 Identify Isochronous Beats. *Proceedings of the th International Conference on the*  
628 *Evolution of Language* 1–2 (2018).
- 629 57. Li, J. *et al.* Primary Auditory Cortex is Required for Anticipatory Motor Response. *Cereb.*  
630 *Cortex* 1–18 (2017). doi:10.1093/cercor/bhx079
- 631 58. Aschersleben, G. Temporal Control of Movements in Sensorimotor Synchronization.  
632 *Brain and Cognition* **48**, 66–79 (2002).
- 633 59. Winkler, I., Háden, G. P., Ladinig, O., Sziller, I. & Honing, H. Newborn infants detect the  
634 beat in music. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 2468–2471 (2009).
- 635 60. Repp, B. H. Perception of timing is more context sensitive than sensorimotor  
636 synchronization. *Perception & Psychophysics* **64**, 703–716 (2002).
- 637 61. Repp, B. H. Sensorimotor synchronization: a review of the tapping literature. *Psychon*  
638 *Bull Rev* **12**, 969–992 (2005).
- 639 62. Takeya, R., Kameda, M., Patel, A. D. & Tanaka, M. Predictive and tempo-flexible  
640 synchronization to a visual metronome in monkeys. *Scientific Reports* **7**, 6127 (2017).
- 641 63. Gámez, J. *et al.* Predictive rhythmic tapping to isochronous and tempo changing  
642 metronomes in the nonhuman primate. *Ann. N. Y. Acad. Sci.* **1423**, 396–414 (2018).
- 643 64. Grube, M., Cooper, F. E., Chinnery, P. F. & Griffiths, T. D. Dissociation of duration-  
644 based and beat-based auditory timing in cerebellar degeneration. *Proc Natl Acad Sci USA*  
645 **107**, 11597–11601 (2010).
- 646 65. Merchant, H., Zarco, W., Pérez, O., Prado, L. & Bartolo, R. Measuring time with different  
647 neural chronometers during a synchronization-continuation task. *Proc Natl Acad Sci USA*  
648 **108**, 19784–19789 (2011).
- 649 66. Breska, A. & Deouell, L. Y. Neural mechanisms of rhythm-based temporal prediction:  
650 Delta phase-locking reflects temporal predictability but not rhythmic entrainment. *PLoS*  
651 *Biol* **15**, e2001665–30 (2017).
- 652 67. Fitch, W. T. Rhythmic cognition in humans and animals: distinguishing meter and pulse



- 653 perception. 1–16 (2013). doi:10.3389/fnsys.2013.00068/abstract
- 654 68. Hauser, M. D. & McDermott, J. The evolution of the music faculty: a comparative  
655 perspective. *Nat Neurosci* **6**, 663–668 (2003).
- 656 69. Fitch, W. T. The biology and evolution of music: A comparative perspective. *Cognition*  
657 **100**, 173–215 (2006).
- 658 70. Jackendoff, R. Parallels and Nonparallels between Language and Music. *Music*  
659 *Perception: An Interdisciplinary Journal* **26**, 195–204 (2009).
- 660 71. Tajadura-Jiménez, A. *et al.* As Light as your Footsteps. in 2943–2952 (ACM Press, 2015).  
661 doi:10.1145/2702123.2702374
- 662 72. Visell, Y. *et al.* Sound design and perception in walking interactions. *Journal of Human*  
663 *Computer Studies* **67**, 947–959 (2009).
- 664 73. Rajendran, V. G., Teki, S. & Schnupp, J. W. H. Temporal Processing in Audition: Insights  
665 from Music. *Neuroscience* (2017). doi:10.1016/j.neuroscience.2017.10.041
- 666 74. Buhusi, C. V. & Meck, W. H. What makes us tick? Functional and neural mechanisms of  
667 interval timing. *Nat Rev Neurosci* **6**, 755–765 (2005).
- 668 75. Singer, Y. *et al.* Sensory cortex is optimized for prediction of future input. *eLife* 1–31  
669 (2018). doi:10.7554/eLife.31557.001
- 670 76. McKinney, M. F., Moelants, D., Davies, M. E. P. & Klapuri, A. Evaluation of Audio Beat  
671 Tracking and Music Tempo Extraction Algorithms. *Journal of New Music Research* **36**,  
672 1–16 (2007).
- 673 77. Peters, M. The relationship between variability of intertap intervals and interval duration.  
674 *Psychological Research* **51**, 38–42 (1989).
- 675 78. Thaut, M. H., Miller, R. A. & Schauer, L. M. Multiple synchronization strategies in  
676 rhythmic sensorimotor tasks: phase vs period correction. *Biol Cybern* **79**, 241–250 (1998).
- 677 79. Ivry, R. B. & Hazeltine, R. E. Perception and production of temporal intervals across a  
678 range of durations: evidence for a common timing mechanism. *J Exp Psychol Hum*  
679 *Percept Perform* **21**, 3–18 (1995).
- 680 80. Kadir, S. N., Goodman, D. F. & Harris, K. D. High-dimensional cluster analysis with the  
681 Masked EM Algorithm. *arXiv preprint arXiv:1309.2848* (2013).
- 682 81. Linden, M. S. J. F. How linear are auditory cortical responses? **15**, 125 (2003).
- 683 82. Ben D B Willmore. Benlib. *github.com* doi:

- 684 83. Ben D B Willmore, Schoppe, O., King, A. J., Schnupp, J. W. H. & Harper, N. S.  
685 Incorporating Midbrain Adaptation to Mean Sound Level Improves Models of Auditory  
686 Cortical Processing. *J Neurosci* **36**, 280–289 (2016).
- 687 84. Harper, N. S. *et al.* Network Receptive Field Modeling Reveals Extensive Integration and  
688 Multi-feature Selectivity in Auditory Cortical Neurons. *PLoS Comput Biol* **12**, e1005113–  
689 30 (2016).
- 690 85. Rabinowitz, N. C., Willmore, B. D. B., Schnupp, J. W. H. & King, A. J. Spectrotemporal  
691 Contrast Kernels for Neurons in Primary Auditory Cortex. *J Neurosci* **32**, 11271–11284  
692 (2012).
- 693