
Research Article: New Research | Sensory and Motor Systems

Temporal correlates to monaural edge pitch in the distribution of inter-spike interval statistics in the auditory nerve

<https://doi.org/10.1523/ENEURO.0292-21.2021>

Cite as: eNeuro 2021; 10.1523/ENEURO.0292-21.2021

Received: 7 July 2021

Accepted: 7 July 2021

This Early Release article has been peer-reviewed and accepted, but has not been through the composition and copyediting processes. The final version may differ slightly in style or formatting and will contain links to any extended data.

Alerts: Sign up at www.eneuro.org/alerts to receive customized email alerts when the fully formatted version of this article is published.

Copyright © 2021 Li and Joris

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

Temporal correlates to monaural edge pitch in the distribution of inter-spike interval statistics in the auditory nerve.

Abbreviated title: Correlates to edge pitch

Authors: Yi-Hsuan Li (1), Philip X. Joris (1)

Affiliations:

(1) Laboratory of Auditory Neurophysiology, Medical School, Campus Gasthuisberg, K. U. Leuven, B-3000 Leuven, Belgium

Corresponding author: Philip X. Joris; philip.joris@med.kuleuven.be

Number of pages: 38

Number of figures: 13

Number of words: Abstract (248), Introduction (650), Discussion (1499)

Acknowledgements: Supported by the Fund for Scientific Research – Flanders to PXJ

(G0B2917N and G085421N). YHL was supported by a Ministry of Education (KU Leuven – Taiwan) scholarship.

Conflict of interest statement: The authors declare no competing financial interests.

Author contributions: YHL and PXJ designed research; YHL and PXJ performed research; YHL analyzed data; YHL and PXJ wrote the paper.

1 **Abstract**

2 Pitch is a perceptual attribute enabling perception of melody. There is no consensus regarding
3 the fundamental nature of pitch and its underlying neural code. A stimulus which has received
4 much interest in psychophysical and computational studies is noise with a sharp spectral edge.
5 High- or low-pass noise gives rise to a pitch near the edge frequency (“monaural edge pitch”,
6 MEP). The simplicity of this stimulus, combined with its spectral and autocorrelation properties,
7 make it an interesting stimulus to examine spectral versus temporal cues that could underly its
8 pitch. We recorded responses of single auditory nerve fibers in chinchilla to MEP-stimuli
9 varying in edge frequency. Temporal cues were examined with shuffled autocorrelogram (SAC)
10 analysis. Correspondence between the population’s dominant interspike interval and reported
11 pitch estimates was poor. A fuller analysis of the population interspike interval distribution,
12 which incorporates not only the dominant but all intervals, results in good matches with
13 behavioral results, but not for the entire range of edge frequencies that generates pitch. Finally,
14 we also examined temporal structure over a slower time scale, intermediate between average
15 firing rate and interspike intervals, by studying the SAC envelope. We found that, in response to
16 a given MEP stimulus, this feature also systematically varies with edge frequency, across fibers
17 with different characteristic frequency. Because neural mechanisms to extract envelope cues are
18 well-established, and because this cue is not limited by coding of stimulus fine-structure, this
19 newly identified slower temporal cue is a more plausible basis for pitch than cues based on fine-
20 structure.

21 **Significance Statement**

22 A longstanding debate concerns the neural underpinnings of pitch, which is a label the brain
23 computes for periodic sounds. Perceptual studies have not resolved whether pitch is based on
24 spectral or temporal cues, or both. Because the neural processing requirements for temporal and
25 place cues are very different, neurophysiological data can in principle resolve this debate. We
26 studied responses of neurons in the auditory nerve to a simple aperiodic stimulus and examined
27 candidate cues that may underly its unusual pitch. We find that fine temporal cues could
28 potentially underly edge pitch, but only for a restricted range over which it is observed
29 behaviorally. The data draw attention to a temporal cue at a slower time scale than is
30 traditionally considered.

31 **Introduction**

32 Pitch is what allows us to hear melodies and is important in communication by speech and to
33 separate multiple sound sources. It is not a physical but a perceptual attribute, i.e. a “label” that
34 the brain creates for sounds with a repeating waveform, produced by vibrating sources. This
35 label must obviously be coded by activity patterns conveyed by the auditory nerve (AN) to the
36 auditory central nervous system (CNS), but there is no agreement on the nature of that code. The
37 century-old pitch debate has centered on the role of spectral *versus* temporal cues. Repetition in a
38 sound waveform not only leads to a temporal cue (periodicity) but also to a spectral cue
39 (harmonicity), and the cues in those two domains are inextricably linked, which explains the
40 difficulty in designing conclusive perceptual experiments. On the other hand, the neural
41 processing requirements for temporal *versus* spectral cues are very different, so that
42 neurophysiological studies ought to be able to settle the pitch debate. The cues accessible to the
43 CNS are those provided by the output of the cochlea. Thus, AN studies are a critical component

44 of the neurophysiological study of pitch (Javel, 1980; Evans, 1983; Cariani and Delgutte, 1996a,
45 1996b; Cedolin and Delgutte, 2010; Bidelman and Heinz, 2011; Kale et al., 2014). These studies
46 show that both temporal and rate cues are viable, albeit with different difficulties and limitations
47 (Winter, 2005; Cedolin and Delgutte, 2010).

48 A stimulus which has long been studied perceptually (von Békésy, 1963) and received
49 renewed attention is simply noise with a flat spectrum with a sharp cutoff. Both high-pass and
50 low-pass noise produce so-called “edge pitch”: a pitch close to the actual edge frequency but
51 slightly mismatched, by a few % (Klein and Hartmann, 1981). For high-pass (HP) noise, the
52 pitch is marginally higher than the edge frequency, while it is somewhat below the edge
53 frequency for low-pass (LP) noise, so the pitch shifts “towards” the passband. The degree of shift
54 decreases with increasing edge frequency.

55 While edge pitch is not very strong, this stimulus has the advantage that few cues are
56 available on which the pitch can be based. The stimulus is not periodic, and its long-term
57 spectrum is flat over its passband. We examined whether fine spike timing could be the cue
58 underlying edge pitch. It has been proposed that the combination of phase-locking, neural delays,
59 and coincidence detection could provide the neural substrate to perform a running
60 autocorrelation (Licklider, 1951; Meddis and Hewitt, 1991), which basically tallies inter-spike
61 intervals. Indeed, for a range of pitch stimuli, the dominant interspike-interval across a
62 population of AN fibers is well-correlated to pitch frequency and strength (Cariani and Delgutte,
63 1996a). In response to noise that exceeds their bandwidth, AN fibers show a distinct temporal
64 response (Ruggero, 1973; Louage et al., 2004) which differs between fibers tuned to different
65 frequencies. However, when the noise edge is within a fiber’s receptive field, its temporal output
66 is modified (Ruggero, 1973). It has been hypothesized (Cariani et al., 2015; Hartmann et al.,

67 2019) that this interaction between frequency tuning and stimulus edge results in an interspike
68 interval distribution which predicts the perceived edge pitch, i.e. slightly shifted from the edge
69 towards the passband.

70 Besides fine spike timing, we also examine two other potential cues. First, lateral neural
71 mechanisms, similar to lateral inhibition in visual and somatosensory systems, could cause a
72 peak in the neural activity pattern (von Békésy, 1963). Such inhibition is not present in the AN
73 but there is a cochlear mechanical source of lateral interaction, which may cause a rate increase
74 in neurons tuned near the edge frequency (Ruggero, 1973; Schalk and Sachs, 1980). Second, the
75 interaction of noise with a neuron's receptive field not only affects spike rate and precise spike
76 timing, but also spike timing on a longer time scale. Such slower aspects of temporal response
77 patterns have also been proposed as perceptual cues (Sinex, 2005; Carney, 2018).

78 **Methods**

79 ***Recording***

80 Our methods for single-unit recording of AN fibers followed those of Louage et al. (2004). Data
81 were collected from thirteen chinchillas (*chinchilla lanigera*) of either sex obtained from a
82 breeding colony at our institution. All procedures were approved by the K.U. Leuven Ethics
83 Committee for Animal Experiments and were in accordance with the National Institutes of
84 Health Guide for the Care and Use of Laboratory Animals. The animals were placed on a heating
85 pad in a double-walled sound-attenuated chamber and were under anesthesia for the duration of
86 the experiment. Induction of anesthesia started with an s.c. injection of 0.05ml xylazine (XYL-M
87 2%), followed by i.m. injection of a mixture of equal volumes of ketamine (Nimatek, 100mg/ml)
88 and medetomidine (Domitor, 1mg/ml) at an initial dose of 0.05 ml i.m. Maintenance of
89 anesthesia was with the same mixture and dose, administered i.m., titrated based on reflexes and

90 vital signs. In some animals, initial anesthesia by ketamine/medetomidine was followed by i.p.
91 injections of 0.05 ml of pentobarbital (Nembutal, 60 mg/ml), diluted by mixing with an equal
92 volume of sterile saline. A tracheotomy was performed to allow for mechanical respiration. The
93 AN was exposed through a posterior fossa approach, involving the removal of a lateral portion of
94 cerebellum. Single AN fibers were isolated with glass micropipettes filled with 3 M KCl,
95 inserted into the nerve trunk under visual guidance. The cartilaginous ear canal was removed to
96 expose the bony meatus. The sound stimulus was compensated for the acoustic transfer function
97 measured at the bony meatus and was delivered with a dynamic speaker (Etymotic Research,
98 ER1 or ER2) coupled to an ear bar inserted into the bony meatus. The neural signal was
99 amplified, filtered, timed, and displayed using standard techniques.

100 *Stimuli*

101 Pure tone stimuli were used while searching for nerve fibers and provided a first estimate of
102 frequency tuning. A threshold tuning curve was obtained by a tracking algorithm (Geisler et al.,
103 1985) which provided spontaneous rate (SR) and characteristic frequency (CF: frequency of
104 lowest threshold). Fibers with $SR < 18$ spikes/s and ≥ 18 spikes/s are referred to as low-SR and
105 high-SR, respectively. Responses were obtained to broadband noise as well as to two groups of
106 MEP stimuli: HP and LP noise. Broadband (0.1 - 4 kHz or 0.1 – 8 kHz) noise was used as a
107 reference stimulus (see Results): it is broadband in relationship to the fibers studied, i.e. it covers
108 their tuning curve, and it is also broadband relative to the “dominance” region for edge pitch,
109 which is roughly between 200 and 1000 Hz (Klein and Hartmann, 1981). HP noise had a fixed
110 upper bound at 4 or 8 kHz, or even higher if warranted by the fiber’s CF, while LP noise had a
111 fixed lower bound at 0.1 kHz. We varied the edge frequency, which was the lower cutoff for HP
112 noise and the upper cutoff for LP noise. Every stimulus was presented for nominally 40

113 repetitions with duration 600 ms. In order to obtain sufficient data for population analysis, we
114 chose five standard edge frequencies: 0.25, 0.5, 1, 1.5, and 2 kHz, for both LP and HP MEP
115 stimuli. Therefore, the standard stimulus battery consisted of broadband noise and 10 MEP
116 stimuli. The overall sound level was usually 60 to 80 dB in 10 dB steps. SPL was varied over a
117 wider range of levels for some fibers, at the beginning of the experiment, to assess the threshold
118 level. For every AN fiber, we started with the broadband noise stimulus and the MEP stimuli (LP
119 and HP) whose edge frequency was closest to CF. If recording time for a given fiber allowed, we
120 also studied its response to variation of edge frequency with fine frequency steps (~10 Hz),
121 and/or tested additional SPLs. In a limited number of fibers, we parametrically varied edge
122 frequency while keeping spectral level constant, usually at 20 dB/Hz.

123 *Analysis of temporal structure*

124 For every stimulus, responses to 40 repetitions of the sound stimulus provided 40 non-
125 identical spike trains in response to the same stimulus. For each fiber, the temporal structure of
126 the spike train in response to a given stimulus was captured by calculating shuffled
127 autocorrelograms (SAC, Figure 1 A) of the spike trains (Joris, 2003; Joris et al., 2006). The SAC
128 provides a summary of the temporal pattern of the spike train “locked” to the stimulus. It is
129 calculated by selecting two non-identical spike trains and counting the number of coincident
130 spikes for different time delays τ (up to 50 ms) between spike trains, and repeating this process
131 for all possible pairs of spike trains while omitting pairs of identical spike trains, to avoid
132 refractory effects. Coincidence was defined as two spikes occurring within a certain time
133 window (default = 50 μ s, decreased to 30 μ s for responses to high edge frequencies). Note that
134 counting coincidences for delay τ is identical to counting inter-spike intervals equal to τ . Each
135 SAC was normalized to the number of stimulus repetitions, stimulus duration and firing rate

136 (Louage et al., 2004). When normalized for all these parameters, uncorrelated spike trains result
137 in a value of 1, and SACs to broadband noise tend toward that value for large delays (Fig. 1A).
138 When calculating fluctuation strength (below, section “Fluctuations”), we use SACs that are not
139 normalized for firing rate: we refer to these as unnormalized SACs. Every peak in the SAC
140 indicates the fiber is firing frequently at this interval, while every trough indicates a paucity of
141 firing: such features are temporal cues potentially used by the brain (see Discussion).

142 To quantify periodicities in the SAC, we use a method of comparison with harmonic lag
143 pulse trains (Fig. 1B)(Tramo et al., 2001; Cariani et al., 2015). In initial analysis, the pulse trains
144 varied in frequency from 100 to 4000 Hz in 1Hz steps but based on the results (Fig. 3C), the
145 upper limit was subsequently lowered to 2000 Hz. Note that the largest peak of the SAC is
146 always the one straddling zero delay, which does not capture any information on periodicity in
147 the response: we ignored this peak by having no pulse at 0 ms in the harmonic lag pulse trains.
148 For each SAC, a correlation profile (Fig. 1C) is obtained by calculating the Pearson correlation
149 of the SAC to harmonic lag pulse trains with the same timespan and binwidth. The periodicity of
150 the harmonic lag pulse train with the highest correlation value was chosen as the dominant
151 interval. For simplicity we label this interval the “estimated pitch frequency” for that fiber: it can
152 be viewed as the pitch for which that single fiber “votes”. The correlation value at this interval is
153 labeled the “pitch salience” (Cariani et al., 2015). As an example, Figure 1A illustrates a SAC
154 with prominent peaks at delays near 2 and 4 ms. Fig. 1B shows harmonic lag pulse trains at
155 different frequencies. Correlation of pulse trains with the SAC results in the correlation profile
156 shown in Fig. 1C, revealing a maximum near 500 Hz. The correlation profiles calculated for
157 individual SACs allow identification of the fibers that most strongly contribute intervals

158 corresponding to a given frequency, thereby revealing the contribution of every AN fiber to the
159 pitch estimate (see Fig. 11).

160 To characterize the distribution of intervals across a population of fibers for a given MEP
161 stimulus, we collected responses from a number of AN fibers of different CF to that stimulus.
162 We then summed all their SACs and refer to this sum as the population interval distribution
163 (PID). As a rule of thumb we required responses from more than ten fibers with CFs distributed
164 over a range re. the edge frequency. Two procedures were used to extract the dominant
165 periodicity of the PID. The first procedure is that of Cariani and Delgutte (1996a), where simply
166 the most frequently occurring interval of the PID is identified, again ignoring the maximum at
167 zero delay. The second procedure is the method using harmonic lag pulse trains as described
168 above for the SAC (Tramo et al., 2001; Cariani et al., 2015), but now applied to the PID of a
169 population of fibers rather than to SACs of individual fibers, yielding a population pitch
170 frequency and pitch salience (see also periodic sieves in Bidelman and Heinz, 2011). Examples
171 of this analysis are shown in Figs. 8,9. Note that this procedure weighs not only the interspike
172 interval corresponding to a certain F_0 but also all subharmonic intervals up to 15 ms.

173 *Fluctuations*

174 We expect not only the fine temporal structure of the response of AN fibers to be affected by
175 MEP stimuli, but also the distribution of spikes over a slower time scale, referred to as
176 “fluctuations” (Carney, 2018). Different analysis methods were explored to quantify fluctuations,
177 which are not easily captured as they are “induced” by cochlear filtering and other cochlear
178 processes (Joris, 2003) and do not simply equate a stimulus parameter. When the cochlear filter
179 is covered by the noise bandwidth, fast fluctuations are expected, while partial coverage is
180 expected to result in slower fluctuations. Because these fluctuations are reflected in the envelope

181 of the SAC, we quantified their time scale by the half-width of the SAC envelope (Fig. 2). In
182 previous studies of our laboratory, the SAC envelope was obtained from the so-called difcor (a
183 subtraction of autocorrelograms to noise and its polarity-inversion) (Joris et al., 2005, 2008). In
184 the present study, responses to polarity-inverted stimuli were not obtained, and the SAC
185 envelope was extracted by fitting a Gaussian curve to the local maxima straddling 0 ms (Fig. 2).
186 To fit the Gaussian, a DC value of 1 was subtracted from all SAC values so that 0 indicates the
187 level of uncorrelated response. To reduce high-frequency noise, the SAC was filtered by a 3-
188 point moving average. For low-CF cells, a minimum distance of 10 sampling points between
189 local maxima was required, and maxima needed to exceed a value of 0.1 with a minimum peak
190 prominence of 0.1 (MATLAB function FINDPEAKS). The time scale of the fluctuation was
191 then defined as the half-width (HW) value of the Gaussian. If only one maximum could be
192 identified in the SAC (in high-CF fibers, Fig. 3B), no Gaussian fit was performed and HW was
193 measured on the SAC itself. To also take into account the magnitude of fluctuation, we
194 multiplied HW with the coincidence rate or CR (Joris et al., 2006), which is the maximum
195 amplitude of the unnormalized SAC, at 0 ms (dimension: (spikes/s)²). High values of HW*CR
196 are obtained for slow but strong fluctuations. This metric, which we refer to as fluctuation
197 strength, is examined for both single fiber and population fiber data.

198 **Results**

199 *Characterization of response to broadband noise*

200 Data to MEP stimuli were obtained for 231 AN fibers. Fig. 3A gives an overview of SR and CF
201 of the fibers studied. Edge pitch is most prominent for edge frequencies below a few kHz, so our
202 sample is biased accordingly. We first discuss responses to broadband noise, which does not

203 evoke a pitch percept but serves as an important baseline to interpret responses to MEP stimuli.
204 Figure 3B shows SACs for 82 AN fibers with CF spanning a range of 0.1 to 9.5 kHz. For AN
205 fibers with $CF < 2$ kHz, SACs show a damped oscillation of increasing frequency with
206 increasing CF and for $CF > 2$ kHz the SACs are increasingly dominated by a single central peak
207 and lose the oscillatory component. These features are consistent with previous reports (Joris,
208 2003; Louage et al., 2004). The top trace in Fig. 3B (red line) is the average of all the SACs: the
209 PID. Except for a peak centered at 0 delay, the PID is featureless and lacks a clear oscillatory
210 component. This indicates that no interspike interval dominates. The peak at zero delay indicates
211 that each fiber tends to fire spikes at a similar point in time across stimulus repetitions. Fig. 3C
212 (top, red trace) shows the correlation profile obtained by correlating the PID with harmonic lag
213 pulse trains over a range of frequencies. As expected, this correlation profile is also featureless
214 (in contrast to e.g. Figs. 7,8): the maximal value, i.e. the estimated pitch frequency, is at 215 Hz
215 (grey dot) but with very low correlation salience (0.015), indicating there is no dominant inter-
216 spike interval.

217 We repeated the same process for individual SACs (cf. Fig. 1C), to obtain an estimated pitch
218 frequency and salience from each fiber. The relation between CF and estimated pitch frequency
219 is shown by the scatterplot in Fig. 3C (bottom) where pitch salience is indicated by the grey
220 scale. A diagonal line indicates identity between single-fiber estimated pitch frequency and CF:
221 the data follow this trend for $CF < 2$ kHz but this relation breaks down when CF is larger than 2
222 kHz. As described earlier (Louage et al., 2004), the reason is that spike timing in AN fibers is
223 dominated by phase-locking to the fine-structure of the filtered and transduced broadband signal
224 for $CF < 2$ kHz and by phase-locking to its envelope for $CF > 2$ kHz. In the further analysis, we
225 use a 2 kHz upper limit on the frequency of the harmonic lag pulse trains. Importantly, there is

226 no consistency in the periodicities and single-fiber pitch estimates across the population, as was
227 already clear from the featureless PID (Fig. 3B, top) and its correlation profile (Fig. 3C, top).

228 The SACs in Fig. 3B show other temporal response features that change with CF and that
229 have been documented earlier (Louage et al., 2004; Joris et al., 2008). First, peaks and troughs
230 occur across a wider range of delays at low than at high CFs. This feature reflects the increasing
231 absolute bandwidth of frequency tuning with CF and is captured here by quantifying the HW of
232 the SAC envelope for CFs < 2 kHz and of the SAC itself for CFs > 2 kHz (see Methods and Fig.
233 2). Second, the peak height of SACs in response to broadband noise tends to decrease with CF.
234 We use fluctuation strength (HW*CR, see Methods), to obtain a metric that reflects the “energy”
235 of fluctuation and which incorporates both width and height. Fig. 3D shows that, in response to
236 broadband noise, fluctuation strength tends to decrease with CF.

237 *Single neuron responses to MEP stimuli*

238 Although the main experimental approach was to look at population responses to a limited
239 number of MEP stimuli, for a small number of fibers we obtained responses to MEP stimuli
240 parametrically changed in cutoff frequency. Such responses are inherently easier to analyze and
241 interpret, and are therefore presented first.

242 Fig. 4 shows responses of two fibers, with a CF in the phase-locking range, to LP (top row)
243 and HP (bottom row) noise with varying cutoff frequency. The first column of panels shows
244 tuning curves and firing rates; subsequent panels show analysis of the fine temporal structure of
245 the response (Fig. 4B,C,G,H) and the fourth and fifth columns show analysis of its slow temporal
246 structure (Fig. 4D,E,I,J).

247 It has been suggested (von Békésy, 1963) that a process akin to lateral inhibition could
248 accentuate the representation of the edge in terms of firing rate and cause edge pitch. AN fibers
249 feature sideband suppression (Katsuki et al., 1959; Sachs and Kiang, 1968). We therefore
250 examined firing rate, measured over the entire stimulus duration, for different edge frequencies.
251 Fig. 4A,F shows the frequency tuning curve (black solid line; top ordinate, CF marked by a
252 dashed line) and the average firing rate (blue symbols and spline fit, bottom ordinate) for
253 changing cutoff frequency. The spectral level is constant (20 dB/Hz, overall level of 56 dB SPL
254 for the wideband stimulus). Entirely as expected, as the edge frequency increases, there is an
255 increase in firing rate to the LP stimuli (Fig. 4A) and a decrease to the HP stimuli (Fig. 4F). A
256 pronounced “hump” in firing rate was not observed for either the LP or HP condition.

257 The fine temporal structure captured by SACs is shown in the second column and the
258 estimated single-fiber pitch frequency in the third column. The SACs of the responses to LP
259 MEP stimuli (Fig. 4B) show clear and systematic changes with edge frequency. When the edge
260 frequency is high so that the noise covers much of the frequency tuning curve, the stimulus is
261 essentially broadband and the SACs are similar to that to broadband noise (Fig. 4B, top trace)
262 with an oscillation close to the fiber’s CF^{-1} . However, when the edge approaches CF and
263 decreases below CF so that only a fraction of the frequency tuning curve is covered, the
264 oscillation frequency of the SAC decreases. Finally, when the edge frequency reaches the edge
265 of the fiber’s frequency tuning, the firing rate drops to the level of spontaneous activity (Fig. 4A,
266 asterisk on ordinate) and the SAC flattens. Pitch estimates obtained by computing the correlation
267 between SACs and harmonic lag pulse trains over a 30 ms window (± 15 ms) are shown in Fig.
268 4C, where the greyscale of the dots shows the pitch salience (Fig. 4C,H, inset). Two dashed lines
269 indicate values of CF (black, vertical) and edge frequency (red, diagonal), respectively; their

270 intersection marks the point where edge frequency equals CF. The pitch estimates show a
271 broken-stick pattern: close to CF at high edge frequencies, and tracking the edge frequency over
272 a wide range at and below CF. Fig. 4H shows the same analysis for responses to HP stimuli for a
273 different fiber. Here, the bottom trace in Fig. 4G shows the response to the broadband condition.
274 As edge frequency increases and the noise covers progressively less of the fiber's frequency
275 tuning curve, there is little change in the SACs until the edge frequency exceeds CF (Fig. 4H,
276 crossing of black and red dashed lines) and the SAC oscillation frequency tracks the edge
277 frequency. However, the frequency range over which such tracking occurs is much narrower
278 than for LP noise (Fig. 4C). Note that the pitch salience reaches higher values (darker symbols)
279 in Fig. 4H than in Fig. 4C. This is due to the high frequency of oscillations for the HP edge near
280 2 kHz compared to the lower frequencies for the LP edges (< 1 kHz), so that more pulses of the
281 harmonic lag pulse trains fall within the analysis window and contribute to the positive
282 correlation.

283 Finally, the two rightmost columns (Fig. 4D,E,I, J) show the analysis of slow temporal
284 fluctuations. It is evident in the SACs as shown in Fig. 4B,G that changes in edge frequency not
285 only affect the periodicity of oscillation, but also cause a profound change in the range of delays
286 over which oscillations are found. Indeed, when the oscillation frequency tracks edge frequency
287 (Fig. 4C,H: pitch estimate follows red dashed line), the SACs show oscillations far outside the
288 portion of the delay axis bounded by the two vertical dashed lines (Fig. 4B,G). Fig. 4D,I repeats
289 the SACs but over a tenfold longer range of delays, and with a Gaussian fit to the SAC envelope.
290 The width of the Gaussian is clearly dependent on edge frequency. Fig. 4E,J shows
291 quantification of the maximum of the unnormalized SACs (CR, green line) and HW (blue line)
292 of the Gaussian envelope. For responses to the LP stimuli, SACs become very wide, with large

293 central peaks, for edge frequencies below CF (horizontal dashed line), while for responses to the
294 HP stimuli they become wide for edge frequencies above CF.

295 Fig. 5 shows the same analysis for responses of a fiber with high CF (8.6 kHz). The changes
296 in firing rate (Fig. 5A,F) are similar to those of the low-CF fibers. As expected, the SACs (Fig.
297 5B, G) only show one central mound and a complete absence of oscillations caused by fine-
298 structure. Pitch estimates based on the harmonic lag pulse train method, which quantifies interval
299 fine-structure up to 2 kHz, are correspondingly completely random with near-zero salience (Fig.
300 5C,H). In response to broadband noise, the SACs show a central mound of limited height and
301 bandwidth (Fig. 5B, top trace, Fig. 5G, bottom trace). However, as the stimulus cutoff
302 approaches the limits of the cochlear filter (on the low-frequency side for the LP condition, Fig.
303 5B,D, and on the high-frequency side for the HP condition, Fig. 5G,I), slow and high-amplitude
304 envelope fluctuations are generated through interaction of the limited noise bandwidth and
305 bandpass cochlear tuning. Thus, at high CFs there is no predominance of interspike intervals
306 close to the period of the cutoff frequency of the LP or HP noise when the stimulus edge is near
307 CF. But nevertheless, there is a temporal cue in the form of the overall interspike interval
308 distribution which covaries with the edge frequency relative to CF: both the width and amplitude
309 (Fig. 5E,J, HW and CR) of the SACs increase when the edge frequency crosses the CF and the
310 noise bandwidth only partially covers the tuning curve.

311 *Correlates of edge-frequency in population fine temporal structure*

312 Having examined how changes in edge frequency affect firing rate, fine temporal structure, and
313 fluctuations at the single-fiber level, we now turn to population analyses where we address the

314 complementary question: what is the response of a population of nerve fibers to a single stimulus
315 and which cues can we identify that may underlie its pitch?

316 Spike trains to each of the 5 LP and HP stimuli were collected from at least 10 AN fibers. For
317 each fiber the SAC was computed. Before showing complete datasets (Fig. 7), we illustrate the
318 main features of responses to 0.5 and 1 kHz edge frequencies for five AN fibers, chosen to have
319 adequately spaced CFs, to allow a direct comparison with the response to broadband noise from
320 the same fibers. The upper row of Figure 6 shows SACs for LP stimuli, the lower row for HP
321 stimuli. The pink background color indicates the frequency band of the stimuli. For each fiber,
322 the SAC to broadband noise is also shown (black line) to contrast with the SAC to the edge-pitch
323 stimulus (blue line). If the edge frequency is such that a fiber's tuning curve is covered by the
324 stimulus passband, the SAC is expected to be identical to that of the response to broadband
325 noise. This is indeed what we observe. Note that the broadband, LP, and HP stimuli were
326 delivered with identical overall SPL (70 dB) and therefore differ in spectrum level. SAC peak
327 amplitude and, more modestly, oscillation frequency depend to some extent on SPL (Louage et
328 al., 2004), which explains the small differences between black and blue lines sometimes seen
329 even for neurons with CFs well within the stimulus passband.

330 For LP noise stimuli, we expect fibers with low CF to be covered by the stimulus passband
331 and to have similar SACs to the LP and broadband noise, which is indeed the case. For fibers
332 with $CF > 0.5$ kHz, the SACs in response to 0.5 kHz LP noise show a slow oscillation, revealing
333 a preponderance of intervals somewhat above 2 ms, i.e. intervals longer than the period of the
334 edge frequency (illustrated with dashed vertical lines in Fig. 6 at that period and its integer
335 multiples) and much longer than in the SAC to broadband noise. A similar pattern is observed
336 for responses to 1-kHz LP noise. For the HP noise, the reverse observation is made: fibers with

337 CF below the edge frequency show faster oscillations in response to HP noise than to broadband
338 noise, at least for CFs quite close to the edge frequency.

339 Figure 7 shows similar plots with all SACs available for the 10 MEP stimuli. Every column
340 shows SACs (blue line) and their population average (PID, red line on top) in response to LP
341 (upper row) and HP (lower row) MEP stimuli with the same edge frequency. We start our
342 description with responses to 0.5 kHz, for which we have good sampling and for which the
343 responses are easily understood from previous figures. In response to LP noise, there is a striking
344 tendency for fibers with CF near the edge frequency to show a periodicity slightly longer than
345 that of the edge frequency (and its multiples, indicated with vertical dashed lines), and this
346 tendency extends to CFs several octaves above the edge frequency. As a result, the PID shows
347 clear peaks somewhat above 2 ms (blue arrows). A similar tendency is observed for HP noise in
348 the response of fibers with CF near the 0.5 kHz edge frequency, but less strikingly so: fibers with
349 CF just below 0.5 kHz show peaks somewhat smaller than 2 ms, and this tendency is much more
350 restricted in spatial extent. For example, fibers more than an octave below 0.5 kHz do not show a
351 preponderance of intervals just below 2 ms. Nevertheless, the PID shows quite prominent peaks
352 just below 2 ms (blue arrows), be it of smaller amplitude than in the PID to LP noise.

353 Note that the PID is an average of all SACs available for each condition, also for CFs
354 straddling the “non-relevant” stimulus edge (the lower edge, 0.1 kHz, of the LP stimulus; or the
355 upper edge, 4 kHz, of the HP stimulus). For the LP stimuli, the fibers with very low CF only
356 contribute a broad central peak to the PID; for the HP stimuli, the fibers with $CF > 4$ kHz only
357 contribute a narrow central peak. Thus, inclusion of these fibers has a negligible effect on the
358 location of local maxima in the PID.

359 For the 1 kHz MEP stimuli, the SACs and PID behave very much like for the 0.5 kHz
360 stimuli, and this seems also the case for the responses to the 1.5 and 2 kHz MEP stimuli, though
361 our CF sampling for these latter stimuli is less complete. For the 0.25 kHz MEP stimuli, the
362 SACs to the LP noise show a quite striking similarity across CFs with an oscillation period
363 slightly higher than 4 ms. However, in response to the HP noise, the SACs average out to a PID
364 that is nearly featureless except for a central peak.

365 *Psychoacoustic versus physiological pitch estimates*

366 From the preceding analysis, we have a single PID to summarize the overall temporal structure
367 of interspike intervals of a population of AN fibers for every MEP stimulus. These PIDs are
368 shown in Fig. 8 for 5 HP conditions (Fig. 8, red) and 5 LP conditions (Fig. 8, blue). Arrowheads
369 identify local maxima. We analyzed the PID in two ways. The first analysis simply takes the
370 local maximum closest to 0 ms (ignoring the main peak at 0 ms). This is usually also the largest
371 local maximum (except at 250 Hz) and is therefore largely in line with the procedure of Cariani
372 and Delgutte (Cariani and Delgutte, 1996b, 1996a). The intervals at which these local maxima
373 occur are systematically higher than the period corresponding to the edge frequency for LP
374 noise, and lower than that period for HP noise. For reference, the period of the edge frequency is
375 illustrated with a vertical dashed line as well as with a cosine at the edge frequency. The latter is
376 a simple prediction of the autocorrelation function of the response to a tone of that frequency
377 (Louage et al., 2004). Fig. 10A summarizes the pitch estimates using this simple metric of the
378 local maximum at the shortest delay: values for LP and HP noise straddle the diagonal but
379 deviate from equality. Fig. 10B zooms in on these deviations by plotting the ratio of pitch to
380 edge frequency for the same data, and includes a comparison with a summary of psychophysical
381 results (Hartmann et al., 2019) shown in light shaded red and blue symbols and lines. Behavioral

382 pitch estimates also deviate systematically from exact edge frequency and tend to be somewhat
383 lower (for LP stimuli) or higher (for HP stimuli) than the stimulus edge frequency, with the
384 deviation being more pronounced with decreasing edge frequency. It is evident that, while the
385 physiological pitch estimates deviate from the edge frequency in the same direction as the
386 psychophysical data, the smallest predominant population interspike interval as a predictor tends
387 to overestimate that deviation.

388 The second analysis to estimate pitch frequency is based on the entire PID rather than on one
389 local maximum (Tramo et al., 2001; Cariani et al., 2015; Hartmann et al., 2019). The underlying
390 assumption is that the CNS fits a harmonic template to the interspike interval distribution and
391 assigns a pitch corresponding to the best-fitting template. As described in the Methods, this
392 procedure consists of correlating the PID with pulse trains of different frequency. Figure 9 shows
393 the correlation coefficient or pitch salience between the PIDs and pulse trains for the 10 MEP
394 stimuli used. The frequency resulting in the largest correlation is marked by an X in Figures 9
395 and 10. The largest values are mostly found near the edge frequency (indicated by a vertical
396 dashed line in Fig. 9): for LP noise they are slightly lower than the edge frequency while for HP
397 noise they are slightly higher. As expected, local maxima also occur near subharmonics of the
398 edge frequencies, since pulse trains at these frequencies also align with maxima in the PID. For
399 LP stimuli, local maxima near the edge frequency and its subharmonics stand out quite clearly
400 except for the LP 2 kHz stimulus, while for HP stimuli local maxima are quite clear except for
401 the HP 250 Hz stimulus. Thus, for low edge frequencies the physiological estimate of pitch
402 salience is higher for LP than for HP stimuli, while the opposite is the case at high edge

403 frequencies. A similar observation is made psychophysically (Small and Daniloff, 1967;
404 Hartmann et al., 2019).

405 The pitch salience profiles for LP 0.25, HP 0.25, and LP 2 kHz conditions show maxima at
406 frequencies much lower than the edge frequency, even though a pattern of peaks can be
407 discerned consistent with frequencies near the edge frequency and its subharmonics. Closer
408 examination of the baselines in Fig. 9 shows that the pitch salience shows a negative drift with
409 increasing pulse frequency, which causes maximal salience values at frequencies much lower
410 than the edge frequency for LP 0.25, HP 0.25, and LP 2 kHz conditions. The negative drift can
411 be removed by using alternative pitch salience metrics, but for simplicity we took the shortcut of
412 defining the ‘best estimated pitch frequency’ as the local maximum closest to the edge
413 frequency, shown by circles in Figure 9. Figure 10C (circle and X symbols) shows the two
414 estimates based on the pitch salience profiles. Except for the outlying values mentioned above,
415 most estimates are close to the edge frequency but again slightly displaced from the diagonal of
416 equality. The results are replotted in panel D with the averaged psychophysical results in lighter
417 color and omitting the outliers for LP250, HP250, and LP2k. The pitch estimates are shifted in
418 the same direction as for the behavioral data and are quantitatively quite similar, in the range of
419 2-5% and with a deviation that decreases with increasing edge frequency.

420 Hartmann et al. (2019) suggested that for low edge frequencies examination of the PID over
421 a wider time window is needed to obtain pitch estimates close to the edge frequency. We tested
422 three time delay windows: ± 15 , 30 and 60 ms, and repeated the same procedure to estimate
423 pitch and found that window size had only minor effects on the estimates obtained (not shown).

424 *Fibers contributing to the fine temporal structure*

425 The pitch frequency estimated from the PID shows a reasonable consistency with psychoacoustic
426 results (Figure 10C,D). However, since it is a population sum, the PID does not give insight into
427 which fibers contribute most strongly to its maxima. We therefore repeated the same analysis
428 with harmonic lag pulse trains applied to SACs of individual fibers rather than to the PID (cf.
429 Figs. 1C). For each fiber, we obtained a pitch salience profile. These profiles, aligned to CF, are
430 shown as light grey traces in Fig. 11. For each fiber, the maximal correlation value (pitch
431 salience) is determined and indicated by a circle superimposed on the profile: frequency and
432 pitch salience are indicated by the position of the circle and its grey scale, respectively.

433 A first glance at the population figures for responses to LP noise (Fig. 11, top row) shows
434 “broken stick” patterns, similar to what we observed at the single fiber level (cf. Fig. 4C). Again,
435 two dashed lines indicate the CF of the fibers (red) and the edge frequency (black), but note that
436 here the meaning of the two lines is reversed relative to the single fiber data (Fig. 4) because here
437 CF varies rather than edge frequency. For LP noise, the fibers with CF lower than the edge
438 frequency (traces below intersection of red and black dashed lines) show pitch estimates close to
439 their own CF which therefore cluster along the red dashed line. This is to be expected (for CFs in
440 the phase-locking range) because the stimulus spectrally covers the tuning curve, and is
441 consistent with the single fiber response (vertical alignment in Fig. 4C). When fiber CF exceeds
442 the edge frequency (traces above the intersection of red and black dashed lines), the pitch
443 estimates cluster slightly below the edge frequency (vertical black dashed line). Here, the LP
444 stimulus covers the tail-end of the tuning curve, which results in a strikingly broad range of CFs
445 for which the dominant periodicity is near the edge frequency. Note that the range of CFs over
446 which such clustering occurs diminishes as edge frequency increases from 0.5 to 2 kHz: this may

447 explain why psychophysically the salience of edge pitch of LP MEP stimuli decreases with
448 increasing edge frequency. Other features of note are that maxima can be observed near
449 subharmonics of the edge frequency (e.g. vertical stacks of symbols near 0.5 kHz for the LP
450 1kHz stimulus), and that the range of contributing fibers is somewhat reduced for the LP 0.25
451 kHz stimulus, due to the higher thresholds for such low frequencies.

452 For HP noise stimuli, the broken stick pattern is less obvious. Here, for AN fibers with CF
453 above the intersection of the two dashed lines, the tuning curve is largely spectrally covered by
454 the stimulus and shows a response similar to that to BB noise, resulting in a dominant periodicity
455 near the fiber's CF so that the pitch estimates line up along the red dashed line indicating CF. For
456 CFs near and to some extent below the point of intersection, pitch estimates of high salience
457 (symbols with dark shading) are apparent somewhat above the edge frequency. However, the
458 range of CFs over which this occurs is more limited than for the corresponding panels to LP
459 stimuli, particularly for edge frequencies of 0.25 and 0.5 kHz. We suspect that this range
460 difference is due to asymmetric frequency tuning (Rose et al., 1971; Kiang and Moxon, 1974).
461 Fibers tuned to frequencies of a few kHz respond to LP noise by virtue of the "tail" of their
462 tuning curve and can contribute firing periodicities near the edge frequency. However, an
463 analogous high-frequency, low-threshold tail is not observed in fibers tuned to very low
464 frequencies.

465 *Correlates of edge-frequency in population slow temporal structure*

466 Although it involves a lot of assumptions, some of which we deem physiologically implausible
467 (see Discussion), the fine-structure-based autocorrelation representation analyzed in the
468 preceding section produces pitch estimates that are close to behavioral observations. However,
469 this representation breaks down for edge frequencies > 2 kHz (Figs. 3,5), and is therefore

470 inadequate to explain edge pitches perceived at higher edge frequencies. Variability of pitch
471 matches within and between subjects increases with increasing edge frequency (Hartmann et al.,
472 2019), but still the fact that the percept persists into frequency regions where fine-structure is
473 unlikely to be a cue, indicates that other cues must be available. We examined slower temporal
474 features by looking at fluctuation strength, i.e. the product of SAC width and height. Fig. 12
475 shows fluctuation strength as a function of CF for all fibers in response to MEP stimuli, with
476 separate symbols for high-SR (red, +) and low-SR (blue, inverted triangle) fibers. The solid line
477 shows a smoothed trend (Matlab, LOESS). A coarse overall tendency towards an inverse
478 relationship between fluctuation strength and CF is present in most panels, as was the case in
479 response to broadband noise (Fig. 3D). Superimposed on this pattern, there is a tendency towards
480 higher fluctuation strengths at CFs near the edge frequency, reflected in a local maximum of the
481 trendline. This is most clearly the case for the population responses to HP MEP stimuli (0.5, 1,
482 and 1.5 kHz), but less prominent for the corresponding LP stimuli, except at 0.5 kHz. For the
483 highest and lowest edge frequencies (0.25 and 2 kHz), the data are too sparse to hint at a local
484 structural pattern. Note that fluctuation strength tends to be somewhat higher in high-SR than in
485 low-SR fibers, which reflects the higher spike rates and hence higher rates of coincidence,
486 obtained in high-SR neurons.

487 *Correlates of edge-frequency in population rate profiles*

488 As mentioned in the Introduction, it has been hypothesized that average firing rate provides the
489 cue for edge pitch through lateral inhibition. Such a process may occur in the CNS, but could be
490 initiated already in the cochlea by the phenomenologically similar process of lateral suppression.
491 Our study was not designed to specifically test this hypothesis, but we can nevertheless examine

492 this cue in the data available. Because there are large differences in spontaneous and maximal
493 firing rate between fibers, we calculated the driven firing rate (firing rate – SR) to the MEP
494 stimulus and normalized it relative to the driven rate of the same fiber to broadband noise (100 –
495 4000 Hz) at an overall level of 70 dB SPL. Note that this stimulus has a lower spectral level than
496 the MEP stimuli, for which we also used an overall level of 70 dB SPL (see legend Fig. 13 for
497 spectral levels). The ratio (firing rate to MEP – SR)/(firing rate to broadband noise – SR) is
498 shown for a population of AN fibers for the 10 MEP stimuli in Figure 13 with symbols
499 differentiating the SR-classes. Responses to broadband noise were not always available, so the
500 number of datapoints is more restricted than in the temporal analyses show in previous figures.

501 If firing rate is only determined by energy of the noise band falling in a neuron’s bandpass
502 filter and lateral suppression is not relevant, we expect rate profiles that mimic the stimulus
503 spectrum: the values should cluster near 1 for CFs within the noise band and decrease for CFs
504 outside that band, with a region of transition when CF approaches the edge frequency (Fig. 13,
505 vertical dashed lines). In contrast, if two-tone suppression affects response rate, we expect a
506 “hump” in firing rate near the edge frequency but with higher frequency for HP noise and lower
507 frequency for LP noise).

508 For LP edge frequencies of 0.25 and 2 kHz, the data are too limited to draw conclusions, but
509 in response to edge frequencies of 0.5, 1, and 1.5 kHz, the rate profile indeed shows a plateau of
510 high values for CFs in the passband with a decrease for CFs above the edge frequency (Fig. 13,
511 upper row). However, the data are inconclusive regarding the presence of a “hump” for CFs just
512 below the corner frequency. Surprisingly, the pattern is less clear for the responses to HP noise,
513 particularly for edge frequencies of 0.25 – 1 kHz, where the ratio of driven rates shows little
514 decrease for fibers with CF below the edge frequency.

515 Note that, for LP stimuli, the “plateau” of normalized driven rates for fibers with CF below
516 the edge frequency is not at but somewhat above a ratio of 1. This is due to the higher spectral
517 level of the MEP stimuli relative to the broadband noise: the difference in spectral level
518 decreases with increasing edge frequency (e.g. 48 vs. 34 dB/Hz for LP 0.25 kHz vs broadband
519 noise, but only 37 vs. 34 dB/Hz for LP 2 kHz vs. broadband noise). For the HP stimuli, the
520 differences in spectral level re. the broadband condition are small: maximally 3 dB, cf. legend
521 Fig. 13), and the datapoints indeed cluster around 1 for CFs above the edge frequency.

522 Based on previous studies (see Discussion) low-SR fibers may be expected to show a clearer
523 “hump” profile than high-SR fibers. It is interesting to observe that low-SR fibers often provided
524 the highest ratios near the edge frequency with a clear drop above or below that frequency (for
525 LP and HP stimuli, respectively). However, many more data would be needed, preferably within
526 an animal, to convincingly demonstrate a systematic pattern near the edge frequency.

527 **Discussion**

528 We recorded responses from nerve fibers to broadband and band-limited noise, to identify
529 possible cues underlying MEP.

530 *Average rate as a pitch cue*

531 In analogy to visual and somatosensory systems, lateral inhibition has been hypothesized (von
532 Békésy, 1963; Hartmann et al., 2019) to cause a response peak in tonotopically-arrayed neural
533 populations near the edge frequency but slightly displaced into the frequency band. Lateral
534 inhibition is not present in the AN but there is a cochlear mechanical source of lateral interaction
535 so that frequency components above or below a neuron’s CF cause suppression of neural activity
536 (Katsuki et al., 1959; Sachs and Kiang, 1968; Robles and Ruggero, 2001). This may cause a rate

537 increase in neurons tuned near the edge frequency (Ruggero, 1973; Schalk and Sachs, 1980). Our
538 population data (Fig. 13) do not strongly support or refute such effects for MEP stimuli.
539 Neither did we see clear effects of lateral suppression on average rate in responses of single
540 neurons to variations of edge frequency with constant spectral level (Figs. 4,5). However, our
541 data to this manipulation are from high-SR fibers, while such effects are reportedly more
542 prominent in low-SR fibers (Schalk and Sachs, 1980). Indeed, there is a hint in the population
543 data (Fig. 13) that rate effects may be more pronounced in low-SR fibers. Also, the cochlear base
544 is more linear than its apex and our recordings were biased towards low CFs, so it is perhaps
545 not surprising that we did not obtain stronger evidence for effects of lateral suppression. Of
546 course, an absence of effects of lateral suppression at the level of the AN does not preclude that
547 lateral inhibition at a subsequent level in the CNS could underly edge pitch. Lateral inhibition
548 provides a simple, mechanistically plausible, tonotopically-based mechanism for edge pitch.
549 Arguing against its role is the observation that MEP decreases in salience above a few kHz,
550 while lateral inhibition is particularly well-documented at higher CFs.

551 *Temporal fine-structure as a pitch cue*

552 Temporal fine-structure has long been postulated to be the basis for pitch (Evans, 1978).
553 Previous AN recordings documented changes in temporal fine-structure when noise stimuli are
554 spectrally restricted to cover only part of a neuron's receptive field (Ruggero, 1973).
555 Furthermore, these changes in temporal fine-structure very systematically affect average firing
556 rate of binaural neurons in the midbrain (Chan et al., 1987).

557 As outlined by Hartmann et al. (2019), edge pitch is an interesting test case because of the
558 unusual aperiodic autocorrelation function of HP and LP MEP stimuli, with peaks systematically
559 displaced from integer multiples of the edge period (one-quarter shift mismatch, Appendix B in

560 Hartmann et al. 2019). These authors took a simple, stimulus-based approach where they used a
561 sinc function ($\sin x/x$) as the basis for temporally-based pitch predictions. We obtained such a
562 summary experimentally by summing autocorrelograms (SACs) of individual nerve fibers (the
563 PID, e.g. Figs. 7,8) to a given MEP stimulus. The first local maximum of the PID was indeed at a
564 time delay that systematically differed from the period of the edge frequency: higher for LP
565 noise and lower HP noise (Fig. 8). Following Cariani and Delgutte (1996a,b), we used the lag at
566 that maximum to estimate pitch frequency: for MEP stimuli, this lag severely under- or
567 overestimates pitch estimates obtained behaviorally (for LP and HP noise, respectively) (Fig.
568 10A,B). As pointed out by Hartmann et al. (2019), inclusion of maxima at longer lags
569 necessarily brings the pitch estimate closer to the edge frequency, because the deviation between
570 such lags and subharmonics of the edge frequency becomes proportionally smaller. We
571 incorporated lags larger than the edge period by sampling the SAC or PID with harmonic lag
572 pulse trains (Fig. 1) (Tramo et al., 2001; Bidelman and Heinz, 2011; Cariani et al., 2015). For
573 edge frequencies up to 2 kHz, this procedure resulted in pitch estimates close to behavioral
574 values (Fig. 10C,D).

575 Hartmann et al. (2019) varied the time window over which lags of the stimulus
576 autocorrelation were considered (15, 30, and 60 ms) and found that the optimum window
577 increased for decreasing edge frequency. We used a ± 15 ms window but also tested longer
578 windows (± 30 and ± 60 ms). We also tested various additional processing steps: the use of
579 unnormalized SACs (Louage et al., 2004), the weighing of different CF regions (in octave bands)
580 to offset unevenness in sampling (see e.g. Fig. 7), and a reweighing of CF (octave bands)
581 towards the presumed human CF distribution (Cariani and Delgutte, 1996a). These steps resulted

582 in minor quantitative changes, but none of them fundamentally changed the results and they are
583 therefore not illustrated.

584 Pitch predicted from the pattern of all-order spike intervals shows several parallels with
585 behavior. Both the sign and degree of mismatch (of pitch re. edge frequency) are quantitatively
586 close to averaged psychophysical results (Fig. 10C, D). Also, pitch salience is higher for LP than
587 HP stimuli at low edge frequencies but *vice versa* at high frequencies (Fig. 9), as it is
588 behaviorally (Small and Daniloff, 1967; Hartmann et al., 2019). However, proposed neural
589 substantiations of autocorrelation face several difficulties (Winter, 2005; Cedolin and Delgutte,
590 2010). The most vexing problem – the requirement for a range of substantial yet precise delays at
591 which spike trains are compared – is much worsened by requiring evaluation and weighing of
592 interspike intervals at lags corresponding to subharmonics. The existence of time delays is well-
593 documented in the binaural system but their neural substrate is highly controversial (Grothe et
594 al., 2010; Karino et al., 2011; Joris and van der Heijden, 2019) and the required magnitude of
595 delays is more than an order of magnitude smaller than for pitch. An argument at an even deeper
596 level, not so much against autocorrelation but rather against the use of fine-structure, is the
597 observation that edge pitch can be perceived at frequencies above 5 kHz. This frequency is often
598 taken as the upper limit of phase-locking to fine-structure but is already higher than the limit
599 estimated for the human auditory nerve (Verschooten et al., 2018, 2019). The observation that
600 the dominant population interspike interval fails to predict pitch (Fig. 10B) and that better
601 autocorrelation-based pitch estimates require much more elaborate and physiologically
602 implausible processing; combined with the inability of these representations to explain the
603 existence of edge pitch far above the limit of phase-locking, suggest that there must be other cues
604 to edge pitch.

605 *Slow temporal fluctuations as a pitch cue*

606 Slow temporal rate fluctuations have been suggested as important cues for complex stimuli such
607 as speech (Sinex, 2005; Carney, 2018). Such fluctuations are also present in responses to
608 stochastic stimuli: SACs of responses to broadband noise reveal both a fine-structure (at low
609 CFs) and an envelope component, which is present at all CFs and reflects a fiber's bandwidth of
610 frequency tuning (Joris, 2003; Mc Laughlin et al., 2008). An even slower change in firing is
611 expected when the noise stimulus covers only part of the receptive field. Indeed, at the level of
612 single fibers, changes in cutoff frequency are accompanied by marked changes in SAC envelope
613 at all CFs (Figs. 4,5 D,J). Our data collection did not anticipate and specifically target
614 examination of this cue, but nevertheless the population data available suffice to support its
615 possible role as a marker for edge frequency and possibly edge pitch (Fig. 12).

616 An appealing feature of the fluctuation cue is that it does not depend on the presence of fine-
617 structure, and thus could underly behavioral responses above the range of phase-locking, but that
618 it nevertheless appears to diminish for increasing edge frequency. High-CF neurons have sharper
619 tuning (Q_{10}) but wider bandwidth than low-CF fibers, and only generate slow fluctuations when
620 a small spectral sliver of the stimulus covers their tuning curve (Fig. 5). In spatial terms, this
621 means that the cochlear sector generating slow fluctuations becomes more restricted with
622 increasing edge frequency. We speculate that that is the reason why edge pitch becomes
623 increasingly weak with increasing edge frequency. So this cue is naturally more prominent at
624 low frequencies, even though it is not intrinsically dependent on the coding of fine-structure.

625 Although we quantify the fluctuation cue with the SAC, use of this slower temporal cue does
626 not hinge on an autocorrelation-type neural operation. Neural mechanisms for the extraction of
627 envelope cues are well established (review: (Joris et al., 2004)).

628 *Combination of pitch cues?*

629 A processing scheme which makes use of all 3 cues proposed, is to converge spike trains to a
630 coincidence detector which integrates over a cochlear sector of limited width. All three cues
631 potentially increase the number of coincidences across a cochlear sector near the edge frequency:
632 by increasing spike rate (lateral interactions), by imposing a common periodicity on spike trains
633 at neighbouring CFs (fine-structure), or by inducing correlated slow fluctuations in spiking
634 (envelope). The output of such a coincidence process could provide a tonotopic marker resulting
635 in the weak form of pitch generated by MEP stimuli. This scheme can be tested with a spatial
636 correlation analysis (Kovacic et al., 2010) applied to responses of populations of AN fibers.

637 **Reference**

- 638 Bidelman GM, Heinz MG (2011) Auditory-nerve responses predict pitch attributes related to
639 musical consonance-dissonance for normal and impaired hearing. *J Acoust Soc Am*
640 130:1488–1502.
- 641 Cariani P, Colburn HS, Hartmann W (2015) Temporal model of edge pitch effects. *J Acoust Soc*
642 *Am* 137:2204.
- 643 Cariani P, Delgutte B (1996a) Neural correlates of the pitch of complex tones. I. Pitch and pitch
644 salience. *J Neurophysiol* 76:1698–1716.
- 645 Cariani P, Delgutte B (1996b) Neural correlates of the pitch of complex tones. II. Pitch shift,
646 pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region
647 for pitch. *J Neurophysiol* 76:1717–1734.
- 648 Carney LH (2018) Supra-Threshold Hearing and Fluctuation Profiles: Implications for
649 Sensorineural and Hidden Hearing Loss. *J Assoc Res Otolaryngol JARO* 19:331–352.
- 650 Cedolin L, Delgutte B (2010) Spatiotemporal representation of the pitch of harmonic complex
651 tones in the auditory nerve. *J Neurosci* 30:12712–12724.
- 652 Chan JK, Yin TCT, Musicant AD (1987) Effects of Interaural Time Delays of Noise Stimuli on
653 Low-Frequency Cells in the Cat's Inferior Colliculus. II. Responses to Band-Pass Filtered
654 Noises. *J Neurophysiol* 58:543–561.
- 655 Evans EF (1978) Place and Time coding of Frequency in the Peripheral Auditory System: Some
656 Physiological Pros and Cons. *Audiology* 17:369–420.
- 657 Evans EF (1983) Pitch and cochlear nerve fibre temporal discharge patterns. In: *Hearing -*
658 *Physiological Bases and Psychophysics*, pp 140–145. Berlin: Springer-Verlag.
- 659 Geisler CD, Deng L, Greenberg SR (1985) Thresholds for primary auditory fibers using
660 statistically defined criteria. *J Acoust Soc Am* 77:1102–1109.
- 661 Grothe B, Pecka M, McAlpine D (2010) Mechanisms of sound localization in mammals. *Physiol*
662 *Rev* 90:983–1012.
- 663 Hartmann WM, Cariani PA, Colburn HS (2019) Noise edge pitch and models of pitch
664 perception. *J Acoust Soc Am* 145:1993–2008.
- 665 Javel E (1980) Coding of AM tones in the chinchilla auditory nerve: Implications for the pitch of
666 complex tones. *J Acoust Soc Am* 68:133–146.
- 667 Joris PX (2003) Interaural time sensitivity dominated by cochlea-induced envelope patterns. *J*
668 *Neurosci* 23:6345–6350.

- 669 Joris PX, Louage DH, Cardoen L, van der Heijden M (2006) Correlation Index: A new metric to
670 quantify temporal coding. *Hear Res* 216–217:19–30.
- 671 Joris PX, Louage DH, van der Heijden M (2008) Temporal damping in response to broadband
672 noise. II. Auditory nerve. *J Neurophysiol* 99:1942–1952.
- 673 Joris PX, Schreiner CE, Rees A (2004) Neural processing of amplitude-modulated sounds.
674 *Physiol Rev* 84:541–577.
- 675 Joris PX, Van De Sande B, van der Heijden M (2005) Temporal damping in response to
676 broadband noise. I. Inferior colliculus. *J Neurophysiol* 93:1857–1870.
- 677 Joris PX, van der Heijden M (2019) Early Binaural Hearing: The Comparison of Temporal
678 Differences at the Two Ears. *Annu Rev Neurosci* 42:433–457.
- 679 Kale S, Micheyl C, Heinz MG (2014) Implications of Within-Fiber Temporal Coding for
680 Perceptual Studies of F0 Discrimination and Discrimination of Harmonic and Inharmonic
681 Tone Complexes. *J Assoc Res Otolaryngol* 15:465–482.
- 682 Karino S, Smith PH, Yin TCT, Joris PX (2011) Axonal branching patterns as sources of delay in
683 the mammalian auditory brainstem: a re-examination. *J Neurosci* 31:3016–3031.
- 684 Katsuki Y, Watanabe T, Suga N (1959) Interaction of auditory neurons in response to two sound
685 stimuli in cat. *J Neurophysiol* 22:603–623.
- 686 Kiang NYS, Moxon EC (1974) Tails of tuning curves of auditory-nerve fibers. *J Acoust Soc Am*
687 55:12.
- 688 Klein MA, Hartmann WM (1981) Binaural edge pitch. *J Acoust Soc Am* 70:51–61.
- 689 Kovacic D, Michelet P, Joris PX (2010) Spatial profiles of correlation in spike timing to
690 broadband noise across auditory nerve fibers. *Assoc Res Otolaryngol Abs* 33:256–257.
- 691 Licklider JCR (1951) A duplex theory of pitch perception. *Experientia* 7:128–134.
- 692 Louage DH, van der Heijden M, Joris PX (2004) Temporal properties of responses to broadband
693 noise in the auditory nerve. *J Neurophysiol* 91:2051–2065.
- 694 Mc Laughlin M, Chabwine JN, van der Heijden M, Joris PX (2008) Comparison of bandwidths
695 in the inferior colliculus and the auditory nerve. II: Measurement using a temporally
696 manipulated stimulus. *J Neurophysiol* 100:2312–2327.
- 697 Meddis R, Hewitt MJ (1991) Virtual pitch and phase sensitivity of a computer model of the
698 auditory periphery. I: Pitch identification. *J Acoust Soc Am* 89:2866–2882.
- 699 Robles L, Ruggero MA (2001) Mechanics of the mammalian cochlea. *Physiol Rev* 81:1305–
700 1352.

- 701 Rose JE, Hind JE, Anderson DJ, Brugge JF (1971) Some Effects of Stimulus Intensity on
702 Response of Auditory Nerve Fibers in the Squirrel Monkey. *J Neurophysiol* 34:685–699.
- 703 Ruggero MA (1973) Response to Noise of Auditory Nerve Fibers in the Squirrel Monkey. *J*
704 *Neurophysiol* 36:569–587.
- 705 Sachs MB, Kiang NYS (1968) Two-Tone Inhibition in Auditory-Nerve Fibers. *J Acoust Soc Am*
706 43:1120–1128.
- 707 Schalk TB, Sachs MB (1980) Nonlinearities in auditory-nerve fiber responses to bandlimited
708 noise. *J Acoust Soc Am* 67:903–913.
- 709 Sinex DG (2005) Spectral processing and sound source determination. *Int Rev Neurobiol*
710 70:371–398.
- 711 Small AM, Daniloff RG (1967) Pitch of Noise Bands. *J Acoust Soc Am* 41:506–512.
- 712 Tramo MJ, Cariani PA, Delgutte B, Braida LD (2001) Neurobiological foundations for the
713 theory of harmony in western tonal music. *Ann N Y Acad Sci* 930:92–116.
- 714 Verschooten E, Desloovere C, Joris PX (2018) High-resolution frequency tuning but not
715 temporal coding in the human cochlea. *PLOS Biol* 16:e2005164.
- 716 Verschooten E, Shamma S, Oxenham AJ, Moore BCJ, Joris PX, Heinz MG, Plack CJ (2019)
717 The upper frequency limit for the use of phase locking to code temporal fine structure in
718 humans: A compilation of viewpoints. *Hear Res* 377:109–121.
- 719 von Békésy G (1963) Hearing Theories and Complex Sounds. *J Acoust Soc Am* 35:588–601.
- 720 Winter IM (2005) The Neurophysiology of Pitch. In: Pitch (Plack CJ, Fay RR, Oxenham AJ,
721 Popper AN, eds), pp 99–146 Springer Handbook of Auditory Research. Springer New
722 York. Available at: http://link.springer.com/chapter/10.1007/0-387-28958-5_4 [Accessed
723 January 30, 2014].

724 **Figure 1 Temporal analysis.** (A) illustrates the Shuffled AutoCorrelogram (SAC) to a single
725 stimulus, in this case broadband noise. It is normalized to the total number of intervals, where a
726 value of 1 indicates absence of correlation in spike timing between different spike trains obtained
727 to the same stimulus. Each SAC value reflects the number of coincident spikes occurring at a
728 given delay. (B) shows a 30 ms segment of harmonic lag pulse trains for 5 different frequencies
729 (bottom to top: 200, 400, 600, 800, 1000 Hz). Note that the pulse at time delay zero is
730 eliminated. (C) is the correlation profile for the SAC shown in panel A. It shows correlation
731 values between the SAC and harmonic lag pulse trains over a wide range of frequencies (100 Hz
732 to 4 kHz, in steps of 1 Hz). The blue dot highlights the largest correlation value, which yields the
733 estimated pitch frequency (here 545 Hz) and correlation salience (here 0.07). There are also
734 smaller peaks in the correlation profile at subharmonic values. Data in A and C were for an AN
735 fiber with characteristic frequency (CF) = 533 Hz and spontaneous rate (SR) of 73 spikes/s.

736

737 **Figure 2. Definition of halfwidth (HW) for analysis of fluctuations.** The brown line is the
738 SAC; the blue line is a Gaussian fit to its envelope. Blue solid circles indicate local maxima. HW
739 is defined as the width of the Gaussian at half height (horizontal black line). Two ordinates are
740 shown. The left y-axis shows the normalized number of coincidences and the right y-axis shows
741 the number of coincidences (unnormalized for firing rate).

742

743 **Figure 3. Responses to broadband noise.** (A) Distribution of CF and SR in the AN population
744 recorded. The marginal histograms have bin widths of 0.15 oct (top) and 5 spike/s (right). (B)
745 SACs of 82 AN fibers (blue lines) sorted by CF. The SACs are normalized and a value of 1 is
746 subtracted. The thick red line (top) is the PID, which is the average of all the SACs. (C, top)

747 shows the correlation profile between PID and subharmonic lag pulse trains. The grey dot
748 identifies estimated pitch frequency (x value) and its pitch salience (y-value). **(C, bottom)** Pitch
749 frequency estimated for individual AN fibers. Each symbol shows an estimate for one fiber: the
750 grey scale shows the pitch salience for values ranging from 0 (white) to 0.2 (black). Scales for
751 top and bottom abscissa are identical but shows the frequency of the subharmonic lag pulse train
752 for the top panel and estimated single-fiber pitch frequency for the bottom panel. Red diagonal
753 dashed line shows diagonal of equality between CF and estimated single-fiber pitch frequency.
754 **(D)** Fluctuation strength of AN fibers as a function of CF, measured as the product of SAC
755 halfwidth (HW in ms) and peak amplitude (coincidence rate, CR, (spikes/s)²).

756

757 **Figure 4. Response of low-CF fibers to LP and HP edge pitch stimuli.** Top row: responses to
758 LP noise for one AN fiber, bottom row: responses to HP noise for another AN fiber. The graphic
759 on the left illustrates the stimuli with one fixed and one varying edge (double-headed arrow). The
760 vertical frequency axis is aligned across all panels in a row but has multiple meanings. **(A)** Blue
761 symbols and curve show firing rate (bottom ordinate) as a function of edge frequency (abscissa).
762 Tuning curve is superimposed (black line, top ordinate.). The abscissa indicates pure tone
763 frequency for the tuning curve and varying edge frequency for firing rate. CF (horizontal dashed
764 line) was 867 Hz, SR was 84 spikes/s (asterisk). **(B)** SACs of responses to different edge
765 frequencies. Vertical dashed lines show period of CF⁻¹. **(C)** estimated pitch frequency versus
766 varying edge frequency. Greyscale (see inset for scale) shows pitch salience. Vertical dashed line
767 (black) indicates CF; diagonal dashed line (red) indicates equality between stimulus edge
768 frequency and estimated pitch. **(D)** SACs shown over wider range of delays, with a Gaussian fit
769 to their envelope. **(E)** Halfwidth (HW in ms, blue) and coincidence rate (CR in (spikes/s)², green)

770 of SACs in D. Note that the numerical abscissa values apply to both metrics. **(F-J)** Same as A-E
771 but for responses to HP noise for a fiber with CF = 1.3 kHz (SR = 25 spikes/s).

772

773 **Figure 5. Response of a high-CF fiber to LP and HP edge pitch stimuli.** Same arrangement as
774 in Fig. 4. CF was 8.6 kHz, SR was 27 spikes/s. Upper cutoff of HP noise was 12 kHz.

775

776 **Figure 6. Comparison of temporal response patterns in fibers with different CF, to**

777 **broadband noise and MEP stimuli.** In each panel, SACs of five selected AN fibers are sorted

778 by CF (y-axis). The top row shows results for LP noise with edge frequency of 0.5 and 1 kHz.

779 The lower row shows the same analysis for responses to HP noise. SACs are in blue in response

780 to MEP stimuli and are in black in response to BB stimuli. The pink background color illustrates

781 the passband of the MEP stimuli and the vertical dashed lines indicate the periodicity of the edge

782 frequency (integer multiples of 2ms and 1 ms for edge frequencies of 0.5 and 1 kHz,

783 respectively). Overall SPL of all stimuli was 70 dB. All SACs use the same X- and Y-scale.

784

785 **Figure 7. SACs and PID to MEP stimuli.** Each panel shows SACs of populations of AN fibers

786 in response to one MEP stimulus. Their average, the PID, is plotted on top in red. The top row

787 shows results for LP noise with edge frequency of 0.5, 1, 1.5, and 2 kHz, and the lower row

788 shows the same analysis for responses to HP noise. The vertical dashed lines show the period of

789 the relevant stimulus edge frequency and its integer multiples. For the PIDs to the 0.5 kHz MEP

790 stimuli, arrows point to the first local maximum straddling the central peak.

791

792 **Figure 8. PID of response to different edge frequencies.** The PIDs of HP (red) and LP (blue)
793 noise at the same edge frequency are shown for edge frequencies of 0.25, 0.5, 1, 1.5 and 2 kHz.
794 The range of delays in the x-axis differs for the different panels and is scaled based on the edge
795 frequency. The grey dashed functions are cosine waves at the edge frequency, with a vertical
796 dashed line marking the corresponding period. The triangle markers indicate local maxima.

797

798 **Figure 9. Pitch estimates based on the pulse train method.** Each panel shows pitch salience,
799 i.e. the correlation between PID and harmonic lag pulse trains, for a range of pulse train
800 frequencies. Two metrics are illustrated. The local maximum closest to the edge frequency is
801 labeled by a circle; the maximum value is labeled by the x symbol.

802

803 **Figure 10. Comparison between physiological and behavioral pitch estimates.** (A) Pitch
804 estimated by local PID maxima closest to 0 ms. (C) Pitch estimated by correlating the PIDs with
805 harmonic lag pulse trains. X symbols show the frequencies resulting in the largest correlation
806 coefficient (largest pitch salience). Full circles show the frequencies of the local maxima closest
807 to the edge frequency (often obscured by X symbols). (B) replotting of data in A as a ratio and
808 with comparison to behavioral data (light colors, taken from Hartmann et al. 2019). (D) same but
809 for data in B. The diagonal black line in A and C and the horizontal black line in B and D
810 indicate equality between edge frequency and pitch estimate.

811

812 **Figure 11. Pitch estimates for individual AN fibers.** Each panel shows a pitch estimate for
813 different fibers, ranked by CF (ordinate). The light grey traces show pitch salience as a function
814 of frequency for each fiber (aligned with the CF ordinate at 0 correlation), and the frequency of
815 maximal salience is indicated with a circle. Grey shading of the circle indicates pitch salience

816 (scale bar on the right). Top row: results for LP noise, bottom row: results for HP noise. The
817 black vertical dash lines indicate edge frequency. Equality between CF and estimated pitch
818 frequency is shown by the red dashed line. The blue traces above each panel are the salience
819 estimates obtained from the PID analysis, retaken from Fig. 9. Ordinate scale of light grey traces
820 is the same for all traces within a panel.

821

822 **Figure 12. Fluctuation strength in response to MEP stimuli for a population of AN fibers.**

823 Each fiber is represented by a single symbol, which differentiates fibers of low- and high SR
824 (inset). Fiber CF is shown in the abscissa; fluctuation strength is shown in the ordinate. The
825 black line is a trendline (Matlab, LOESS, span: 0.4) and the vertical dashed lines indicate
826 stimulus edge frequency.

827

828 **Figure 13. Population profiles of normalized driven rate vs. CF, for 10 MEP stimuli.** The
829 ordinate shows normalized driven rate, which is driven rate (firing rate – SR) to the MEP
830 stimulus divided by that to broadband noise. The vertical dashed lines indicate the stimulus edge
831 frequency; the horizontal line indicates a normalized driven rate equal to one, which means the
832 driven firing rate to the MEP stimuli equals that to broadband noise. The black lines are a
833 trendline through the data points (Matlab, LOESS function, span: 0.5). Blue inverted triangles
834 indicated data for low-SR fibers; red + symbols for high-SR fibers. All stimuli were at an overall
835 level of 70 dB SPL, resulting in the following spectral levels (in dB/Hz), broadband noise: 34,
836 LP 0.25 kHz: 48, LP 0.5 kHz: 43, LP 1 kHz: 40, LP 1.5 kHz: 38, LP 2 kHz: 37, HP 0.25 kHz:
837 34.2, HP 0.5 kHz: 34.6, HP 1 kHz: 35.2, HP 1.5 kHz: 36, HP 2 kHz: 37.

838

























