



**HAL**  
open science

# Reduction in sound discrimination in noise is related to envelope similarity and not to a decrease in envelope tracking abilities

Samira Souffi, Léo Varnet, Meryem Zaidi, Brice Bathellier, Chloé Huetz, Jean-Marc Edeline

## ► To cite this version:

Samira Souffi, Léo Varnet, Meryem Zaidi, Brice Bathellier, Chloé Huetz, et al.. Reduction in sound discrimination in noise is related to envelope similarity and not to a decrease in envelope tracking abilities. *The Journal of Physiology*, 2023, 601 (1), pp.123-149. 10.1113/JP283526 . hal-03853055

**HAL Id: hal-03853055**

**<https://hal.science/hal-03853055>**

Submitted on 15 Nov 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

1  
2 **Reduction in sound discrimination in noise is related to envelope**  
3 **similarity and not to a decrease in envelope tracking abilities**  
4  
5

6 Samira Souffi<sup>1, §</sup>, Léo Varnet<sup>2</sup>, Meryem Zaidi<sup>1</sup>, Brice Bathellier<sup>3</sup>, Chloé Huetz<sup>1</sup> and  
7 Jean-Marc Edeline\*<sup>1</sup>  
8

9 <sup>1</sup> Paris-Saclay Institute of Neuroscience (Neuro-PSI, UMR 9197), CNRS - Université Paris-Saclay, Centre CEA Saclay  
10 - Bât 151, 91400 Saclay, France.

11 <sup>2</sup> Laboratoire des systèmes perceptifs, UMR CNRS 8248, Département d'Etudes Cognitives, Ecole Normale Supérieure,  
12 Université Paris Sciences & Lettres, Paris, France.

13 <sup>3</sup> Institut de l'Audition, Institut Pasteur, Université de Paris, INSERM, Paris F-75012, France.  
14  
15  
16  
17

18 Number of pages: 33

19 Number of figures: 7

20 Number of words in abstract: 166

21 Number of words in introduction: 994

22 Number of words in discussion: 2429  
23

24 Running title: Envelope tracking of noisy sounds in the auditory system  
25  
26

27 \* *Corresponding Author:*

28 Jean-Marc Edeline

29 UMR 9197, Neuro-PSI (Institut des Neurosciences Paris-Saclay)

30 CNRS - Université Paris-Saclay,

31 Centre CEA Saclay, Bâtiment 151

32 91400 Saclay cedex, France

33 email: jean-marc.edeline@universite-paris-saclay.fr  
34

35 § current address: Edmond and Lily Safra Center for Brain Sciences, The Hebrew University of Jerusalem, Edmond J.  
36 Safra Campus, Givat Ram, Jerusalem 9190401, Israel.  
37  
38  
39  
40  
41  
42  
43  
44

45 **First author profile**  
46

47 Samira Souffi completed her PhD in France at the Institute of Neuroscience in Paris-Saclay  
48 University (NeuroPSI) under the supervision of Dr. Jean-Marc Edeline. Her research was focused  
49 on understanding the effects of background noises on the neuronal discrimination in the whole  
50 auditory system. Her postdoctoral research in Prof. Israel Nelken laboratory (ELSC, Hebrew  
51 University of Jerusalem, Israel) describes the neuronal networks involved in developing sound  
52 preferences (e.g. music preferences) on mice by recording the calcium activity in both the reward  
53 and the auditory systems.

54 **Key points:**

55

56 \* In quiet, envelope tracking in the low amplitude modulation range (<20 Hz) is correlated with the  
57 neuronal discrimination between communication sounds as quantified by mutual information from  
58 the cochlear nucleus up to the auditory cortex.

59 \* At each level of the auditory system, auditory neurons keep their abilities to track the  
60 communication sound envelopes in situations of acoustic degradation such as vocoding and the  
61 addition of masking noises up to a signal-to-noise ratio of -10 dB.

62 \* In noise, the increase in between-stimulus envelope similarity explains both the reduction in  
63 behavioral and neuronal discrimination in the auditory system.

64 \* Envelope tracking can be viewed as a universal mechanism allowing neural and behavioral  
65 discrimination as long as the temporal envelope of communication sounds display some differences.

66

67

68 **Abstract**

69

70 Humans and animals constantly face challenging acoustic environments such as various background  
71 noises that impair the detection, discrimination and identification of behaviorally relevant sounds.  
72 Here, we disentangled the role of temporal envelope tracking on the reduction in neuronal and  
73 behavioral discrimination between communication sounds in situations of acoustic degradations. By  
74 collecting neuronal activity from six different levels of the auditory system, from auditory nerve up  
75 to secondary auditory cortex in anesthetized guinea-pigs, we found that tracking of slow changes of  
76 the temporal envelope is a general functional property of auditory neurons for encoding  
77 communication sounds in quiet and also in adverse, challenging, conditions. Results from a Go/No-  
78 Go sound discrimination task on mice support the idea that the loss of distinct slow envelope cues  
79 in noisy conditions impacted the discrimination performance. Together, these results suggest that  
80 envelope tracking is potentially a universal mechanism operating in the central auditory system,  
81 which allows the detection of any between-stimuli difference in slow envelope and thus cope with  
82 degraded conditions.

83

84

85

86

87

88 **Keywords:** Auditory system; Temporal envelope tracking; Neuronal and behavioral discrimination;  
89 Degraded acoustic conditions.

90	<b>Abbreviation list</b>
91	
92	ABR: auditory brainstem responses
93	A1: primary auditory cortex
94	AM: amplitude modulation
95	CN: cochlear nucleus
96	CNIC: central nucleus of inferior colliculus
97	E: envelope
98	ERB: equivalent rectangular bandwidth
99	FM: frequency modulation
100	H: high
101	L: low
102	M: middle
103	MGv: ventral part of medial geniculate body
104	MI: mutual information
105	MUA: multiunit activity
106	PSTH: peristimulus histogram
107	R: correlation coefficient
108	$R_{\max_{E-PSTH}}$ : maximal value of correlation coefficient between envelope and PSTH
109	$R_{\text{Random}}$ : correlation coefficient between envelope and shuffled PSTH
110	RMS: root mean square
111	sANF: simulated auditory nerve fibers
112	SNR: signal-to-noise ratio
113	SR: spontaneous rate
114	TFRP: time-frequency response profile
115	VRB: ventral-rostral belt (secondary auditory cortex)
116	
117	



## 118 **Introduction**

119

120 In humans, speech signals are characterized by rhythmic streams of amplitude and frequency  
121 modulations (AM and FM) that convey phoneme, syllable, word, and phrase information (Rosen  
122 1992, Varnet et al., 2017, Ding et al., 2017). It is known for several decades that the low-frequency  
123 modulations of the temporal envelope carry essential cues for speech perception (Drullman et al.,  
124 1994a,b; Shannon et al., 1995; Zeng et al., 2005): even in challenging conditions (including in  
125 various types of noise), the human auditory system has the capacity to process highly degraded  
126 speech as long as the temporal envelope modulations below 20 Hz are preserved (Drullman et al.,  
127 1994a, b; Shannon et al., 1995; Zeng et al., 2005). This is consistent with electroencephalographic  
128 (EEG) and magnetoencephalographic (MEG) studies in which cortical responses were found in  
129 phase with the temporal envelope of speech signals and strongly correlated with the average level of  
130 speech comprehension both for normal and compressed speech (Ahissar et al., 2001; Luo and  
131 Poeppel, 2007; Ding et al., 2014). In a recent study, Ortiz-Barajas and colleagues (2021) have found  
132 that newborns possess the neural capacity to track the amplitude and the phase of the speech  
133 envelope in their native language (French), as well as in rhythmically similar and different  
134 unfamiliar languages (Spanish and English). These results support the hypothesis that speech  
135 envelope tracking may be a necessary prerequisite, although not sufficient, for speech  
136 comprehension (Köseme et al., 2016, 2017).

137 In animals, the synchronization of auditory cortex responses with the temporal envelope of guinea  
138 pig vocalizations has been observed in several studies (Wallace et al., 2005; Wallace & Palmer,  
139 2009; Grimsley et al., 2011, 2012), some of them even suggesting that cortical responses could be  
140 isomorphic to the vocalization envelope (Figure 2A in Grimsley et al., 2012). Using speech stimuli  
141 with different levels of degradation (clear, conversational and compressed), Abrams and colleagues  
142 (2017) recorded responses of auditory cortex neurons in guinea pigs and showed that populations of  
143 cortical neurons encode both the periodicity and the temporal broadband envelope of the speech  
144 signal. These temporal representations in auditory cortex were quite resistant to the degradations  
145 (conversational and compressed speech) and additional studies have pointed out that cortical  
146 neurons can still respond to target stimuli in important levels of noise (between -5 and 0 dB SNR,  
147 Nagarayan et al., 2002; Narayan et al., 2007; Shetake et al., 2011; Homma et al., 2020). At the  
148 subcortical level, several studies revealed in both mammals and birds that the average responses of  
149 inferior colliculus neurons can reflect the communication sound envelope (Suta et al., 2003;  
150 Woolley et al., 2006; Rode et al., 2013).

151 Here, we used acoustic degradations that differentially affected the similarities between acoustic  
152 envelopes: vocoders strongly altered the spectral cues but preserved most of the temporal

153 information whereas noise addition produced spectrotemporal degradations, reduced the temporal  
154 cues while introducing irrelevant envelope fluctuations and altered the spectral cues (Fig. 1 of  
155 Souffi et al., 2020). We used a stationary noise which strongly increased the acoustic similarity  
156 between the envelopes and a chorus noise which differed between the four envelopes and therefore  
157 masked the vocalizations while not inducing an increase in the overall similarity of the stimuli. We  
158 previously showed that the addition of a stationary noise strongly impaired the neuronal  
159 discrimination performance at the subcortical and cortical levels, whereas it was less impaired in the  
160 vocoding conditions (Fig. 6-9 of Souffi et al., 2020). To go further, our main goal in the current  
161 study was to determine whether the similarities between acoustic envelopes or the loss in envelope  
162 tracking ability by auditory neurons reduce or even prevent the neuronal and behavioral  
163 discrimination in situations of acoustic degradations. In a condition-independent scenario, the  
164 neurons keep the same intrinsic ability to track the stimulus envelopes whatever the acoustic  
165 conditions (in quiet and in degraded conditions): As long as the stimulus envelopes differ, the  
166 neurons will discriminate the stimuli. In contrast, in a condition-dependent scenario, the acoustic  
167 degradations reduce the neurons' ability to track the stimulus envelopes. This deleterious effect can  
168 potentially occur when the neurons are strongly driven by the acoustic degradations (such as noise  
169 addition) leading to limited dynamic ranges for coding the target stimuli. This occurs for example  
170 for the responses of auditory nerve fibers (ANF) to tones in continuous noise. Even if the responses  
171 to 120-300Hz periodic AM stimuli were still preserved at 0 and +6 dB SNR (Figure 6 of Frisina et  
172 al., 1996), many studies reported that the rate-level functions of ANF tested with pure tones were  
173 largely altered in noise: in many cases the responses did not reach the same saturation level (Rhode  
174 et al., 1978; Geisler and Sinex, 1980; Costalupes et al., 1984; Frisina et al., 1996) or the whole  
175 curve was shifted toward the right (Rhode et al., 1978; Costalupes et al., 1984) indicating that the  
176 thresholds were higher and the dynamic ranges were smaller than in quiet. This was also a function  
177 of the bandwidth of the noise and the types of ANF (i.e., the effects differed between low, medium  
178 or high spontaneous rate fibers, see for example Reiss et al., 2011). Based on these studies, it seems  
179 that, as early as the auditory nerve, the detection of AM cues contained in target stimuli, and  
180 therefore the tracking abilities of central auditory neurons, can be reduced. Similar results have  
181 been observed in the inferior colliculus (Ramachandran et al., 2000).

182 In an attempt to dissociate between these two scenarios, we evaluated the relationship between the  
183 envelope tracking of sounds and the neuronal discrimination in the entire auditory system. We  
184 simulated auditory nerve fiber (sANF) responses (with a widely-used model, Bruce et al., 2018) and  
185 recorded the neuronal activity in five auditory structures in response to four conspecific  
186 vocalizations presented in quiet, using three tone-vocoders and two types of noise (a stationary and  
187 a chorus noise at three SNRs +10, 0 and -10 dB) in anesthetized guinea pigs. We found that

188 subcortical and cortical neurons track the envelopes in the low AM range ( $<20\text{Hz}$ ), with a high  
189 degree of fidelity in original and degraded conditions, suggesting that the auditory system maintains  
190 a robust temporal representation from the auditory nerve to the auditory cortex. Behaving mice  
191 were also able to discriminate between these communication sounds, and performed the task above  
192 chance level in all noisy conditions. Overall, our results demonstrate that the between-stimulus  
193 envelope similarity, which increases in noise, negatively correlates both with the neuronal  
194 discrimination and the behavioral performance.

## 195 **Materials and Methods**

196 Most of the Methods are similar to those described in Souffi and colleagues (2020).

197

### 198 **Subjects for the electrophysiological and behavioral experiments**

199 These experiments were performed under the national license A-91-557 (project 2014-25,  
200 authorization 05202.02) and using the procedures N° 32-2011 and 34-2012 validated by the Ethic  
201 committee N°59 (CEEA, (Comité d’Ethique pour l’Expérimentation Animale) Paris Centre et Sud).  
202 All surgical procedures were performed in accordance with the guidelines established by the  
203 European Communities Council Directive (2010/63/EU Council Directive Decree).

204 Extracellular recordings were obtained from 47 adult pigmented guinea pigs (aged 3 to 16 months  
205 old, 36 males, 11 females) at five different levels of the auditory system: the cochlear nucleus (CN),  
206 the inferior colliculus (IC), the medial geniculate body (MGB), the primary (A1) and secondary  
207 auditory cortex (area VRB). Animals, weighing from 515 to 1100 g (mean 856 g), came from our  
208 own colony housed in a humidity (50-55%) and temperature (22-24°C)-controlled facility on a 12  
209 h/12 h light/dark cycle (light on at 7:30 A.M.) with free access to food and water.

210 Two days before the electrophysiological experiment the animal’s pure-tone audiogram was  
211 determined by testing auditory brainstem responses (ABR) under isoflurane anesthesia (2.5%) as  
212 described in Gourévitch and colleagues (2009). A software (RTLlab, Echodia, Clermont-Ferrand,  
213 France) allowed averaging 500 responses during the presentation of each pure-tone frequency and  
214 each intensity (between 0.5 and 32 kHz, duration: 10 ms, rise-fall time: 2 ms) delivered by a  
215 speaker (Knowles Electronics) placed in the animal’s right ear canal. The threshold of each ABR  
216 was defined as the lowest intensity where a small ABR wave could still be detected (usually wave  
217 III). For each frequency, the threshold was determined by gradually decreasing the sound intensity  
218 (from 80 dB down to -10 dB SPL). There was a perfect agreement between the thresholds visually  
219 determined by two co-authors (SS, JME). Based upon a large database of more than 250 guinea  
220 pigs, we considered that all animals used in this study had normal pure-tone audiograms  
221 (Gourévitch et al., 2009; Gourévitch and Edeline, 2011).

222 Behavioral experiments were performed on nine eight-weeks old C57Bl/6J female mice (see  
223 *Behavioral Go/No-Go discrimination task* part for more details).

224

### 225 **Acoustic stimuli**

226 The acoustic stimuli were the same as in Souffi and colleagues (2020, 2021). They were generated  
227 using MatLab, transferred to a RP2.1-based sound delivery system (TDT) and sent to a Fostex  
228 speaker (FE87E). The speaker was placed at 2 cm from the guinea pig’s right ear, a distance at

229 which the speaker produced a flat spectrum ( $\pm 3$  dB) between 140 Hz and 36 kHz. Calibration of  
230 the speaker was made using noise and pure tones recorded by a Bruel and Kjaer microphone 4133  
231 coupled to a preamplifier BandK 2169 and a digital recorder Marantz PMD671.

232 Time-Frequency Response Profiles (TFRP) were determined using 129 pure-tone frequencies  
233 covering eight octaves (0.14-36 kHz) and presented at 75 dB SPL. The tones had a gamma  
234 envelope given by  $\gamma(t) = (\frac{t}{4})^2 e^{-\frac{t}{4}}$ , where  $t$  is time in ms. At a given stimulus level, each frequency  
235 was repeated eight times at a rate of 2.35 Hz in pseudorandom order. The duration of these tones  
236 over half-peak amplitude was 13.6 ms and at 50ms the sound intensity was 6.7 dB SPL. There was  
237 no overlap between tones.

238 A set of four conspecific vocalizations was used to assess the neuronal responses to communication  
239 sounds. These vocalizations were recorded from animals of our colony. Pairs of animals were  
240 placed in the acoustic chamber and their vocalizations were recorded by a Bruel & Kjaer  
241 microphone 4133 coupled to a preamplifier B&K 2169 and a digital recorder Marantz PMD671. A  
242 large set of whistle calls was loaded in the Audition software (Adobe Audition 3) and four  
243 representative examples of whistles were selected (Figure 1A, left panel). As shown in Figure 1B  
244 (left panel), their overall envelopes clearly differed, W2 and W4 envelopes being the closest from  
245 each other. The four whistles were presented in two frozen noises ranging from 10 to 24 000 Hz. To  
246 generate these noises, audio-recordings were performed in the colony room where a large group of  
247 guinea pigs were housed (30-40 animals; 2-4 animals/cage). Several 4-seconds of audio recordings  
248 were added up to generate the "chorus noise", whose power spectrum was computed using the  
249 Fourier transform. The chorus noise masking each target vocalization was slightly different in terms  
250 of spectro-temporal content. The chorus noise spectrum was then used to shape the spectrum of a  
251 Gaussian white noise. The resulting "vocalization-shaped stationary noise" therefore matched the  
252 "chorus-noise" audio spectrum. Figure 1B displays the overall envelopes of the four whistles in the  
253 vocalization-shaped stationary noise (third panel) and in the chorus noise (fourth panel) with  
254 signal-to-noise ratios (SNR) of +10, 0 and -10 dB.

255 The four selected whistles were also processed by three tone vocoders (Gnansia et al., 2009, 2010).  
256 In the following figures, the unprocessed whistles will be referred to as the original versions, and  
257 the vocoded versions as Voc38, Voc 20, and Voc10 using 38, 20, and 10 bands, respectively. In  
258 contrast to previous studies that used noise-excited vocoders (Nagarajan et al., 2002; Ranasinghe et  
259 al., 2012; Ter-Mikaelian et al., 2013), a tone vocoder was used here, because noise vocoders were  
260 found to introduce random (i.e., non-informative) intrinsic temporal-envelope fluctuations  
261 distorting the crucial spectro-temporal modulation features of communication sounds (Shamma and  
262 Lorenzi, 2013; Kates, 2011; Stone et al., 2011). Figure 1B displays the overall envelopes of the 38-  
263 band vocoded (first row, second panel), the 20-band vocoded (second row, second panel) and the

264 10-band vocoded (third row, second panel) versions of the four whistles. The three vocoders  
265 differed only in terms of the number of frequency bands (i.e., analysis filters) used to decompose  
266 the whistles (38, 20 or 10 bands). The 38-band vocoding process is briefly described below, but the  
267 same principles apply to the 20-band or the 10-band vocoders. Each digitized signal was passed  
268 through a bank of 38 fourth-order Gammatone filters (Patterson, 1987) with center frequencies  
269 uniformly spaced along a guinea-pig adapted ERB (Equivalent Rectangular Bandwidth) scale  
270 ranging from 50 to 35505 Hz (Sayles and Winter, 2010).

#### 271 *Overall envelope extraction*

272 In each frequency band, the temporal envelope was extracted using full-wave rectification and low-  
273 pass filtering at 64 Hz with a zero-phase, sixth-order Butterworth filter. The resulting envelopes  
274 were used to amplitude modulate sine-wave carriers with frequencies at the center frequency of the  
275 Gammatone filters, and with random starting phase. Impulse responses were peak-aligned for the  
276 envelope (using a group delay of 16 ms) and the acoustic temporal fine structure across frequency  
277 channels (Hohmann, 2002). The modulated signals were finally weighted and summed over the 35  
278 frequency bands (see section "*Quantification of the envelope tracking*"). The weighting  
279 compensated for imperfect superposition of the bands' impulse responses at the desired group  
280 delay. The weights were optimized numerically to achieve a flat frequency response.

281

#### 282 **Surgical procedures**

283 All guinea pigs were anesthetized by an initial injection of urethane (1.2 g/kg, i.p.) supplemented by  
284 additional doses of urethane (0.5 g/kg, i.p.) when reflex movements were observed after pinching  
285 the hind paw (usually 2-4 times during the recording session). A single dose of atropine sulfate  
286 (0.06mg/kg, s.c.) was given to reduce bronchial secretions and a small dose of buprenorphine was  
287 administered (0.05mg/kg, s.c.) as urethane has no analgesic properties. After placing the animal in a  
288 stereotaxic frame, a craniotomy was performed and a local anesthetic (Xylocain 2%) was injected in  
289 the wound.

290 For auditory cortex recordings (area A1 and VRB), a craniotomy was performed above the left  
291 temporal cortex. The dura above the auditory cortex was removed under binocular control and the  
292 cerebrospinal fluid was drained through the cisterna to prevent the occurrence of oedema. For the  
293 recordings in MGB, a craniotomy was performed above the most posterior part of the MGB (8mm  
294 posterior to Bregma) to reach the left auditory thalamus at a location where the MGB is mainly  
295 composed of its ventral, tonotopic, part (Redies et al., 1989; Edeline et al., 1999; Anderson et al.,  
296 2007; Wallace et al., 2007). For IC recordings, a craniotomy was performed above the IC and  
297 portions of the cortex were aspirated to expose the surface of the left IC (Malmierca et al., 1995,  
298 1996; Rees et al., 1997). For CN recordings, after opening the skull above the right cerebellum,

299 portions of the cerebellum were aspirated to expose the surface of the right CN (Paraouty et al.,  
300 2018).

301 After all surgeries, a pedestal in dental acrylic cement was built to allow an atraumatic fixation of  
302 the animal's head during the recording session. The stereotaxic frame supporting the animal was  
303 placed in a sound-attenuating chamber (IAC, model AC1). At the end of the recording session, a  
304 lethal dose of Exagon (pentobarbital >200 mg/kg, i.p.) was administered to the animal.

305

### 306 **Recording procedures**

307 Data from multi-unit recordings were collected in 5 auditory structures, the non-primary cortical  
308 area VRB, the primary cortical area A1, the medial geniculate body (MGB), the inferior colliculus  
309 (IC) and the cochlear nucleus (CN). In a given guinea pig, neuronal recordings were only collected  
310 in one auditory structure.

311 Cortical extracellular recordings were obtained from arrays of 16 tungsten electrodes (TDT,  
312 TuckerDavis Technologies;  $\varnothing$ : 33  $\mu\text{m}$ , <1 M $\Omega$ ) composed of two rows of 8 electrodes separated by  
313 1000  $\mu\text{m}$  (350  $\mu\text{m}$  between electrodes of the same row). A silver wire, used as ground, was inserted  
314 between the temporal bone and the dura mater on the contralateral side. The location of the primary  
315 auditory cortex was estimated based on the pattern of vasculature observed in previous studies  
316 (Wallace et al., 2000; Gaucher et al., 2013, 2020; Gaucher and Edeline, 2015). The non-primary  
317 cortical area VRB was located ventral to A1 and distinguished by its longer response latencies to  
318 pure tones (Rutkowski et al., 2002; Grimsley et al., 2012). For each experiment, the position of the  
319 electrode array was set in such a way that the two rows of eight electrodes sampled neurons  
320 responding from low to high frequency when progressing in the rostro-caudal direction [see  
321 examples in Figure 1 of Gaucher et al., (2012) and in Figure 6A of Occelli et al., (2016)].

322 In the MGB, IC and CN, the recordings were obtained using 16-channel multi-electrode arrays  
323 (NeuroNexus) composed of one shank (10 mm) of 16 electrodes spaced by 110  $\mu\text{m}$  and with  
324 conductive site areas of 177 $\mu\text{m}^2$ . The electrodes were advanced vertically (for MGB and IC) or with  
325 a 40° angle to the CN surface until responses to pure tones could be detected on at least 10  
326 electrodes.

327 All thalamic recordings were from the ventral part of MGB (see above surgical procedures) and all  
328 displayed response latencies < 9ms. At the collicular level, we distinguished the lemniscal and non-  
329 lemniscal divisions of IC based on depth and the latencies of pure tone responses. We excluded the  
330 most superficial recordings (until a depth of 1500 $\mu\text{m}$ ) and those exhibiting latencies  $\geq$  20ms in an  
331 attempt to select recordings from the central nucleus of IC (CNIC). At the level of the cochlear  
332 nucleus, the recordings were collected from both the dorsal and ventral divisions.

333 The raw signal was amplified 10,000 times (TDT Medusa). It was then processed by an RX5  
334 multichannel data acquisition system (TDT). The signal collected from each electrode (sampling  
335 rate 25kHz on each channel) was filtered (610-10000 Hz) to extract multi-unit activity (MUA). The  
336 trigger level was set for each electrode to select the largest action potentials from the signal with a  
337 1ms precision. On-line and off-line examination of the waveforms suggests that the MUA collected  
338 here was made of action potentials generated by a few neurons at the vicinity of the electrode.  
339 However, as we did not use tetrodes, the result of several clustering algorithms (Pouzat et al., 2002;  
340 Quiroga et al., 2004; Franke et al., 2015) based on spike waveform analyses were not reliable  
341 enough to isolate single units with good confidence. Although these are not direct proofs, the facts  
342 that the electrodes were of similar impedance (0.5-1 MOhm) and that the spike amplitudes had  
343 similar values (100-300  $\mu$ V) for the cortical and the subcortical recordings were two indications  
344 suggesting that the cluster recordings obtained in each structure included a similar number of  
345 neurons. Even if a similar number of neurons were recorded in the different structures, we cannot  
346 discard the possibility that the homogeneity of the multi-unit recordings (in terms of number of cells  
347 contributing to each recording) differ between structures. By collecting several hundreds of  
348 recordings in each structure, these potential differences should be attenuated in the present study.

349

350

### 351 **Simulations of auditory nerve fiber responses**

352

353 A computational model of auditory nerve fiber responses was used to assess whether the envelope-  
354 tracking properties measured in the central auditory system could be a mere consequence of the  
355 processing taking place at peripheral levels. For this purpose, we used a well-established and  
356 widely-used model of the auditory periphery (Bruce et al., 2018). This model provides a  
357 phenomenological description of the major functional stages of the auditory periphery, from the  
358 middle ear up to the auditory nerve (Osses et al., 2022). The implementation used in the present  
359 study is available as the routine 'bruce2018' within the AMT toolbox (v1.0) for MATLAB (Majdak  
360 et al., 2022).

361 In order to make the simulated data as comparable as possible to the neuronal responses collected in  
362 the electrophysiological experiments, the distribution of cochlear center frequencies was chosen to  
363 be similar to the best frequencies obtained from the CN data. Default parameters were used for the  
364 later stages of the model. For each cochlear channel, five auditory-nerve fibers were simulated with  
365 different spontaneous rates (SR): 1 low-SR fiber (SR = 0.1 spikes/s), 1 medium-SR fiber (SR = 4  
366 spikes/s) and 3 high-SR fibers (SR = 100 spikes/s). The outcome of the model corresponds to the  
367 aggregated responses of these 5 simulated auditory nerve fibers (sANF) in an attempt (i) to keep the  
368 physiological ratio between low, medium and high threshold fibers and (ii) to roughly match the



369 number of cells contributing to the MUA collected in the central auditory structures (<6 neurons at  
370 the vicinity of the electrode).

371 The responses to twenty repetitions of each vocalization in the original and degraded conditions  
372 were simulated and analyzed in the same way as recorded data.

373

### 374 **Experimental protocol**

375 As inserting an array of 16 electrodes in a brain structure unavoidably induces a deformation of this  
376 structure, a 30-min recovery time was allowed for the structure to return to its initial shape, then the  
377 array was slowly lowered. Time-frequency response profiles (TFRPs) were used to assess the  
378 quality of our recordings and to adjust electrode depth. For auditory cortex recordings (A1 and  
379 VRB), the recording depth was 500-1000  $\mu\text{m}$ , which corresponds to layer III and the upper part of  
380 layer IV according to Wallace and Palmer (2008). For thalamic recordings, the NeuroNexus probe  
381 was lowered about 7mm below the pia before the first responses to pure tones were detected. For  
382 the collicular and cochlear nucleus recordings, the NeuroNexus probe was visually inserted into the  
383 structure and after a 15 minutes stabilization period, auditory stimuli were presented.

384 When a clear frequency tuning was obtained for at least 10 of the 16 electrodes, the stability of the  
385 tuning was assessed: we required that the recorded neurons displayed at least three (each lasting 6  
386 minutes) successive similar TFRPs (i.e., with similar best frequencies) before starting the protocol.

387 When the stability was satisfactory, the protocol was started by presenting the acoustic stimuli in  
388 the following order: We first presented the four whistles at 75 dB SPL in their original versions (in  
389 quiet), then the vocoded whistles (Voc38, Voc20 and Voc10 versions) were presented at 75 dB SPL  
390 followed by the masked vocalizations presented against the chorus then against the vocalization-  
391 shaped stationary noise at 65, 75 and 85 dB SPL. Thus, the level of the original vocalizations was  
392 kept constant (75 dB SPL), and the noise level was increased (65, 75 and 85 dB SPL). In all cases,  
393 each vocalization was repeated 20 times and all the loudness levels are in RMS value. Presentation  
394 of this entire stimulus set lasted 45 minutes. The protocol was re-started either after moving the  
395 electrode array on the cortical map or after lowering the NeuroNexus probe by at least 300  $\mu\text{m}$  for  
396 subcortical structures.

397

### 398 **Behavioral Go/No-Go discrimination task**

399 Nine eight-weeks old C57Bl/6J mice were water-deprived (33 ml/g per day) and trained daily for  
400 200–300 trials in a Go/No-Go task involving two of the guinea pig whistles (W1 and W3 in figure  
401 1), one (the S+) signaling the reward (a drop of water) and the other not (the S-). The training  
402 procedures were similar to those described in previous studies (Deneux et al., 2016; Ceballo et al.,  
403 2019). Mice were head-fixed and held in a plastic tube on aluminum foil. Mice first performed 1-3

404 habituation sessions to learn to obtain a water reward (~5  $\mu$ l) by licking on a stainless steel water  
405 spout at least 8 times after the positive stimulus S+. A trial only started when the mice were not  
406 licking the spout for at least 3 seconds. Licks were detected by changes in resistance between the  
407 aluminum foil and the water spout. After habituation, the fraction of collected rewards was ~80%.  
408 The learning protocol then started in which mice received the S- for which they had to lick less  
409 than 3 times to avoid a 5s time-out. One of the two whistles (the S+ or the S-) was presented every  
410 10-20 s (uniform distribution) followed by a 1s test period during which the mouse had to lick at  
411 least 5-8 times to receive the reward. Positive and negative stimuli were played in a pseudorandom  
412 order with the constraint that exactly 4 positive and 4 negative sounds must be played every 8 trials.  
413 Once a mouse showed at least 80% of correct discrimination between the S+ and the S- for two  
414 successive days in the original condition, it was trained in noisy conditions, first with the stationary  
415 noise (successively at +10, 0 and -10 dB SNR), then with the chorus noise (successively at +10, 0  
416 and -10 dB SNR). Each mouse had to perform at least one day at 80% in a given SNR to be tested  
417 on the following day at a lower SNR. Behavioral analyses were all automated; thus no animal  
418 randomization or experimenter blinding was used.

419

## 420 **Data analysis**

421

422 All the analyses were performed on MATLAB 2021 (MathWorks).

### 423 **Quantification of responses to pure tones**

424 The TFRPs were obtained by constructing post-stimulus time histograms for each frequency with 1  
425 ms time bins. The firing rate evoked by each frequency was quantified by summing all the action  
426 potentials from the tone onset up to 100 ms after this onset. Thus, TFRPs were matrices of 100 bins  
427 in abscissa (time) multiplied by 129 bins in ordinate (frequency). All TFRPs were smoothed with a  
428 uniform 5x5 bin window for visualization (not for the data analyses). For each TFRP, the Best  
429 Frequency (BF) was defined as the frequency at which the highest firing rate was recorded. Peaks  
430 of significant response were automatically identified using the following procedure: A positive peak  
431 in the TFRP was defined as a contour of firing rate above the average level of the baseline activity  
432 (100ms of spontaneous activity taken before each tone onset) plus six times the standard deviation  
433 of the baseline activity. Recordings without significant peak of responses or with inhibitory  
434 responses (decreases in firing rate 3 standard deviations below spontaneous activity) were excluded  
435 from the data analyses.

## 436 **Quantification of the envelope tracking**

437 We first extracted the envelope as explained on a previous section (see above “Overall envelope  
438 extraction”) and then filtered them (original, vocoded and noisy vocalizations) using a bank of 35  
439 gammatone filters with center frequencies uniformly spaced along a guinea pig - adapted ERB  
440 (equivalent rectangular bandwidth) scale ranging from 20 to 30 000 Hz. Then, three ranges of  
441 amplitude modulation (AM) were investigated: the low (L, < 20 Hz), middle (M, between 20 and  
442 100 Hz) and high (H, between 100 and 200 Hz) AM ranges. For all the AM filtering, we used  
443 Butterworth filters at -6 dB per octave. Second, the envelopes were downsampled to a resolution of  
444 1 ms to match the sampling rate of the PSTHs. Finally, we applied a half-wave rectification  
445 followed by a normalization with the corresponding RMS value.

446 The neuronal responses (i.e. the PSTHs) were also filtered with the same three frequency bands as  
447 the envelopes followed by a normalization with the corresponding RMS value. The rationale for  
448 this filtering step was that we wanted to isolate and quantify the correspondence between temporal  
449 aspects of the stimuli in particular frequency ranges and PSTHs.

450 Next, we performed normalized cross-correlations between the filtered envelopes and PSTHs for  
451 each AM range. We selected seven gammatones, as a trade-off between accurately representing the  
452 envelopes along the audio spectrum and minimizing redundancy between envelopes. Maximal  
453 values in the correlograms were automatically detected in each structure to account for propagation  
454 delays in the auditory system. The lags were selected according to the distributions of the latencies  
455 obtained in response to pure tones at 75 dB SPL. The different lags identified were: 1-10ms for CN,  
456 5-20ms for CNIC, 6-15ms for MGv, 9-30ms for A1 and 9-40ms for VRB. In all analyses, we  
457 decided to keep the maximal correlation coefficient out of the seven selected gammatone filters  
458 ( $R_{\max_{E-PSTH}}$ ).

459  
460 *Evaluation of the correlation significance by shuffling the evoked activity*

461 It is known that significant correlation between neuronal events and sensory stimuli can be obtained  
462 by chance (see for review Harris, 2020). Therefore, it was crucial to run drastic controls to reduce  
463 the probability that the correlations detected here result from spurious correlations.

464 To determine a significance threshold for the correlation, we shuffled only the evoked activity in  
465 the original condition on a time-scale of 1 ms, in order to preserve the global shape of the whole  
466 response (i.e., the four response peaks due to the starting of each stimulus separated by a period of  
467 silence). Specifically, for the original condition, we only shuffled the spikes obtained during the  
468 presentation of each whistle to avoid adding spikes in the silence period. The obtained shuffled  
469 PSTHs were then processed using the same procedure as for unshuffled PSTHs: filtering in the  
470 three AM ranges and half-wave rectification followed by a normalization with the corresponding

471 RMS value. Then, we computed the cross-correlation ( $R_{\text{Random}}$ ) between each shuffled PSTH and  
 472 each envelope. We performed this procedure 1000 times and set, for each correlation value PSTH-  
 473 E, a significance threshold of the R value that is the mean of the  $R_{\text{Random}}$  values plus two fold the  
 474 standard deviations ( $\mu(R_{\text{Random}}) \pm 2\sigma$ ). Based upon this criterion, percentages of recordings were  
 475 discarded in each structure and for each AM range: in VRB, 30%, 10%, 47% of recordings were  
 476 discarded in the L, M and H range respectively; in A1, 51%, 38%, 73% of recordings were  
 477 discarded in the L, M and H range respectively; in MGv, 61%, 49%, 63% of recordings were  
 478 discarded in the L, M and H range respectively; in CNIC, 29%, 26%, 33% of recordings were  
 479 discarded in the L, M and H range respectively; in CN, 33%, 43%, 50% of recordings were  
 480 discarded in the L, M and H range respectively; in sANF, 35%, 86%, 77% of recordings were  
 481 discarded in the L, M and H range respectively. Although this drastic procedure discarded a non-  
 482 negligible proportion of recordings, it reduced the probability that the correlations described here  
 483 were obtained by chance.

484

#### 485 **Quantification of mutual information from the responses to vocalizations**

486 The method developed by Schnupp and colleagues (2006) was used to quantify the amount of  
 487 information contained in the responses to vocalizations obtained with natural, vocoded or noisy  
 488 stimuli. This method allows quantifying how well the vocalization's identity can be inferred from  
 489 neuronal responses. Neuronal responses were represented using different time scales ranging from  
 490 the duration of the whole response (total spike count) to a 1-ms precision (precise temporal  
 491 patterns), which allows analyzing how much the spike timing contributes to the information. As this  
 492 method is exhaustively described in Schnupp and colleagues (2006) and in Gaucher and colleagues  
 493 (2013a), we only present below the main principles.

494 The method relies on a pattern-recognition algorithm that is designed to “guess which stimulus  
 495 evoked a particular response pattern” (Schnupp et al., 2006) by going through the following steps:  
 496 From all the responses of a subcortical or cortical site to the different stimuli, a single response (test  
 497 pattern) is extracted and represented as a PSTH with a given bin size. Then, a mean response  
 498 pattern is computed from the remaining responses for each stimulus class. The test pattern is then  
 499 assigned to the stimulus class of the closest mean response pattern. This operation is repeated for all  
 500 the responses, generating a confusion matrix where each response is assigned to a given stimulus  
 501 class. From this confusion matrix, the Mutual Information (MI) is given by Shannon's formula:

$$502 \quad MI = \sum_{x,y} p(x,y) \times \log_2 \left( \frac{p(x,y)}{p(x) \times p(y)} \right)$$

503 where x and y are the rows and columns of the confusion matrix, or in other words, the values taken  
 504 by the random variables “presented stimulus class” and “assigned stimulus class”.

505 In our case, we used responses to the four whistles and selected the first 280 ms of these responses  
506 to work on spike trains of exactly the same duration (the shortest whistle being 280 ms long). In a  
507 scenario where the responses do not carry information, the assignments of each response to a mean  
508 response pattern is equivalent to chance level (here 0.25 because we used 4 different stimuli and  
509 each stimulus was presented the same number of times) and the MI would be close to zero. In the  
510 opposite case, when responses are very different between stimulus classes and very similar within a  
511 stimulus class, the confusion matrix would be diagonal and the mutual information would tend to  
512  $\log_2(4) = 2$  bits. This algorithm was applied with different bin sizes ranging from 1 to 280 ms (see  
513 figure 2B in Souffi and colleagues (2020) for the evolution of MI with temporal precisions ranging  
514 from 1 to 40 ms). The value of 8 ms was selected for the data analysis because in each structure the  
515 MI reached its maximum at this value of temporal precision.

516 The MI estimates are subject to non-negligible positive sampling biases. Therefore, as in Schnupp  
517 and colleagues (2006), we estimated the expected size of this bias by calculating MI values for  
518 “shuffled” data, in which the response patterns were randomly reassigned to stimulus classes. The  
519 shuffling was repeated 100 times, resulting in 100 MI estimates of the bias ( $MI_{\text{bias}}$ ). These  $MI_{\text{bias}}$   
520 estimates are then used as estimators for the computation of the statistical significance of the MI  
521 estimate for the real (unshuffled) datasets: the real estimate is considered as significant if its value is  
522 statistically different from the distribution of  $MI_{\text{bias}}$  shuffled estimates. Significant MI estimates  
523 were computed for MI calculated from neuronal responses under one electrode and for each  
524 condition. Therefore, there was a  $MI_{\text{bias}}$  value for each MI estimate. The range of  $MI_{\text{bias}}$  values was  
525 very similar between brain structures: depending on the conditions (original, vocoded and noisy  
526 vocalizations), it ranged from 0.102 to 0.107 bits in the CN, from 0.107 to 0.110 bits in the IC, from  
527 0.105 to 0.114 bits in the MGB, 0.107 to 0.111 bits in the A1 and from 0.106 to 0.116 bits in VRB.  
528 There was no significant difference between the mean values of  $MI_{\text{bias}}$  in the different structures  
529 (Students’ t test unpaired, all  $p > 0.25$ ).

530

### 531 **Quantification of acoustic envelope similarity**

532 For each acoustic condition and each AM range, we quantified the acoustic similarity between each  
533 pair of stimuli as the correlation between their envelopes across the seven selected gammatones.  
534 Then, we averaged the six correlation values (related to all possible combinations with the four  
535 stimuli) to obtain an estimate of the similarity between the four stimuli for each condition (original,  
536 vocoding and noisy conditions) and each AM range (see Fig. 7A, dark lines). More precisely, we  
537 averaged Fisher z-transformed coefficients and reported the back-transformed averages on the  
538 figure 7A. In order to confirm that there is no bias in our gammatone selection, we carried out the

539 same analysis on the output of the 35 gammatones and obtained similar results (see Fig. 7A, light  
540 lines).

541

#### 542 **Statistical analysis**

543

544 We used an analysis of variance (ANOVA) for multiple factors to reveal the main effects in the  
545 whole data set (vocoding conditions: three levels, masking noise conditions: three levels for each  
546 noise; auditory structures: six levels; AM ranges: three levels). Post-hoc pairwise tests were  
547 performed between the original condition and the vocoding or noisy conditions, or between  
548 structures to assess the significance of the multiple comparisons. They were corrected for multiple  
549 comparisons using Bonferroni corrections and were considered as significant if their p-value was  
550 below 0.05.

551 **Results**

552

553 We simulated auditory nerve fiber (sANF) responses and collected neuronal recordings from five  
554 auditory structures: the cochlear nucleus (CN, 10 animals), the central nucleus of the inferior  
555 colliculus (CNIC, 11 animals), the ventral part of the medial geniculate (MGv, 10 animals), the  
556 primary auditory cortex (A1, 11 animals) and a secondary auditory area (VRB, 5 animals).

557 All analyses were performed on a set of recordings (or simulated recordings) selected using  
558 stringent criteria and  $n$  values correspond to the number of selected recordings. Note that all the R  
559 values presented below are considered as significant (see Method section for more details).

560 Figures 1A-B illustrate the spectrograms and the overall envelopes of all stimuli in the original,  
561 vocoded and noisy conditions. In the following, the term stimulus refers either to the four original  
562 or vocoded whistles, or to the four whistles embedded in noise. The four overall envelopes of the  
563 stimuli were clearly different between each other in the original and vocoded conditions, however,  
564 they progressively became more similar in noisy conditions as the SNR decreased, especially in  
565 stationary noise.

566

567 **Auditory neurons track the envelopes in the low AM range better than in middle and high**  
568 **AM ranges**

569

570 We first determined which ranges of amplitude modulations are tracked by the subcortical and  
571 cortical neurons. To address this question, we filtered both the envelopes and the neuronal  
572 responses in three AM ranges: the low ( $< 20$  Hz), middle (between 20 and 100 Hz) and high  
573 (between 100 and 200 Hz) AM ranges. Figure 1C presents the seven selected Butterworth-filtered  
574 envelopes (among the 35) of the four whistles after first having been filtered using a gammatone  
575 filterbank and brings out that the low AM range contained larger envelope fluctuations than the  
576 middle and high AM ranges. Figures 2A-B present individual examples (Fig. 2A) and populations  
577 (Fig. 2B) of PSTHs constructed from the responses to presentation of the original vocalizations in  
578 each structure. Based on these PSTHs, it appears that the evoked responses tended to be more  
579 phasic in the two cortical areas (A1 and VRB) than in the subcortical structures.

580 Figures 2C-E show the PSTHs from individual recordings (in black) and stimulus envelopes (E, in  
581 red) obtained in each structure (and for the sANF) in the original condition, both filtered in the  
582 same AM ranges. For each example of E-PSTH, we indicated the cross-correlation value (R) on the  
583 top left of each panel. In the following results, the correlation value selected for each recording at a  
584 given AM range was the maximum over the seven gammatone filters ( $R_{\max_{E-PSTH}}$ ). Note that  
585 similar results were obtained when we used the correlation value obtained with the gammatone

586 filter the closest to the best frequency of each neuronal recording (data not shown). Whatever the  
587 structure, in these individual recordings, the higher R values were in the low AM range rather than  
588 in the middle and high AM ranges (Fig. 2C-E). Figure 2F presents the distribution, the mean and  
589 the interquartile range of the  $R_{\max_{E-PSTH}}$  values for each structure in the three AM ranges (L, M and  
590 H) in the original condition. Overall, we found a statistically significant difference in average  
591  $R_{\max_{E-PSTH}}$  values for both the three AM ranges and the six structures (two-way ANOVA,  $p <$   
592  $0.05$ ) with a significant interaction between these two factors. For all structures, the mean  $R_{\max_{E-}}$   
593  $PSTH}$  values were much higher in the L range compared to the M and H ranges. In the low AM  
594 range, sANF, CN and CNIC recordings displayed significantly higher mean  $R_{\max_{E-PSTH}}$  values than  
595 MGv and cortical recordings (one-way ANOVA,  $p < 0.0001$   $F_{(5, 1165)} = 138.25$  with Students' t test  
596 unpaired, sANF vs. MGv, A1 or VRB  $p < 0.0001$ , CN vs. MGv, A1 or VRB  $p < 0.0001$ , CNIC vs.  
597 MGv, A1 or VRB  $p < 0.0001$ ; mean ( $\pm$  STD)  $R_{\max_{E-PSTH}}$  values:  $R_{sANF(n=217)} = 0.81 \pm 0.03$ ,  $R_{CN(n=}$   
598  $336)} = 0.80 \pm 0.08$ ,  $R_{CNIC(n=274)} = 0.78 \pm 0.11$ ,  $R_{MGv(n=102)} = 0.68 \pm 0.13$ ,  $R_{A1(n=171)} = 0.60 \pm 0.13$  and  
599  $R_{VRB(n=66)} = 0.63 \pm 0.14$ ). In the middle and high AM ranges, the differences between the structures  
600 were less clear but the CNIC recordings still exhibited slightly higher mean  $R_{\max_{E-PSTH}}$  values  
601 compared to the other structures (mean ( $\pm$  STD)  $R_{\max_{E-PSTH}}$  values in the middle AM range:  
602  $R_{sANF(n=44)} = 0.32 \pm 0.03$ ,  $R_{CN(n=285)} = 0.31 \pm 0.07$ ,  $R_{CNIC(n=285)} = 0.34 \pm 0.07$ ,  $R_{MGv(n=133)} = 0.31 \pm$   
603  $0.07$ ,  $R_{A1(n=220)} = 0.27 \pm 0.05$  and  $R_{VRB(n=85)} = 0.29 \pm 0.06$ ; mean ( $\pm$  STD)  $R_{\max_{E-PSTH}}$  values in the  
604 high AM range:  $R_{sANF(n=77)} = 0.31 \pm 0.03$ ,  $R_{CN(n=249)} = 0.31 \pm 0.05$ ,  $R_{CNIC(n=257)} = 0.33 \pm 0.05$ ,  $R_{MGv(n=}$   
605  $97)} = 0.29 \pm 0.04$ ,  $R_{A1(n=196)} = 0.27 \pm 0.05$  and  $R_{VRB(n=50)} = 0.30 \pm 0.05$ ). This poor ability to follow  
606 fast AM changes was expected for auditory cortex neurons but not expected for subcortical neurons  
607 and for sANF (which can synchronize at higher AM rates when tested with periodic artificial  
608 stimuli, review in Joris et al., 2004). This suggests that only a partial encoding of high AM rates  
609 contained in complex natural sounds is performed by subcortical neurons.

610 To summarize, in the original condition, the neurons' PSTHs were more strongly correlated with  
611 the stimulus envelope in the low AM range than in the middle and high AM ranges, both at the  
612 subcortical and cortical levels.

613

614 **In the original condition, the better cortical and subcortical neurons track the slow envelope**  
615 **(<20 Hz), the higher the value of mutual information**

616

617 Does envelope tracking allow auditory neurons to discriminate the four vocalizations in the original  
618 condition? To address this question, we examined whether there is a relationship between the  
619 neuronal discrimination performance and the neurons' abilities to follow the stimulus envelope  
620 (Fig. 3). The distribution, the mean and the interquartile range of the neuronal discrimination



621 (quantified by the mutual information, MI) are presented for each structure in Figure 3A. As  
622 previously reported (Souffi et al., 2020), subcortical neurons (CN, CNIC and MGv neurons) were  
623 better in discriminating the original whistles compared to cortical neurons (A1 and VRB neurons)  
624 and here we extended this result to sANF (one-way ANOVA  $p < 0.0001$   $F_{(5, 1538)} = 266.46$  with  
625 Students' t test unpaired, sANF vs. A1 or VRB  $p < 0.0001$ , CN vs. A1 or VRB  $p < 0.0001$ , CNIC vs.  
626 A1 or VRB  $p < 0.0001$ , MGv vs. A1 or VRB  $p < 0.0001$ ; mean ( $\pm$  STD) MI values:  $MI_{sANF(n=77)} =$   
627  $1.84 \pm 0.21$  bits,  $MI_{CN(n=249)} = 0.92 \pm 0.47$  bits,  $MI_{CNIC(n=257)} = 1.00 \pm 0.5$  bits,  $MI_{MGv(n=97)} = 1.19 \pm$   
628  $0.55$  bits,  $MI_{A1(n=196)} = 0.68 \pm 0.37$  bits and  $R_{VRB(n=50)} = 0.55 \pm 0.29$  bits, Fig. 3A). The scattergrams  
629 presented in Figure 3B display the  $R_{maxE-PSTH}$  values as a function of the MI values in each  
630 structure and AM range.

631 Figure 3C summarizes the correlation values between  $R_{maxE-PSTH}$  and MI parameters, in each  
632 structure and in the three AM ranges. All significant correlation values between these two variables  
633 are reported in red. In all but one case (in CNIC in the middle AM range), significant positive  
634 correlations between  $R_{maxE-PSTH}$  and MI values were obtained in all AM ranges in subcortical  
635 structures ( $p^L_{CN} < 0.0001$ ,  $p^M_{CN} < 0.0001$ ,  $p^H_{CN} = 0.01$ ,  $p^L_{CNIC} < 0.0001$ ,  $p^M_{CNIC} = 0.24$ ,  $p^H_{CNIC} = 0.005$ ,  
636  $p^L_{MGv} = 0.006$ ,  $p^M_{MGv} < 0.0001$ ,  $p^H_{MGv} = 0.01$ ). For the sANF, the range of MI values was too limited  
637 to compute reliable correlations (most MI values were close to the maximum of 2 bits). At the  
638 subcortical level, the highest correlation values between  $R_{maxE-PSTH}$  and MI values as a whole,  
639 were found in MGv. At the cortical level, significant correlations between  $R_{maxE-PSTH}$  and MI  
640 values were detected in the low and middle AM ranges in A1 ( $p^L_{A1} = 0.01$ ,  $p^M_{A1} = 0.05$ ,  $p^H_{A1} = 0.32$ )  
641 and there was no significant correlation in VRB (may be as a consequence of fewer recordings in  
642 this area,  $p^L_{VRB} = 0.31$ ,  $p^M_{VRB} = 0.51$ ,  $p^H_{VRB} = 0.99$ ). Interestingly, at each level except in VRB, there  
643 was a positive and significant correlation value in the low AM range suggesting that the neuronal  
644 ability for tracking the slow envelopes (<20 Hz) better explains the neuronal discrimination in the  
645 entire auditory system than the tracking of higher AM rates.

646 To summarize, it appeared that in the original condition, the better the tracking of the temporal  
647 envelope, the better the between-stimuli neuronal discrimination. In addition, for cortical neurons,  
648 the correlation between the  $R_{maxE-PSTH}$  and MI values was stronger in the lower AM range,  
649 whereas for subcortical neurons, there were also still significant correlations in the higher AM  
650 ranges.

651  
652 **The acoustic degradations decreased the values of mutual information but did not affect the**  
653 **envelope tracking performed by the auditory neurons**

654

655 In almost all situations of acoustic degradations, the neurons' ability to discriminate between the  
 656 four vocalizations was decreased. Figures 4A-F present the distributions of the MI values for the  
 657 original condition and the three levels of degradation conditions: the three tone-vocoders (38, 20  
 658 and 10 frequency bands) or the three SNRs (+10, 0 and -10 dB SNR) in the two types of noise  
 659 (stationary and chorus noises). In general, there were modest effects on the MI values in the  
 660 vocoding and chorus noise conditions compared to the stationary noise conditions for all structures  
 661 except sANF. The decrease was significant only for the 10-band vocoded vocalizations in MGv and  
 662 A1 (Fig. 4D, one-way ANOVA,  $p = 0.05$   $F_{(3, 811)} = 2.58$  with Students' t test paired,  $MGv_{Ori}$  vs.  
 663  $MGv_{Voc10}$   $p < 0.0001$ , mean ( $\pm$  STD) MI values:  $MI_{Ori} = 1.19 \pm 0.55$  bits,  $MI_{Voc38} = 1.13 \pm 0.57$  bits,  
 664  $MI_{Voc20} = 1.15 \pm 0.55$  bits,  $MI_{Voc10} = 1.03 \pm 0.53$  bits; Fig. 4E, one-way ANOVA,  $p = 0.001$   $F_{(3, 722)}$   
 665  $= 3.73$  with Students' t test paired,  $A1_{Ori}$  vs.  $A1_{Voc10}$   $p < 0.0001$ , mean ( $\pm$  STD) MI values:  $MI_{Ori} =$   
 666  $0.68 \pm 0.37$  bits,  $MI_{Voc38} = 0.62 \pm 0.33$  bits,  $MI_{Voc20} = 0.61 \pm 0.33$  bits,  $MI_{Voc10} = 0.56 \pm 0.30$  bits),  
 667 and no significant difference was detected in VRB (Fig.4F, one-way ANOVA,  $p = 0.75$   $F_{(3, 186)} =$   
 668  $0.41$ , mean ( $\pm$  STD) MI values:  $MI_{Ori} = 0.55 \pm 0.29$  bits,  $MI_{Voc38} = 0.55 \pm 0.28$  bits,  $MI_{Voc20} = 0.55 \pm$   
 669  $0.29$  bits,  $MI_{Voc10} = 0.61 \pm 0.31$  bits). However, the decrease was already significant with 38-band  
 670 vocoded vocalizations in sANF or with 20-band vocoded vocalizations in CN and CNIC (Fig.4A,  
 671 one-way ANOVA,  $p < 0.0001$   $F_{(3, 1302)} = 111.3$  with Students' t test paired,  $sANF_{Ori}$  vs.  $sANF_{Voc38}$   $p$   
 672  $< 0.0001$ , mean ( $\pm$  STD) MI values:  $MI_{Ori} = 1.84 \pm 0.21$  bits,  $MI_{Voc38} = 1.61 \pm 0.32$  bits,  $MI_{Voc20} =$   
 673  $1.39 \pm 0.52$  bits,  $MI_{Voc10} = 1.29 \pm 0.52$  bits; Fig.4B, one-way ANOVA,  $p < 0.0001$   $F_{(3, 1424)} = 12.42$   
 674 with Students' t test paired,  $CN_{Ori}$  vs.  $CN_{Voc20}$   $p < 0.0001$ , mean ( $\pm$  STD) MI values:  $MI_{Ori} = 0.92 \pm$   
 675  $0.47$  bits,  $MI_{Voc38} = 0.96 \pm 0.46$  bits,  $MI_{Voc20} = 0.86 \pm 0.45$  bits,  $MI_{Voc10} = 0.75 \pm 0.43$  bits; Fig.4C,  
 676 one-way ANOVA,  $p < 0.0001$   $F_{(3, 1231)} = 13.17$  with Students' t test paired,  $CNIC_{Ori}$  vs.  $CNIC_{Voc20}$   
 677  $p < 0.0001$ , mean ( $\pm$  STD) MI values:  $MI_{Ori} = 1.00 \pm 0.49$  bits,  $MI_{Voc38} = 0.99 \pm 0.46$  bits,  $MI_{Voc20} =$   
 678  $0.90 \pm 0.46$  bits,  $MI_{Voc10} = 0.79 \pm 0.39$  bits). Note that there was also a significant increase in MI  
 679 values with the 38-band vocoded vocalizations in CN (Fig.4B, one-way ANOVA,  $p < 0.0001$   $F_{(3,$   
 680  $1424)} = 12.42$  with Students' t test paired,  $CN_{Ori}$  vs.  $CN_{Voc38}$   $p = 0.0073$ ).  
 681 In chorus noise, in CN and CNIC, there was no significant decrease in mean MI values (Fig. 4B,  
 682 one-way ANOVA,  $p = 0.05$   $F_{(3, 1176)} = 2.65$ , mean ( $\pm$  STD) MI values:  $MI_{Ori} = 0.92 \pm 0.47$  bits,  
 683  $MI_{+10dB} = 0.85 \pm 0.48$  bits,  $MI_{0dB} = 0.83 \pm 0.50$  bits,  $MI_{-10dB} = 0.83 \pm 0.49$  bits; Fig. 4C, one-way  
 684 ANOVA,  $p = 0.36$   $F_{(3, 1188)} = 1.06$ ,  $MI_{Ori} = 1.00 \pm 0.49$  bits,  $MI_{+10dB} = 1.05 \pm 0.50$  bits,  $MI_{0dB} = 1.06$   
 685  $\pm 0.52$  bits,  $MI_{-10dB} = 1.06 \pm 0.52$  bits), whereas in sANF and MGv, the mean MI values  
 686 significantly decreased at +10 dB or 0 dB SNR respectively (Fig. 4A, one-way ANOVA,  $p <$   
 687  $0.0001$   $F_{(3, 1331)} = 232.86$  with Students' t test paired,  $sANF_{Ori}$  vs.  $sANF_{+10dB}$   $p < 0.0001$ , mean ( $\pm$   
 688 STD) MI values:  $MI_{Ori} = 1.84 \pm 0.21$  bits,  $MI_{+10dB} = 1.48 \pm 0.37$  bits,  $MI_{0dB} = 1.20 \pm 0.45$  bits,  $MI_{-$   
 689  $10dB} = 1.09 \pm 0.50$  bits; Fig. 4D, one-way ANOVA,  $p < 0.0001$   $F_{(3, 753)} = 7.3$  with Students' t test

690 paired,  $MGv_{Ori}$  vs.  $MGv_{0dB}$   $p < 0.0001$ , mean ( $\pm$  STD) MI values:  $MI_{Ori} = 1.19 \pm 0.55$  bits,  $MI_{+10dB} = 1.08 \pm 0.52$  bits,  $MI_{0dB} = 0.99 \pm 0.50$  bits,  $MI_{-10dB} = 0.98 \pm 0.48$  bits). At the cortical level, there  
691 was a significant decrease in A1 at 0 dB SNR and no significant change of mean MI values in VRB  
692 (Fig. 4E, one-way ANOVA,  $p = 0.0039$   $F_{(3, 697)} = 4.5$  with Students' t test paired,  $A1_{Ori}$  vs.  $A1_{0dB}$   $p$   
693  $< 0.0001$ , mean ( $\pm$  STD) MI values:  $MI_{Ori} = 0.68 \pm 0.37$  bits,  $MI_{+10dB} = 0.64 \pm 0.36$  bits,  $MI_{0dB} =$   
694  $0.55 \pm 0.30$  bits,  $MI_{-10dB} = 0.60 \pm 0.31$  bits; Fig. 4F, one-way ANOVA,  $p = 0.31$   $F_{(3, 179)} = 1.19$ , mean  
695 ( $\pm$  STD) MI values:  $MI_{Ori} = 0.55 \pm 0.29$  bits,  $MI_{+10dB} = 0.63 \pm 0.32$  bits,  $MI_{0dB} = 0.54 \pm 0.27$  bits,  
696  $MI_{-10dB} = 0.53 \pm 0.24$  bits).

697 Stationary noise strongly reduced the MI values compared to the vocoding and the chorus noise  
698 addition. The mean MI value in sANF, CN and MGv was significantly reduced already at +10 dB  
699 SNR (Fig. 4A, one-way ANOVA,  $p < 0.0001$   $F_{(3, 1153)} = 767.64$  with Students' t test paired,  $sANF_{Ori}$   
700 vs.  $sANF_{+10dB}$   $p < 0.0001$ , mean ( $\pm$  STD) MI values:  $MI_{Ori} = 1.84 \pm 0.21$  bits,  $MI_{+10dB} = 1.09 \pm 0.39$   
701 bits,  $MI_{0dB} = 0.68 \pm 0.37$  bits,  $MI_{-10dB} = 0.66 \pm 0.39$  bits; Fig. 4B, one-way ANOVA,  $p < 0.0001$   $F_{(3,$   
702  $812)} = 61.22$  with Students' t test paired,  $CN_{Ori}$  vs.  $CN_{+10dB}$   $p < 0.0001$ , mean ( $\pm$  STD) MI values:  
703  $MI_{Ori} = 0.92 \pm 0.47$  bits,  $MI_{+10dB} = 0.68 \pm 0.38$  bits,  $MI_{0dB} = 0.53 \pm 0.26$  bits,  $MI_{-10dB} = 0.41 \pm 0.17$   
704 bits; Fig. 4D, one-way ANOVA,  $p < 0.0001$   $F_{(3, 630)} = 62.03$  with Students' t test paired,  $MGv_{Ori}$  vs.  
705  $MGv_{+10dB}$   $p < 0.0001$ , mean ( $\pm$  STD) MI values:  $MI_{Ori} = 1.19 \pm 0.55$  bits,  $MI_{+10dB} = 1.00 \pm 0.50$  bits,  
706  $MI_{0dB} = 0.74 \pm 0.35$  bits,  $MI_{-10dB} = 0.46 \pm 0.19$  bits), whereas the mean MI value in CNIC was  
707 significantly reduced at 0 dB SNR (Fig. 4C, one-way ANOVA,  $p < 0.0001$   $F_{(3, 1078)} = 32.08$  with  
708 Students' t test paired,  $CNIC_{Ori}$  vs.  $CNIC_{0dB}$   $p < 0.0001$ , mean ( $\pm$  STD) MI values:  $MI_{Ori} = 1.00 \pm$   
709  $0.49$  bits,  $MI_{+10dB} = 1.00 \pm 0.49$  bits,  $MI_{0dB} = 0.87 \pm 0.44$  bits,  $MI_{-10dB} = 0.63 \pm 0.28$  bits). At the  
710 cortical level, noise significantly reduced the mean MI value in A1 only at -10 dB SNR (Fig. 4E,  
711 one-way ANOVA,  $p < 0.0001$   $F_{(3, 669)} = 13.99$  with Students' t test paired,  $A1_{Ori}$  vs.  $A1_{-10dB}$   $p <$   
712  $0.0001$ , mean ( $\pm$  STD) MI values:  $MI_{Ori} = 0.68 \pm 0.37$  bits,  $MI_{+10dB} = 0.62 \pm 0.34$  bits,  $MI_{0dB} = 0.59$   
713  $\pm 0.31$  bits,  $MI_{-10dB} = 0.42 \pm 0.18$  bits), whereas the mean MI values in VRB remained unchanged in  
714 all conditions (Fig. 4F, one-way ANOVA,  $p = 0.26$   $F_{(3, 164)} = 1.34$ , mean ( $\pm$  STD) MI values:  $MI_{Ori}$   
715  $= 0.55 \pm 0.29$  bits,  $MI_{+10dB} = 0.55 \pm 0.28$  bits,  $MI_{0dB} = 0.51 \pm 0.22$  bits,  $MI_{-10dB} = 0.42 \pm 0.17$  bits).

716 What can be the scenarios explaining the decrease in neuronal discrimination and involving the  
717 envelope tracking in situations of acoustic degradations? At least two scenarios can be envisioned:  
718 First, in a condition-independent envelope tracking scenario, a neuron keeps the same intrinsic  
719 capacity to track the stimulus envelopes whatever the acoustic conditions, i.e., both in quiet and in  
720 conditions of acoustic degradations. In that case, as long as the stimulus envelopes present some  
721 differences, the neuron will detect these differences and will discriminate the stimuli. Second, in a  
722 condition-dependent scenario, the acoustic degradations reduce the neurons' ability to track the  
723 stimulus envelope. In that case, despite differences between the stimulus envelopes, the intense  
724

725 activity occurring when the neurons are strongly driven by the noise prevents the recorded neuron  
726 from tracking the stimulus envelopes. To determine which of these two scenarios actually operates,  
727 we investigated whether the  $R_{\max_{E-PSTH}}$  values were changed in the conditions of acoustic  
728 degradations such as the vocoding or the noise addition (Figures 5-6). Figure 5 shows, for  
729 individual recordings, the superpositions of the PSTH and the envelope (E) in the low AM range for  
730 the original condition and all the degraded conditions (vocoding, stationary and chorus noise) in  
731 each auditory structure. The  $R_{\max_{E-PSTH}}$  values are indicated on the top left of each panel. In these  
732 individual recordings, the  $R_{\max_{E-PSTH}}$  values presented very little changes in all degraded  
733 conditions compared to the original condition.

734 We next quantified for each recording, the  $R_{\max_{E-PSTH}}$  variations compared with the original  
735 condition ( $\Delta R_{\max_{E-PSTH}}$ ). This was quantified in each structure, for all the degraded conditions and  
736 each AM range (Fig. 6). Compared with the  $R_{\max_{E-PSTH}}$  values obtained in the original condition,  
737 there was little or no change in the degraded conditions for all structures. More precisely, in sANF  
738 and CN, we observed a maximal increase in mean ( $\pm$  STD)  $R_{\max_{E-PSTH}}$  values of 0.16 ( $\pm$ 0.03) (and  
739 0.11 ( $\pm$ 0.13) for CN) and, a maximal decrease of 0.10 ( $\pm$ 0.04) (and 0.05 ( $\pm$ 0.06) for CN) depending  
740 on the degraded conditions and the AM range (Fig. 6). In CNIC and MGv, the mean ( $\pm$  STD)  
741  $R_{\max_{E-PSTH}}$  changes in degraded conditions were very small (Fig. 6, between -0.06 ( $\pm$ 0.03) and  
742 0.006 ( $\pm$ 0.06) for CNIC and between -0.08 ( $\pm$ 0.08) and -0.0002 ( $\pm$ 0.15) for MGv). In A1, the  
743 changes in mean ( $\pm$  STD)  $R_{\max_{E-PSTH}}$  values varied between -0.09 ( $\pm$ 0.11) and 0.07 ( $\pm$ 0.15) and in  
744 VRB it varied between -0.13 ( $\pm$ 0.13) and 0.07 ( $\pm$  0.15).

745 These results clearly provide evidence that the abilities of neurons for tracking the temporal  
746 envelope cues were preserved at each level of the auditory system and in all the situations of  
747 acoustic degradations.

748

#### 749 **The increase in between-envelope similarity explains the decrease in neuronal discrimination**

750

751 If the neurons are still able to track the stimulus envelopes in all conditions of acoustic  
752 degradations, what could explain the pronounced MI decrease in these situations? The most  
753 parsimonious explanation is that the noise addition increases the similarity between stimulus  
754 envelopes, which in turn reduces the neuronal discriminative efficiency based on the envelope  
755 tracking. We thus quantified the acoustic similarity between the stimulus envelopes in the original  
756 condition and in all situations of acoustic degradations. The changes of the between-envelope  
757 similarity in each AM range are presented in figure 7A. In the M and H ranges, the between-  
758 stimulus envelope similarity was low and remained low in all degraded conditions. In contrast,  
759 large changes occurred in the low AM range: in these frequency ranges, the envelope similarity

760 increased progressively with the acoustic degradations. In the following results, we will focus on  
761 this AM range.

762 In the vocoding conditions, the similarity between the four whistle envelopes was relatively  
763 constant, except for the 10-band vocoded condition where this similarity was slightly higher. In the  
764 stationary noise, the four stimulus envelopes became similar and reached a correlation value above  
765 0.8 at the -10 dB SNR condition (which is very close to the maximal value of the acoustic  
766 similarity). In the chorus noise conditions, the four stimulus envelopes remained different (because  
767 spectro-temporal differences were present in the frozen chorus noise) with the highest similarity in  
768 the -10 dB SNR condition.

769 Figure 7B points out that in the condition where the between-stimulus envelope similarity was  
770 higher (at -10 dB SNR in stationary noise), the envelope tracking remained similar (the  $\Delta R_{\max_{E-}}$   
771  $P_{STH}$  values remained stable) whereas the neuronal discrimination decreased compared to the  
772 original condition (most of the  $\Delta MI$  values were largely negative). This clearly demonstrates the  
773 dissociation between changes in  $R_{\max}$  value and those in  $MI$  value. Figure 7C highlights the close  
774 relationship between the acoustic similarity of the four stimulus envelopes and the abilities of  
775 auditory neurons to discriminate between them. Both in subcortical and cortical structures, as the  
776 acoustic distance between the four stimulus envelopes in the low AM range progressively  
777 decreased, the neuronal discrimination decreased (Fig. 7C).

778 Together, these results indicate that it is not a loss in neuronal envelope tracking which leads to a  
779 reduction of the neuronal discriminative abilities in the degraded conditions. Rather, it is the  
780 increase in envelope similarity in situation of acoustic degradations that is one of the important  
781 factors responsible for the decrease in discrimination abilities. Thus, the between-stimulus envelope  
782 similarity in the lower AM range (<20 Hz) can predict the evolution of the discrimination in the  
783 entire auditory system.

784

785 **The increase in between-envelope similarity also correlates with the behavioral performance**  
786 **in noise**

787

788 To examine whether the discrimination performance of auditory neurons might provide a neuronal  
789 basis for behavioral performance, we tested whether behaving animals can discriminate between  
790 whistles when engaged in an operant conditioning task involving the same stimuli. We opted to  
791 train mice in a behavioral task rather than guinea pigs for two main reasons: (1) guinea pigs are  
792 poor and slow learners in instrumental tasks (2) this avoided that the stimuli used for the behavioral  
793 task have innate particular meanings because whistles are alert signals for guinea pigs.

794 The behavioral task was a Go/No-Go task involving the discrimination between two of the four  
795 whistles used in our electrophysiological studies (W1 and W3 see figure 1): Licks to the S+ were  
796 rewarded by a 5 $\mu$ L drop of water and licks to the S- were punished by a 5-second time-out period.  
797 Mice were first trained for 5-10 initial sessions to perform the discrimination in the original  
798 condition until they reached 80% of correct responses for two successive days (N = 9). Then, the  
799 mice were sequentially trained in the stationary noise at the +10, 0 and -10 dB SNR for at least four  
800 sessions. The performance at the last four sessions at each SNR are displayed on figure 7D. For all  
801 mice, the average performance decreased at the 0 and -10 dB SNR, even if two mice were still at  
802 80% of correct performance, the others were slightly above the chance level. In the chorus noise,  
803 the performance of most of the mice were relatively stable, which can be explained by the fact that  
804 acoustically the chorus noise surrounding the two target vocalizations differed between the two  
805 whistles, so that there were more acoustic cues to discriminate between the target stimuli in these  
806 conditions (note that this could also come from the fact that the mice were already extensively  
807 trained to perform the discrimination task in stationary noise when they started the chorus noise).  
808 Despite this pitfall, the main result of this behavioral study was that mice can discriminate the target  
809 vocalizations above chance level even at -10 dB SNR in stationary noise. Furthermore, the decrease  
810 in behavioral performance was strongly related to the reduction of the differences between the two  
811 temporal envelopes in the low AM range (inset in Fig. 7D). These results provide evidence that the  
812 behavioral performance of mice is correlated with the changes in the slow temporal envelope cues.

## 813 **Discussion**

814

815 Our first major result is that the neuronal discrimination performance in the original condition was  
816 correlated with the capacity for tracking the envelopes in the low AM range both for subcortical and  
817 cortical neurons, except in the secondary auditory cortex (VRB) (Fig. 3C). Our second major result  
818 is that, under acoustic degraded conditions and in each structure, the ability for envelope tracking  
819 only slightly changed compared to the original condition (Fig. 5, 6, 7B). Finally, our findings  
820 revealed that the increased similarity between the stimulus envelopes in the low AM range (<20 Hz,  
821 Fig. 7C-D) is one of the important factors responsible for the decrease in neuronal and behavioral  
822 discrimination.

823

### 824 **Slow envelope tracking: a general property of auditory neurons**

825

826 At the level of the auditory nerve, previous studies reported conflicting results concerning the noise  
827 resistance: Frisina and colleagues (1996) found that all AN units partially preserve their AM coding  
828 even in the presence of loud (0 or +6 dB SNR) background noise. However, many others  
829 electrophysiological studies, and the present simulated data, showed a low resistance to noise in the  
830 auditory nerve (Rhode et al., 1978; Geisler and Sinex, 1980; Palmer and Evans, 1982; Costalupes et  
831 al., 1984; Costalupes, 1985; Young and Barta, 1986). This can be explained by several factors  
832 including (i) the type of target stimuli (artificial vs. natural stimuli), (ii) the noise type (stationary  
833 vs. non-stationary noise), (iii) the type of the auditory nerve fibers (low, middle and high SR fibers),  
834 and also the noise levels which have been tested. For example, electrophysiological studies have  
835 shown that low and medium SR fibers with best frequencies around the frequency of a pure tone  
836 exhibited tone-evoked rate changes in presence of a stationary noise at positive SNRs (Rhode et al.,  
837 1978; Geisler and Sinex, 1980; Palmer and Evans, 1982; Costalupes et al., 1984; Costalupes, 1985;  
838 Young and Barta, 1986). High SR fibers, in contrast, exhibited much weaker tone-evoked rate  
839 changes at positive SNRs limited by the high rate response to noise. Thus, as noise level increases,  
840 the discharge rate approaches a fiber's saturation rate and ultimately eliminates the fiber's ability to  
841 respond to tested tones. Low (and middle) SR fibers that have higher thresholds and wider dynamic  
842 range, are significantly more resistant to saturation by high noise levels than high SR fibers.  
843 Therefore, a different ratio between low, middle and high SR fibers could have changed our results  
844 in a way that more sANF responses showed a higher resistance to noise. Here, as we wanted to be  
845 as close as possible to the multi-unit activity recorded in the auditory structures (we assumed that  
846 we recorded about 5 neurons under each electrode), we decided to choose five fibers with a  
847 classical ratio of 1 low SR fiber, 1 medium SR fiber and 3 high SR fibers. However, we should

848 keep in mind that all these previous studies have used tones (or amplitude modulated tones) in noise  
849 at positive SNRs, but natural sounds can potentially trigger more complex encoding as early as the  
850 auditory nerve. Recently, using a similar model of auditory nerve as in the present study,  
851 Rabinowitz and colleagues (2013) showed a poor adaptation to the noise statistics by simulated  
852 fibers when natural environment sounds were used as target stimuli whereas auditory cortex and IC  
853 neurons showed a better adaptation to noise. In our study, we also found a high sensitivity to the  
854 noise for sANF as early as +10 dB SNR in both noises (see examples on Figure 5) and also in the  
855 vocoding conditions potentially due to a higher sensitivity to the spectrotemporal alterations  
856 compared to the other structures (see figure 4A).

857 A few electrophysiological studies have shown that subcortical neurons can display responses very  
858 close to the envelopes of natural stimuli (inferior colliculus: Suta et al., 2003; Rode et al., 2013;  
859 MGB: Tanaka and Taniguchi, 1991; Philibert et al., 2005; Suta et al., 2007). Rode and colleagues  
860 (2013) found that between 15 and 60% of collicular neurons displayed high correlations for at least  
861 one of the three vocalization envelopes, and a subset of collicular neurons even followed the  
862 envelopes of the three guinea pig vocalizations with high correlations ( $>0.85$ ). A similar range of  
863 correlations (between 0.6-0.9) in CNIC was obtained in the present study and, as in their study, we  
864 also did not find a relation between the gammatone filter eliciting the highest  $R_{E-PSTH}$  value and the  
865 best frequency of the neurons.

866 Unlike to other previous cortical studies (Wang et al., 1995; Bar-Yosef et al., 2002; Nagarajan et  
867 al., 2002; Grimsley et al., 2012; Abrams et al., 2017), we filtered envelopes and neuronal responses  
868 in the same frequency bands - from low ( $<20$  Hz) to high (100 and 200 Hz) ranges - to obtain a  
869 direct quantification of the envelope tracking abilities in particular frequency ranges. Furthermore,  
870 we compared the degree of envelope tracking performed by subcortical and cortical neurons in  
871 challenging situations where the envelope is either relatively well preserved or strongly degraded.  
872 Nagarajan and colleagues (2002) found that the synchronization between A1 responses and the  
873 temporal envelope of vocalizations was highly significant and, interestingly, this property was  
874 underestimated based on responses to amplitude-modulated tones. In addition, they pointed out that  
875 A1 responses were quite resistant to spectral degradations (generated by a noise-vocoder) and to  
876 noise addition up to 0 dB SNR. More importantly, the responses were similar when the vocalization  
877 envelope was preserved between 2 and 30 Hz, whereas the responses were strongly reduced when  
878 the envelope was low-pass filtered at 4 or 10 Hz. We confirmed these cortical results on several  
879 aspects: first, the highest correlation coefficients were detected in the lower AM range ( $<20$  Hz) for  
880 each acoustic condition, second, we showed that the envelope tracking ability was little affected by  
881 the presence of noise addition or by the vocoding. We extended these results to a non-primary  
882 cortical area (VRB), to each subcortical level, and even to sANF. Note that the envelope tracking



883 ability is not specific to the processing of conspecific vocalizations: similar results were found with  
884 speech in noise in the auditory cortex of guinea pigs (Abrams et al., 2017).

885 Together, these results highlight that subcortical and cortical auditory neurons maintain their  
886 capacity to track the slow envelope of natural sounds both when they are composed of noise-free  
887 vocalizations or a mixture of noise and vocalizations, suggesting that this property is immutable and  
888 unchanged by the acoustic degradations.

889 In the low AM range ( $<20$  Hz), we noticed a decrease in mean correlation ( $R_{\text{maxE-PSTH}}$ ) values from  
890 midbrain to thalamus to cortex (Fig. 2F) reflecting that the further away from the periphery, the less  
891 precise is the phase-locking ability on the AM cues. For higher AM rates, we expected higher  
892 correlations between the neuronal responses and the envelopes for subcortical structures  
893 (Creutzfeldt et al., 1980; Frisina et al., 1990; Rhode and Greenberg, 1994; Neuert et al., 2001; for  
894 review, Joris et al., 2004). Surprisingly, such a hierarchy was not detected in our results, the mean  
895 correlations in higher AM rates ( $>20$  Hz) being similarly low for each structure including the sANF.  
896 These lower correlation coefficients obtained for the middle- and high-AM frequency bands for all  
897 structures, might result because the envelopes have much lower amplitudes in these bands than in  
898 the low-AM frequency band (see Figure 1C). Another hypothesis is that shorter segments of  
899 neuronal responses could be highly correlated to the higher AM ranges of the envelopes. If so,  
900 reducing the time window on which the correlation is computed should increase the correlations in  
901 the higher AM ranges. We computed the cross-correlation for each whistle (around 300 ms) and  
902 still found low correlations in higher AM ranges (data not shown). This suggests that if higher  
903 correlations exist in higher AM ranges, smaller temporal windows (less than several hundreds of  
904 milliseconds) are required to reveal them. The fact that Abrams and colleagues (2017) found some  
905 residues of the fundamental frequency (between 100-120 Hz, relative to the pitch) in segments of  
906 A1 responses no longer than 100 ms argues in favor of this possibility.

907 The main hypothesis of our study was that the tracking abilities of auditory neurons is one of the  
908 mechanisms explaining the neuronal discrimination. Another possibility is that for higher AM cues,  
909 some auditory neurons respond by increasing their firing rate. This hypothesis relies on the  
910 existence of a rate-place code for periodicity: whereas the temporal tracking abilities decrease along  
911 the auditory pathway, higher periodicities can be encoded by a rate-place code as this has been  
912 demonstrated using amplitude modulated sounds in different species (Langner & Schreiner, 1988;  
913 Schreiner & Langner, 1988; Langner, 1992; Lu et al., 2001a, b; Liang et al., 2002; Lu and Wang,  
914 2004). According to this possibility, neurons increasing their firing rate for coding higher AM cues  
915 should be located at particular locations in IC and auditory cortex (Langner et al., 2009; Schnupp et  
916 al., 2015). However, in order to explain the neuronal discrimination, it seems necessary that each of  
917 the four whistles activated different locations in the periodicity maps of the different auditory

918 structures. As shown in figure 1C, each of the four whistles contained about the same energy in  
919 low, middle and high AM modulations, and as a consequence similar locations should be activated  
920 in these periodicity maps leading to a low discrimination level. Thus, although we cannot discard  
921 this hypothesis, the possibility that the neural discrimination relies on a rate-place code for  
922 particular AM cues seems unlikely. At the cortical level (both in A1 and in VRB), it is also possible  
923 that despite the fact individual neurons cannot keep tracking the detailed envelope fluctuations  
924 (because of their low-pass properties regarding AM cues and their prominent onset responses), they  
925 may, as a large population, track the envelope changes if each neuron is sensitive to a particular rate  
926 of change of the stimulus envelope (a particular rate of transients). Note that according to this  
927 hypothesis, which has been formulated almost twenty years ago (Heil, 2003), this tracking  
928 mechanism would also lose accuracy with increasing levels of background noise.

929

930 **The decrease in neuronal discrimination can be explained by the increase of between-**  
931 **envelopes similarities in the low AM range**

932

933 In the original condition, the better neurons track the slow envelope (<20 Hz), the higher the  
934 neuronal discrimination performance for all structures (Fig. 3B-C). In situations of acoustic  
935 degradation, the envelopes of the original stimuli were altered leading to situations where the  
936 envelopes were mostly dominated by the noise envelopes. However, the three situations of acoustic  
937 degradations used here notably differed. In the tone-vocoder situation, the spectral content is  
938 strongly degraded but the slow temporal envelope is relatively well preserved (Shannon et al., 1995;  
939 Kates, 2011; Souffi et al., 2020). In the chorus noise, there was only a small increase in acoustic  
940 similarity in the low AM range (Fig. 7A) because the chorus noise itself contains strong temporal  
941 variations which differ from one whistle to another. As a consequence of this pitfall, when the  
942 target vocalizations were inserted in the chorus noise, specific regions in the spectro-temporal  
943 domain were dominated by the target vocalizations, while in other regions, it was dominated by the  
944 chorus noise. Consequently, the target vocalizations embedded in the chorus noise generated stimuli  
945 that can be discriminated at all SNRs either based on the vocalization envelopes or based on the  
946 chorus noise envelope itself. In all structures, the neuronal discrimination showed little decrease in  
947 the chorus noise (see Fig. 4) and so was the behavioral performance (Fig. 7D). Only in the  
948 stationary noise, the four slow envelopes became closer as the level of degradation increased (< 20  
949 Hz, see Fig. 7A). This was detrimental for discriminating the vocalizations in noisy conditions in  
950 which envelope tracking become inefficient and worst, can strongly reduce the neuronal  
951 discrimination along the auditory system. Therefore, reducing or increasing the envelope  
952 differences in the low AM range would constrain or facilitate the neuronal discrimination in

953 subcortical and cortical levels. Furthermore, the behavioral performance of mice revealed that they  
954 can discriminate the target vocalizations in quiet (with > 90% correct performance) and can even  
955 discriminate the vocalizations up to 0 dB SNR in stationary noise (with 70-80% correct  
956 performance), suggesting that the between-stimulus envelope differences could explain the  
957 behavioral performance during a discrimination task. Previous studies have reported good  
958 behavioral discrimination performance in conditions of acoustic degradations such as vocoded  
959 consonants or vowels (Ranasinghe et al., 2012a, b), consonants in various levels of background  
960 noise (Shetake et al., 2011), bird songs embedded in stationary and chorus noise (Narayan et al.,  
961 2007) or in broadband dynamic moving ripples (Homma et al., 2020). In all these studies, the  
962 discrimination performance of auditory cortex neurons, based upon spike-timing, has been found to  
963 match relatively well the behavioral performance (Narayan et al., 2007; Ranasinghe et al., 2012a;  
964 Homma et al., 2020) and sometimes even with performance of human subjects (Walker et al.,  
965 2008). Altogether, our results indicate that it is not a loss in neuronal envelope tracking which leads  
966 to a reduction of the neuronal discriminative abilities in the degraded conditions, rather, it is the  
967 direct consequence of the acoustic distance changes between stimulus envelopes.

968

#### 969 **Comparison with human studies: case of newborn infants**

970

971 Speech envelope corresponds to the slow amplitude fluctuations of the signal over time, with peaks  
972 occurring roughly at the syllabic rate. The two pioneer results supporting the view that the envelope  
973 plays a key role in speech comprehension are (1) that comprehension is impaired when the speech  
974 envelope is filtered out (Drullman et al., 1994a, b), and (2) that adult listeners readily understand  
975 degraded speech in which only the envelope is preserved, at least when speech is presented in  
976 silence (Shannon et al., 1995). Additionally, studies have shown that when adults listen to speech,  
977 their neuronal activity synchronized with specific features of the envelope, a phenomenon known as  
978 speech envelope tracking (Ahissar et al., 2001; Luo and Poeppel, 2007; Abrams et al., 2008;  
979 Nourski et al., 2009). Several recent electrophysiological results have provided new insights into  
980 this putative speech envelope tracking mechanism. First, oscillations whose frequency corresponds  
981 to the modulation frequency of the speech envelope (4-5 Hz) have been found to be independent of  
982 comprehension: brain responses in the theta band track the speech envelope even when speech is  
983 time-compressed at a rate that renders it incomprehensible for adult listeners (Zoefel and  
984 VanRullen, 2016; Kösem et al., 2016, 2017; Pefkou et al., 2017). Results from newborns and young  
985 infants have also brought new insights. For example, combining hemodynamic (near-infrared  
986 spectroscopy) and EEG recordings, Cabrera and Gervain (2020) showed that infants (9-10 months  
987 old) detect consonant changes on the basis of envelope cues (without the temporal fine structure)

988 and they can even do so on the basis of the slow temporal variation alone (AM <8 Hz). More  
989 recently, Ortiz-Barajas and colleagues (2021) found that the cortical networks of newborns  
990 (exclusively exposed to French before birth) have the capacity to track the amplitude and the phase  
991 of the speech envelope in their native languages as well as in unfamiliar languages (Spanish and  
992 English). Altogether, these results suggest that amplitude - and phase-tracking take place in the  
993 absence of attention and comprehension.  
994 Thus, envelope tracking can be viewed as a universal mechanism used in all species to discriminate  
995 between communication sounds in a large diversity of acoustic situations ranging from quiet to  
996 adverse, challenging, conditions.

997 **References**

- 998  
 999 Abrams DA, Nicol T, Zecker S, Kraus N (2008) Right-hemisphere auditory cortex is dominant for coding syllable  
 1000 patterns in speech. *J Neurosci.* Apr 9;28(15):3958-65. doi: 10.1523/JNEUROSCI.0187-08.2008.  
 1001 Abrams DA, Nicol T, White-Schwoch T, Zecker S, Kraus N (2017) Population responses in primary auditory cortex  
 1002 simultaneously represent the temporal envelope and periodicity features in natural speech. *Hear Res.*  
 1003 May;348:31-43. doi: 10.1016/j.heares.2017.02.010.  
 1004 Ahissar E, Nagarajan S, Ahissar M, Protopapas A, Mahncke H, & Merzenich MM (2001) Speech comprehension is  
 1005 correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National*  
 1006 *Academy of Sciences of the United States of America*, 98(23), 13367–13372. doi:10.1073/pnas.201400998  
 1007 Anderson LA, Wallace MN & Palmer AR (2007) Identification of subdivisions in the medial geniculate body of the  
 1008 guinea pig. *Hearing Research* 228, 156–167.  
 1009 Aushana Y, Souffi S, Edeline JM, Lorenzi C, Huetz C (2018) Robust Neuronal Discrimination in Primary Auditory  
 1010 Cortex Despite Degradations of Spectro-temporal Acoustic Details: Comparison Between Guinea Pigs with  
 1011 Normal Hearing and Mild Age-Related Hearing Loss. *J. Assoc. Res. Otolaryngol.* 19, 163–180.  
 1012 Bar-Yosef O, Rotman Y, Nelken I (2002) Responses of neurons in cat primary auditory cortex to bird chirps: effects of  
 1013 temporal and spectral context. *J Neurosci.* Oct 1;22(19):8619-32. doi: 10.1523/JNEUROSCI.22-19-  
 1014 08619.2002.  
 1015 Bruce IC, Erfani Y, and Zilany MSA (2018) "A phenomenological model of the synapse between the inner hair cell and  
 1016 auditory nerve: Implications of limited neurotransmitter release sites," *Hearing Research* 360:40–54.  
 1017 Cabrera L, Gervain J (2020) Speech perception at birth: The brain encodes fast and slow temporal information. *Sci*  
 1018 *Adv.* Jul 22;6(30):eaba7830. doi: 10.1126/sciadv.aba7830.  
 1019 Ceballo S, Piwowska Z, Bourg J, Daret A, Bathellier B (2019) Targeted Cortical Manipulation of Auditory Perception.  
 1020 *Neuron.* Dec 18;104(6):1168-1179.e5. doi: 10.1016/j.neuron.2019.09.043.  
 1021 Costalupes JA (1985) Representation of tones in noise in the responses of auditory nerve fibers in cats. I. Comparison  
 1022 with detection thresholds. *J Neurosci.* Dec;5(12):3261-9. doi: 10.1523/JNEUROSCI.05-12-03261.1985.  
 1023 Costalupes JA, Young ED, Gibson DJ (1984) Effects of continuous noise backgrounds on rate response of auditory  
 1024 nerve fibers in cat. *J Neurophysiol.* Jun;51(6):1326-44. doi: 10.1152/jn.1984.51.6.1326.  
 1025 Creutzfeldt O, Hellweg FC, Schreiner C (1980) Thalamocortical transformation of responses to complex auditory  
 1026 stimuli. *Exp Brain Res*;39(1):87-104. doi: 10.1007/BF00237072.  
 1027 Deneux T, Kempf A, Daret A, Ponsot E, Bathellier B (2016) Temporal asymmetries in auditory coding and perception  
 1028 reflect multi-layered nonlinearities. *Nat Commun.* Sep 1;7:12682. doi: 10.1038/ncomms12682.  
 1029 Ding N, Melloni L, Yang A, Wang Y, Zhang W, Poeppel D (2017) Characterizing Neural Entrainment to Hierarchical  
 1030 Linguistic Units using Electroencephalography (EEG). *Front Hum Neurosci.* Sep 28;11:481. doi:  
 1031 10.3389/fnhum.2017.00481.  
 1032 Ding N, Chatterjee M, Simon JZ. (2014) Robust cortical entrainment to the speech envelope relies on the spectro-  
 1033 temporal fine structure. *Neuroimage.* Mar;88:41-6. doi: 10.1016/j.neuroimage.2013.10.054.  
 1034 Drullman R, Festen JM, Plomp R (1994b) Effect of reducing slow temporal modulations on speech reception. *J Acoust*  
 1035 *Soc Am.* May;95(5 Pt 1):2670-80. doi: 10.1121/1.409836.  
 1036 Drullman R, Festen JM & Plomp R (1994a) Effect of temporal envelope smearing on speech reception. *The Journal of*  
 1037 *the Acoustical Society of America* 95, 1053–1064.  
 1038 Edeline JM, Manunta Y, Nodal F, Bajo V (1999) Do auditory responses recorded from awake animals reflect the  
 1039 anatomical parcellation of the auditory thalamus? *Hearing Research*, 131, 135-152.  
 1040 Franke F, Quiñero R, Hierlemann A, Obermayer K (2015) Bayes optimal template matching for spike sorting -  
 1041 combining fisher discriminant analysis with optimal filtering. *J Comput Neurosci.* 38(3):439-59.  
 1042 Frisina RD, Karcich KJ, Tracy TC, Sullivan DM, Walton JP, Colombo J. (1996) Preservation of amplitude modulation  
 1043 coding in the presence of background noise by chinchilla auditory-nerve fibers. *J Acoust Soc Am.*  
 1044 Jan;99(1):475-90. doi: 10.1121/1.414559.  
 1045 Frisina RD, Smith RL, Chamberlain SC (1990) Encoding of amplitude modulation in the gerbil cochlear nucleus: I. A  
 1046 hierarchy of enhancement. *Hear. Res.* 44, 99–122.  
 1047 Gaucher Q, Edeline JM (2015) Stimulus-specific effects of noradrenaline in auditory cortex: implications for the  
 1048 discrimination of communication sounds. *J. Physiol. (Lond.)* 593, 1003–1020.  
 1049 Gaucher Q, Edeline JM, Gourévitch B (2012) How different are the local field potentials and spiking activities? Insights  
 1050 from multi-electrodes arrays. *J. Physiol. Paris* 106, 93–103.

- 1051 Gaucher Q, Huetz C, Gourévitch B, Edeline JM (2013a) Cortical inhibition reduces information redundancy at  
1052 presentation of communication sounds in the primary auditory cortex. *J. Neurosci.* 33, 10713–10728.
- 1053 Geisler CD, Sinex DG (1980) Responses of primary auditory fibers to combined noise and tonal stimuli. *Hear Res.*  
1054 (4):317-34. doi: 10.1016/0378-5955(80)90026-x.
- 1055 Gnansia D, Péan V, Meyer B, Lorenzi C (2009) Effects of spectral smearing and temporal fine structure degradation on  
1056 speech masking release. *J. Acoust. Soc. Am.* 125, 4023–4033.
- 1057 Gnansia D, Pressnitzer D, Péan V, Meyer B, Lorenzi C (2010) Intelligibility of interrupted and interleaved speech for  
1058 normal-hearing listeners and cochlear implantees. *Hearing Research* 265, 46–53.
- 1059 Gourévitch B, Edeline JM (2011) Age-related changes in the guinea pig auditory cortex: relationship with brainstem  
1060 changes and comparison with tone-induced hearing loss. *Eur. J. Neurosci.* 34, 1953–1965.
- 1061 Gourévitch B, Doisy T, Avillac M, Edeline JM (2009) Follow-up of latency and threshold shifts of auditory brainstem  
1062 responses after single and interrupted acoustic trauma in guinea pig. *Brain Res.* 1304, 66–79.
- 1063 Grimsley JM, Palmer AR, Wallace MN (2011) Different representations of tooth chatter and purr call in guinea pig  
1064 auditory cortex. *Neuroreport.* Aug 24;22(12):613-6. doi: 10.1097/WNR.0b013e3283495ae9.
- 1065 Grimsley JMS, Shanbhag SJ, Palmer AR, Wallace MN (2012) Processing of Communication Calls in Guinea Pig  
1066 Auditory Cortex. *PLoS ONE* 7, e51646.
- 1067 Harris KD (2020) Nonsense correlations in neuroscience. *bioRxiv* 2020.11.29.402719; doi:  
1068 <https://doi.org/10.1101/2020.11.29.402719>.
- 1069 Heil P, (2003) Coding of temporal onset envelope in the auditory system. *Speech Commun.* 41 (1), 123–134.
- 1070 Hohmann V (2002) Frequency analysis and synthesis using a Gammatone filterbank. *Acust Acta Acust* 88:433–442.
- 1071 Homma NY, Hullett PW, Atencio CA, Schreiner CE (2020) Auditory Cortical Plasticity Dependent on Environmental  
1072 Noise Statistics. *Cell Rep.* Mar 31;30(13):4445-4458.e5. doi: 10.1016/j.celrep.2020.03.014.
- 1073 Joris PX, Schreiner CE, Rees A (2004) Neural processing of amplitude-modulated sounds. *Physiol. Rev.* 84, 541–577.
- 1074 Kates JM (2011) Spectro-temporal envelope changes caused by temporal fine structure modification. *The Journal of the*  
1075 *Acoustical Society of America* 129, 3981–3990.
- 1076 Kösem A & Van Wassenhove V (2017) Distinct contributions of low-and high-frequency neural oscillations to speech  
1077 comprehension. *Language, Cognition and Neuroscience*, 32(5), 536-544.
- 1078 Kösem A, Basirat A, Azizi L, & van Wassenhove V (2016) High-frequency neural activity predicts word parsing in  
1079 ambiguous speech streams. *Journal of neurophysiology*, 116(6), 2497-2512.
- 1080 Langner G, Schreiner CE (1988) Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. *J*  
1081 *Neurophysiol.* Dec;60(6):1799-822. doi: 10.1152/jn.1988.60.6.1799.
- 1082 Langner G. (1992) Periodicity coding in the auditory system. *Hear Res.* 60(2):115-42. doi: 10.1016/0378-  
1083 5955(92)90015-f.
- 1084 Langner G, Dinse HR, Godde B (2009) A map of periodicity orthogonal to frequency representation in the cat auditory  
1085 cortex. *Front Integr Neurosci.* 3:27. doi: 10.3389/neuro.07.027.2009.
- 1086 Liang L, Lu T, Wang X. (2002) Neural representations of sinusoidal amplitude and frequency modulations in the  
1087 primary auditory cortex of awake primates. *J Neurophysiol.* 87(5):2237-61. doi: 10.1152/jn.2002.87.5.2237.
- 1088 Lu T, Liang L, Wang X. (2001a) Temporal and rate representations of time-varying signals in the auditory cortex of  
1089 awake primates. *Nat Neurosci.*;4(11):1131-8. doi: 10.1038/nn737.
- 1090 Lu T, Liang L, Wang X (2001b) Neural representations of temporally asymmetric stimuli in the auditory cortex of  
1091 awake primates. *J Neurophysiol.* 85(6):2364-80. doi: 10.1152/jn.2001.85.6.2364.
- 1092 Lu T, Wang X (2004) Information content of auditory cortical responses to time-varying acoustic stimuli. *J*  
1093 *Neurophysiol.* 91(1):301-13. doi: 10.1152/jn.00022.2003.
- 1094 Luo H, & Poeppel D (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory  
1095 cortex. *Neuron*, 54(6), 1001–1010. doi:10.1016/j.neuron.2007.06.004
- 1096 Majdak P, Hollomey C, and Baumgartner R (2022) AMT 1.0: the toolbox for reproducible research in auditory  
1097 modeling. *Acta Acustica*.
- 1098 Nagarajan SS et al. (2002) Representation of Spectral and Temporal Envelope of Twitter Vocalizations in Common  
1099 Marmoset Primary Auditory Cortex. *Journal of Neurophysiology* 87, 1723–1737.
- 1100 Narayan, R. et al. Cortical interference effects in the cocktail party problem. *Nature Neuroscience* 10, 1601–1607  
1101 (2007).
- 1102 Neuert V, Pressnitzer D, Patterson RD, Winter IM (2001) The responses of single units in the inferior colliculus of the  
1103 guinea pig to damped and ramped sinusoids. *Hear Res.* Sep;159(1-2):36-52. doi: 10.1016/s0378-  
1104 5955(01)00318-5.

1105 Nourski KV, Reale RA, Oya H, Kawasaki H, Kovach CK, Chen H, Howard MA 3rd, Brugge JF (2009) Temporal  
1106 envelope of time-compressed speech represented in the human auditory cortex. *J Neurosci.* Dec  
1107 9;29(49):15564-74. doi: 10.1523/JNEUROSCI.3065-09.2009.

1108 Occelli F, Suied C, Pressnitzer D, Edeline JM, Gourévitch B (2016) A Neural Substrate for Rapid Timbre Recognition?  
1109 Neural and Behavioral Discrimination of Very Brief Acoustic Vowels. *Cereb. Cortex* 26, 2483–2496.

1110 Ortiz Barajas MC, Guevara R, Gervain J (2021) The origins and development of speech envelope tracking during the  
1111 first months of life. *Dev Cogn Neurosci.* Apr;48:100915. doi: 10.1016/j.dcn.2021.100915.

1112 Osses Vecchi A, Varnet L, Carney LH., Dau T, Bruce IC, Verhulst S, and Majdak P (2022) A comparative study of  
1113 eight human auditory models of monaural processing. *Acta Acustica.* 6, 17.

1114 Palmer AR, Evans EF (1982) Intensity coding in the auditory periphery of the cat: responses of cochlear nerve and  
1115 cochlear nucleus neurons to signals in the presence of bandstop masking noise. *Hear Res.* Aug;7(3):305-23.  
1116 doi: 10.1016/0378-5955(82)90042-9.

1117 Paraouty N, Stasiak A, Lorenzi C, Varnet L, Winter IM (2018) Dual Coding of Frequency Modulation in the Ventral  
1118 Cochlear Nucleus. *J. Neurosci.* 38, 4123–4137.

1119 Patterson RD (1987) A pulse ribbon model of monaural phase perception. *J. Acoust. Soc. Am.* 82, 1560–1586.

1120 Pefkou M, Arnal LH, Fontolan L, Giraud AL. (2017)  $\theta$ -Band and  $\beta$ -Band Neural Activity Reflects Independent Syllable  
1121 Tracking and Comprehension of Time-Compressed Speech. *J Neurosci.* Aug 16;37(33):7930-7938. doi:  
1122 10.1523/JNEUROSCI.2882-16.2017.

1123 Philibert B, Laudanski J, Edeline JM (2005) Auditory thalamus responses to guinea-pig vocalizations: a comparison  
1124 between rat and guinea-pig. *Hear Res.* Nov;209(1-2):97-103. doi: 10.1016/j.heares.2005.07.004.

1125 Pouzat C, Delescluse M, Viot P, Diebolt J (2004) Improved spike-sorting by modeling firing statistics and burst-  
1126 dependent spike amplitude attenuation: a Markov chain Monte Carlo approach. *J Neurophysiol.* 91(6):2910-  
1127 28.

1128 Quiroga RQ, Nadasdy Z, Ben-Shaul Y (2004) Unsupervised spike detection and sorting with wavelets and  
1129 superparamagnetic clustering. *Neural Comput.* 16(8):1661-87.

1130 Rabinowitz NC, Willmore BD, King AJ, Schnupp JW (2013) Constructing noise-invariant representations of sound in  
1131 the auditory pathway. *PLoS Biol.* Nov;11(11):e1001710. doi: 10.1371/journal.pbio.1001710.

1132 Ramachandran R, Davis KA, May BJ (2000) Rate representation of tones in noise in the inferior colliculus of  
1133 decerebrate cats. *J Assoc Res Otolaryngol.* 1(2):144-60. doi: 10.1007/s101620010029.

1134 Ranasinghe KG, Vrana WA, Matney CJ, Kilgard MP (2012b) Neural Mechanisms Supporting Robust Discrimination of  
1135 Spectrally and Temporally Degraded Speech. *Journal of the Association for Research in Otolaryngology* 13,  
1136 527–542.

1137 Ranasinghe KG, Carraway RS, Borland MS, Moreno NA, Hanacik EA, Miller RS, Kilgard MP (2012a) Speech  
1138 discrimination after early exposure to pulsed-noise or speech. *Hear Res.* Jul;289(1-2):1-12. doi:  
1139 10.1016/j.heares.2012.04.020.

1140 Redies H, Brandner S, Creutzfeldt OD (1989) Anatomy of the auditory thalamocortical system of the guinea pig. *The*  
1141 *Journal of Comparative Neurology* 282, 489–511.

1142 Reiss LA, Ramachandran R, May BJ (2011) Effects of signal level and background noise on spectral representations in  
1143 the auditory nerve of the domestic cat. *J Assoc Res Otolaryngol.* 12(1):71-88. doi: 10.1007/s10162-010-  
1144 0232-5.

1145 Rhode WS, Greenberg S (1994b) Encoding of amplitude modulation in the cochlear nucleus of the cat. *J. Neurophysiol.*  
1146 71, 1797–1825.

1147 Rhode WS, Geisler CD, Kennedy DT (1978) Auditory nerve fiber response to wide-band noise and tone combinations.  
1148 *J Neurophysiol.* May;41(3):692-704. doi: 10.1152/jn.1978.41.3.692.

1149 Rode T, Hartmann T, Hubka P, Scheper V, Lenarz M, Lenarz T, Kral A, Lim HH (2013) Neural representation in the  
1150 auditory midbrain of the envelope of vocalizations based on a peripheral ear model. *Front Neural Circuits.*  
1151 Oct 21;7:166. doi: 10.3389/fncir.2013.00166.

1152 Rosen S (1992) Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos Trans R Soc Lond B*  
1153 *Biol Sci.* Jun 29;336(1278):367-73. doi: 10.1098/rstb.1992.0070.

1154 Rutkowski RG, Shackleton TM, Schnupp JWH, Wallace MN, Palmer AR (2002) Spectrotemporal Receptive Field  
1155 Properties of Single Units in the Primary, Dorsocaudal and Ventrostral Auditory Cortex of the Guinea Pig.  
1156 *Audiology and Neurotology* 7, 214–227.

1157 Sayles M, Winter IM (2010) Equivalent-rectangular bandwidth of single units in the anaesthetized guinea-pig ventral  
1158 cochlear nucleus. *Hear. Res.* 262, 26–33.

- 1159 Shamma S, Lorenzi C (2013) On the balance of envelope and temporal fine structure in the encoding of speech in the  
1160 early auditory system. *J Acoust Soc Am.* May;133(5):2818-33. doi: 10.1121/1.4795783.
- 1161 Shannon RV, Zeng FG, Kamath V, Wygonski J & Ekelid M (1995) Speech Recognition with Primarily Temporal Cues.  
1162 *Science* 270, 303–304.
- 1163 Shetake JA, Wolf JT, Cheung RJ, Engineer CT, Ram SK, Kilgard MP (2011) Cortical activity patterns predict robust  
1164 speech discrimination ability in noise. *Eur J Neurosci.* Dec;34(11):1823-38. doi: 10.1111/j.1460-  
1165 9568.2011.07887.x.
- 1166 Schreiner CE, Langner G (1988) Periodicity coding in the inferior colliculus of the cat. II. Topographical organization.  
1167 *J Neurophysiol.* 60(6):1823-40. doi: 10.1152/jn.1988.60.6.1823.
- 1168 Schnupp JW, Garcia-Lazaro JA, Lesica NA (2015) Periodotopy in the gerbil inferior colliculus: local clustering rather  
1169 than a gradient map. *Front Neural Circuits.*;9:37. doi: 10.3389/fncir.2015.00037.
- 1170 Schnupp JWH, Hall TM, Kokelaar RF, Ahmed B (2006) Plasticity of temporal pattern codes for vocalization stimuli in  
1171 primary auditory cortex. *J. Neurosci.* 26, 4785–4795.
- 1172 Souffi S, Lorenzi C, Huetz C, Edeline JM (2021) Robustness to Noise in the Auditory System: A Distributed and  
1173 Predictable Property. *eNeuro.* Mar 18;8(2):ENEURO.0043-21.2021. doi: 10.1523/ENEURO.0043-21.2021.
- 1174 Souffi S, Lorenzi C, Varnet L, Huetz C, Edeline JM (2020) Noise-Sensitive But More Precise Subcortical  
1175 Representations Coexist with Robust Cortical Encoding of Natural Vocalizations. *J Neurosci.* Jul  
1176 1;40(27):5228-5246. doi: 10.1523/JNEUROSCI.2731-19.2020..
- 1177 Stone MA, Füllgrabe C, Mackinnon RC, Moore BCJ (2011) The importance for speech intelligibility of random  
1178 fluctuations in ‘steady’ background noise. *J. Acoust. Soc. Am.* 130, 2874–2881.
- 1179 Suta D, Popelár J, Kvasnák E, Syka J (2007) Representation of species-specific vocalizations in the medial geniculate  
1180 body of the guinea pig. *Exp Brain Res.* Nov;183(3):377-88. doi: 10.1007/s00221-007-1056-3.
- 1181 Suta D, Kvasnák E, Popelár J, Syka J (2003) Representation of species-specific vocalizations in the inferior colliculus  
1182 of the guinea pig. *J Neurophysiol.* Dec;90(6):3794-808. doi: 10.1152/jn.01175.2002.
- 1183 Tanaka H, Taniguchi I (1991) Responses of medial geniculate neurons to species-specific vocalized sounds in the  
1184 guinea pig. *Jpn J Physiol.*41(6):817-29. doi: 10.2170/jjphysiol.41.817.
- 1185 Ter-Mikaelian M, Semple MN, Sanes DH (2013) Effects of spectral and temporal disruption on cortical encoding of  
1186 gerbil vocalizations. *Journal of Neurophysiology* 110, 1190–1204.
- 1187 Varnet L, Ortiz-Barajas MC, Erra RG, Gervain J, Lorenzi C (2017) A cross-linguistic study of speech modulation  
1188 spectra. *The Journal of the Acoustical Society of America* 142, 1976–1989.
- 1189 Walker KM, Ahmed B, Schnupp JW (2008) Linking cortical spike pattern codes to auditory perception. *J Cogn  
1190 Neurosci.* Jan;20(1):135-52. doi: 10.1162/jocn.2008.20012.
- 1191 Wallace MN, Palmer AR (2008) Laminar differences in the response properties of cells in the primary auditory cortex.  
1192 *Exp Brain Res* 184, 179–191.
- 1193 Wallace MN, Shackleton TM, Anderson LA, Palmer AR (2005) Representation of the purr call in the guinea pig  
1194 primary auditory cortex. *Hear Res.* Jun;204(1-2):115-26. doi: 10.1016/j.heares.2005.01.007.
- 1195 Wallace MN, Rutkowski RG, Palmer AR (2000) Identification and localisation of auditory areas in guinea pig cortex.  
1196 *Experimental Brain Research* 132, 445–456.
- 1197 Wallace MN, Palmer AR (2009) Functional subdivisions in low-frequency primary auditory cortex (AI) *Exp Brain Res.*  
1198 *Apr*;194(3):395-408. doi: 10.1007/s00221-009-1714-8.
- 1199 Wallace MN, Anderson LA, Palmer AR (2007) Phase-Locked Responses to Pure Tones in the Auditory Thalamus.  
1200 *Journal of Neurophysiology* 98, 1941–1952.
- 1201 Wang X, Merzenich MM, Beitel R, Schreiner CE (1995) Representation of a species-specific vocalization in the  
1202 primary auditory cortex of the common marmoset: temporal and spectral characteristics. *Journal of  
1203 Neurophysiology* 74, 2685–2706.
- 1204 Woolley SM, Gill PR, Theunissen FE (2006) Stimulus-dependent auditory tuning results in synchronous population  
1205 coding of vocalizations in the songbird midbrain. *J Neurosci.* Mar 1;26(9):2499-512. doi:  
1206 10.1523/JNEUROSCI.3731-05.2006.
- 1207 Young ED, Barta PE (1986) Rate responses of auditory nerve fibers to tones in noise near masked threshold. *J Acoust  
1208 Soc Am.* Feb;79(2):426-42. doi: 10.1121/1.393530.
- 1209 Zeng FG, Nie K, Stickney GS, Kong YY, Vongphoe M, Bhargava A, Wei C, Cao K. (2005) Speech recognition with  
1210 amplitude and frequency modulations. *Proc. Natl. Acad. Sci. USA*, 102, 2293-2298.
- 1211 Zoefel B, VanRullen R (2016) EEG oscillations entrain their phase to high-level features of speech sound. *Neuroimage.*  
1212 *Jan 1;124(Pt A):16-23.* doi: 10.1016/j.neuroimage.2015.08.054.



1213 **Additional information section**

1214

1215 **Data availability statement**

1216 The data that support the findings of this study are available from the corresponding author upon  
1217 reasonable request.

1218

1219 **Competing interests**

1220 The authors declare no competing financial interests.

1221

1222 **Author contributions**

1223 This work was performed at the Paris-Saclay Institute of Neurosciences (NeuroPSI). S.S. and J-  
1224 M.E. designed the experiments. S.S. and M.Z. performed the experiments. S.S., L.V., B.B. and C.H.  
1225 have participated in the acquisition, analysis or interpretation of data. All authors drafting the work  
1226 or revising it critically for important intellectual content. All authors approved the final version of  
1227 the manuscript. All authors agree to be accountable for all aspects of the work in ensuring that  
1228 questions related to the accuracy or integrity of any part of the work are appropriately investigated  
1229 and resolved; and all persons designated as authors qualify for authorship, and all those who qualify  
1230 for authorship are listed.

1231

1232 **Funding**

1233 JME was supported by grants from the French Agence Nationale de la Recherche (ANR) (ANR-14-  
1234 CE30-0019-01). SS was supported by the Fondation pour la Recherche Médicale (FRM) grant  
1235 number ECO20160736099 and by the Entendre Foundation.

1236

1237 **Acknowledgements**

1238 We thank Nihaad Paraouty for training us on the cochlear nucleus surgery. We thank Prf. Christian  
1239 Lorenzi and Prf. Shihab Shamma for helpful comments on a previous version of the MS, and are  
1240 particularly grateful to Dr. Virginie van Wassenhove for suggesting several improvements of the  
1241 MS. We also wish to thank Céline Dubois, Mélanie Dumont and Aurélie Bonilla, for taking care of  
1242 the guinea-pig colony.

1243

1244

## 1245 **Figure legends**

1246

1247 **Figure 1. Overall and filtered envelopes in three amplitude modulation ranges.**

1248 **A.** Spectrograms of original and degraded stimuli.

1249 **B.** Overall envelopes of original and degraded stimuli. The envelopes of the four original whistles  
1250 are presented on the left panel. Two whistles were used for a Go/No-Go behavioral discrimination  
1251 task (see Fig. 7D): whistle 1 as the “Go or S+” stimulus and whistle 3 as the “No-Go or S-”  
1252 stimulus. From left to right, the four envelopes of these stimuli are presented first, in the vocoding  
1253 conditions (with 38, 20 and 10 frequency bands from top to bottom), then in stationary noise (at  
1254 +10, 0 and -10 dB SNR from top to bottom) and in chorus noise conditions (at +10, 0 and -10 dB  
1255 SNR from top to bottom).

1256 **C.** Examples of the filtered envelopes for the original vocalizations using a bank of 35 gammatone  
1257 filters with center frequencies uniformly spaced along a guinea pig - adapted ERB (equivalent  
1258 rectangular bandwidth) scale ranging from 20 to 30 000 Hz. Three ranges of amplitude modulation  
1259 (AM) have been investigated here: the low (<20 Hz), middle (between 20 and 100 Hz) and high  
1260 (between 100 and 200 Hz) AM ranges. The red curves indicate the seven filtered envelopes selected  
1261 along the signal for the subsequent analyses.  
1262

1263 **Figure 2. Correlations in the original condition between peri-stimulus time histograms**  
1264 **(PSTHs) of subcortical and cortical recordings and the stimulus envelope both filtered in the**  
1265 **three selected AM ranges.**

1266 **A.** Individual examples of original PSTHs obtained in each structure (from bottom to top, sANF:  
1267 simulated auditory nerve fibers, CN: cochlear nucleus, CNIC: central nucleus of the inferior  
1268 colliculus, MGv: ventral division of the medial geniculate, A1: primary auditory cortex, VRB:  
1269 ventro-rostral belt).

1270 **B.** Population responses ranked from the lowest to the highest best frequencies with the color code  
1271 representing the normalized firing rate. On the bottom of each panel, the population firing rate  
1272 represents the instantaneous summed activity of the whole virtual population, and on the right, the  
1273 total firing rate along the different best frequencies.

1274 **C-E.** Examples of correlations between the PSTH (in black) and the envelope (in red). In each  
1275 panel, the PSTHs and the stimulus envelopes are filtered in the same frequency range. For each  
1276 recording, the correlation value between the PSTH and the envelope is shown on the top left. In a  
1277 given AM range, the stimulus envelopes differ between examples because we selected the  
1278 gammatone envelope (out of seven gammatones) which induced the highest correlation. Note that  
1279 the PSTHs are not lagged compared to the envelopes as during the analysis.

1280 **F.** Box plots showing the distributions of the  $R_{\max_{E-PSTH}}$  values for the six auditory structures  
1281 (sANF to VRB) in the three AM ranges. The red dots in the box plots correspond to the mean  
1282  $R_{\max_{E-PSTH}}$  values and the boxes correspond to the interquartile ranges. Note the higher  $R_{\max_{E-PSTH}}$   
1283 values in the low (L) AM range compared with the middle (M) and high (H) AM ranges. The black  
1284 lines represent significant differences between the mean  $R_{\max_{E-PSTH}}$  values. In the low AM range,  
1285 sANF, CN and CNIC recordings displayed significantly higher mean  $R_{\max_{E-PSTH}}$  values than MGv  
1286 and cortical recordings (low AM range: one-way ANOVA,  $p < 0.0001$   $F_{(5, 1165)} = 138.25$  with  
1287 Students' t test unpaired, sANF vs. MGv, A1 or VRB  $p < 0.0001$ , CN vs. MGv, A1 or VRB  
1288  $p < 0.0001$ , CNIC vs. MGv, A1 or VRB  $p < 0.0001$ ; mean ( $\pm$  STD)  $R_{\max_{E-PSTH}}$  values:  $R_{sANF(n=217)} =$   
1289  $0.81 \pm 0.03$ ,  $R_{CN(n=336)} = 0.80 \pm 0.08$ ,  $R_{CNIC(n=274)} = 0.78 \pm 0.11$ ,  $R_{MGv(n=102)} = 0.68 \pm 0.13$ ,  $R_{A1(n=}$   
1290  $171)} = 0.60 \pm 0.13$  and  $R_{VRB(n=66)} = 0.63 \pm 0.14$ ). In the middle and high AM ranges, the CNIC  
1291 recordings still exhibited slightly higher mean  $R_{\max_{E-PSTH}}$  values compared to the other structures  
1292 (mean ( $\pm$  STD)  $R_{\max_{E-PSTH}}$  values in the middle AM range:  $R_{sANF(n=44)} = 0.32 \pm 0.03$ ,  $R_{CN(n=285)} =$   
1293  $0.31 \pm 0.07$ ,  $R_{CNIC(n=285)} = 0.34 \pm 0.07$ ,  $R_{MGv(n=133)} = 0.31 \pm 0.07$ ,  $R_{A1(n=220)} = 0.27 \pm 0.05$  and  $R_{VRB(n=}$   
1294  $85)} = 0.29 \pm 0.06$ ; mean ( $\pm$  STD)  $R_{\max_{E-PSTH}}$  values in the high AM range:  $R_{sANF(n=77)} = 0.31 \pm 0.03$ ,  
1295  $R_{CN(n=249)} = 0.31 \pm 0.05$ ,  $R_{CNIC(n=257)} = 0.33 \pm 0.05$ ,  $R_{MGv(n=97)} = 0.29 \pm 0.04$ ,  $R_{A1(n=196)} = 0.27 \pm 0.05$

1296 and  $R_{VRB(n=50)} = 0.30 \pm 0.05$ ). Note that n values correspond to the selected simulations or  
1297 recordings.

1298

1299 **Figure 3. In the original condition, the better cortical and subcortical neurons track the slow**  
1300 **envelope (<20 Hz), the higher the value of mutual information.**

1301 **A.** Box plots showing the distributions of the MI values obtained in the six levels of the auditory  
1302 system in the original condition. The red dots in the box plots correspond to the mean MI values  
1303 and the boxes correspond to the interquartile ranges. Note the lower significant values obtained at  
1304 the cortical level in A1 and VRB compared to those obtained in sANF and subcortical structure  
1305 (one-way ANOVA  $p < 0.0001$   $F_{(5, 1538)} = 266.46$  with Students' t test unpaired, sANF vs. A1 or  
1306 VRB  $p < 0.0001$ , CN vs. A1 or VRB  $p < 0.0001$ , CNIC vs. A1 or VRB  $p < 0.0001$ , MGv vs. A1 or  
1307 VRB  $p < 0.0001$ ; mean ( $\pm$  STD) MI values:  $MI_{sANF(n=77)} = 1.84 \pm 0.21$  bits,  $MI_{CN(n=249)} = 0.92 \pm$   
1308  $0.47$  bits,  $MI_{CNIC(n=257)} = 1 \pm 0.5$  bits,  $MI_{MGv(n=97)} = 1.19 \pm 0.55$  bits,  $MI_{A1(n=196)} = 0.68 \pm 0.37$  bits  
1309 and  $R_{VRB(n=50)} = 0.55 \pm 0.29$  bits). Note that n values correspond to the selected simulations or  
1310 recordings. The black lines represent the first significant differences between the mean values. Note  
1311 that for the sake of clarity, not all significant differences are indicated by black lines. For example,  
1312 the difference between sANF and MGv was the first one to be significant but it was also significant  
1313 between sANF and the two cortical areas A1 and VRB.

1314 **B.** Scattergrams showing the relationships between  $R_{maxE-PSTH}$  and MI values for the six structures  
1315 in the three AM ranges. Black lines correspond to the linear regression lines.

1316 **C.** Matrix summarizing the correlation coefficients between  $R_{maxE-PSTH}$  and MI in each structure  
1317 and AM range. The values in red indicate that the correlation was significant. In all but one case (in  
1318 CNIC in the middle AM range), significant positive correlations between  $R_{maxE-PSTH}$  and MI  
1319 values were obtained in all AM ranges in subcortical structures ( $p^L_{CN} < 0.0001$ ,  $p^M_{CN} < 0.0001$ ,  
1320  $p^H_{CN} = 0.01$ ,  $p^L_{CNIC} < 0.0001$ ,  $p^M_{CNIC} = 0.24$ ,  $p^H_{CNIC} = 0.005$ ,  $p^L_{MGv} = 0.006$ ,  $p^M_{MGv} < 0.0001$ ,  $p^H_{MGv} =$   
1321  $0.01$ ). For the sANF, the range of MI values was too limited to compute reliable correlations (most  
1322 MI values were close to the maximum of 2 bits). At the subcortical level, the highest correlation  
1323 values between  $R_{maxE-PSTH}$  and MI values were found in MGv. At the cortical level, significant  
1324 correlations between  $R_{maxE-PSTH}$  and MI values were detected in the low and middle AM ranges in  
1325 A1 ( $p^L_{A1} = 0.01$ ,  $p^M_{A1} = 0.05$ ,  $p^H_{A1} = 0.32$ ) and there was no significant correlation in VRB ( $p^L_{VRB} =$   
1326  $0.31$ ,  $p^M_{VRB} = 0.51$ ,  $p^H_{VRB} = 0.99$ ).

1327

1328 **Figure 4. Neuronal discrimination performance in all degraded conditions along the auditory**  
1329 **system.**

1330 **A-F.** Neuronal discrimination performance (quantified by the mutual information, in bits) in  
1331 original condition and in the three situations of acoustic degradation (top panels: vocoding, middle  
1332 panels: stationary noise and bottom panels: chorus noise). In each box plot, the horizontal line  
1333 corresponds to the median value and the boxes correspond to the interquartile ranges. For all  
1334 structures except in sANF, note the largest decrease in MI value in the stationary noise in the  
1335 subcortical structures compared with the relative stability of these values in vocoding and chorus  
1336 noise. Note also the much smaller decreases observed at the cortical level in the three situations of  
1337 acoustic alterations. The stars represent the first significant differences between the mean original  
1338 values and those obtained in degraded conditions. Note that for the sake of clarity, not all  
1339 significant differences are indicated by stars. The decrease was significant only for the 10-band  
1340 vocoded vocalizations in MGv and A1 (Fig. 4D, one-way ANOVA,  $p = 0.05$   $F_{(3, 811)} = 2.58$  with  
1341 Students' t test paired,  $MGv_{Ori}$  vs.  $MGv_{Voc10}$   $p < 0.0001$ ; Fig. 4E, one-way ANOVA,  $p = 0.001$   $F_{(3,$   
1342  $722)} = 3.73$  with Students' t test paired,  $A1_{Ori}$  vs.  $A1_{Voc10}$   $p < 0.0001$ ), and no significant difference  
1343 was detected in VRB (Fig. 4F, one-way ANOVA,  $p = 0.75$   $F_{(3, 186)} = 0.41$ ). The decrease was already  
1344 significant with 38-band vocoded vocalizations in sANF or with 20-band vocoded vocalizations in  
1345 CN and CNIC (Fig. 4A, one-way ANOVA,  $p < 0.0001$   $F_{(3, 1302)} = 111.3$  with Students' t test paired,  
1346  $sANF_{Ori}$  vs.  $sANF_{Voc38}$   $p < 0.0001$ ; Fig. 4B, one-way ANOVA,  $p < 0.0001$   $F_{(3, 1424)} = 12.42$  with  
1347 Students' t test paired,  $CN_{Ori}$  vs.  $CN_{Voc20}$   $p < 0.0001$ ; Fig. 4C, one-way ANOVA,  $p < 0.0001$   $F_{(3,$

1231) = 13.17 with Students' t test paired, CNIC<sub>Ori</sub> vs. CNIC<sub>Voc20</sub>  $p < 0.0001$ ). Note that there was also a significant increase in MI values with the 38-band vocoded vocalizations in CN (Fig.4B, one-way ANOVA,  $p < 0.0001$   $F_{(3, 1424)} = 12.42$  with Students' t test paired, CN<sub>Ori</sub> vs. CN<sub>Voc38</sub>  $p = 0.0073$ ). In chorus noise, in CN and CNIC, there was no significant decrease in mean MI values (Fig. 4B, one-way ANOVA,  $p = 0.05$   $F_{(3, 1176)} = 2.65$ ; Fig. 4C, one-way ANOVA,  $p = 0.36$   $F_{(3, 1188)} = 1.06$ ), whereas in sANF and MGv, the mean MI values significantly decreased at +10 dB or 0 dB SNR respectively (Fig. 4A, one-way ANOVA,  $p < 0.0001$   $F_{(3, 1331)} = 232.86$  with Students' t test paired, sANF<sub>Ori</sub> vs. sANF<sub>+10dB</sub>  $p < 0.0001$ ; Fig. 4D, one-way ANOVA,  $p < 0.0001$   $F_{(3, 753)} = 7.3$  with Students' t test paired, MGv<sub>Ori</sub> vs. MGv<sub>0dB</sub>  $p < 0.0001$ ). At the cortical level, there was a significant decrease in A1 at 0 dB SNR and no significant change of mean MI values in VRB (Fig. 4E, one-way ANOVA,  $p = 0.0039$   $F_{(3, 697)} = 4.5$  with Students' t test paired, A1<sub>Ori</sub> vs. A1<sub>0dB</sub>  $p < 0.0001$ ; Fig.4F, one-way ANOVA,  $p = 0.31$   $F_{(3, 179)} = 1.19$ ). In stationary noise, the mean MI value in sANF, CN and MGv was significantly reduced already at +10 dB SNR (Fig. 4A, one-way ANOVA,  $p < 0.0001$   $F_{(3, 1153)} = 767.64$  with Students' t test paired, sANF<sub>Ori</sub> vs. sANF<sub>+10dB</sub>  $p < 0.0001$ ; Fig. 4B, one-way ANOVA,  $p < 0.0001$   $F_{(3, 812)} = 61.22$  with Students' t test paired, CN<sub>Ori</sub> vs. CN<sub>+10dB</sub>  $p < 0.0001$ ; Fig. 4D, one-way ANOVA,  $p < 0.0001$   $F_{(3, 630)} = 62.03$  with Students' t test paired, MGv<sub>Ori</sub> vs. MGv<sub>+10dB</sub>  $p < 0.0001$ ), whereas the mean MI value in CNIC was significantly reduced at 0 dB SNR (Fig. 4C, one-way ANOVA,  $p < 0.0001$   $F_{(3, 1078)} = 32.08$  with Students' t test paired, CNIC<sub>Ori</sub> vs. CNIC<sub>0dB</sub>  $p < 0.0001$ ). At the cortical level, stationary noise significantly reduced the mean MI value in A1 only at -10 dB SNR (Fig. 4E, one-way ANOVA,  $p < 0.0001$   $F_{(3, 669)} = 13.99$  with Students' t test paired, A1<sub>Ori</sub> vs. A1<sub>-10dB</sub>  $p < 0.0001$ ), whereas the mean MI values in VRB remained unchanged in all conditions (Fig. 4F, one-way ANOVA,  $p = 0.26$   $F_{(3, 164)} = 1.34$ ).

1371 **Figure 5. Individual examples of correlations between neuronal responses and envelopes in all**  
 1372 **acoustic conditions and all structures.**

1373 Note that for the subcortical and cortical structures, we presented the results from the low AM  
 1374 range. The correlation value between the PSTH (in black) and the envelope (in red) is shown on the  
 1375 left. In all structures, the correlation values for these individual recordings remained similar  
 1376 between the acoustic conditions.  
 1377

1378 **Figure 6. The envelope tracking is only slightly affected by the different situations of acoustic**  
 1379 **degradation.**

1380 Mean changes of the  $R_{\max_{E-PSTH}}$  ( $\pm$ STD) quantified from the original condition in the different  
 1381 situations of acoustic degradations (vocoding, stationary noise (SN) and chorus noise (CN)) for all  
 1382 recordings obtained in the six structures. Note that, in each structure and AM range, the  $R_{\max_{E-PSTH}}$   
 1383 values were only slightly changed between the original and the degraded conditions. In sANF and  
 1384 CN, we observed a maximal increase in mean ( $\pm$  STD)  $R_{\max_{E-PSTH}}$  values of 0.16 ( $\pm$ 0.03) (and 0.11  
 1385 ( $\pm$ 0.13) for CN) and, a maximal decrease of 0.10 ( $\pm$ 0.04) (and 0.05 ( $\pm$ 0.06) for CN) depending on  
 1386 the degraded conditions and the AM range (Fig. 6). In CNIC and MGv, the mean ( $\pm$  STD)  $R_{\max_{E-PSTH}}$   
 1387 changes in degraded conditions were very small (Fig. 6, between -0.06 ( $\pm$ 0.03) and 0.006  
 1388 ( $\pm$ 0.06) for CNIC and between -0.08 ( $\pm$ 0.08) and -0.0002 ( $\pm$ 0.15) for MGv). In A1, the changes in  
 1389 mean ( $\pm$  STD)  $R_{\max_{E-PSTH}}$  values varied between -0.09 ( $\pm$ 0.11) and 0.07 ( $\pm$ 0.15) and in VRB it  
 1390 varied between -0.13 ( $\pm$ 0.13) and 0.07 ( $\pm$  0.15).  
 1391  
 1392

1393 **Figure 7. In all situations of acoustic alteration, the decrease in neuronal discrimination**  
 1394 **performance can be explained by the increase in envelope similarity in the low range.**

1395 **A.** Acoustic similarity ( $R_{Env}$ ) between the envelopes of the four whistles in the original condition  
 1396 (Ori) and in the three situations of acoustic alterations (vocoding, stationary noise and chorus noise)  
 1397 for the low (L, red lines), middle (M, yellow lines) and high (H, purple lines) AM ranges. Dark  
 1398 lines correspond to the  $R_{Env}$  values based on the 7 selected gammatones, whereas the light lines

1399 correspond to the  $R_{Env}$  values based on the 35 gammatones. Note that in the stationary noise, the  
1400 correlation between the stimulus envelopes largely increased in the L range, indicating that the  
1401 stimuli tended to be similar to each other in these AM ranges, which was not the case in the middle  
1402 and high ranges (M and H). This between-stimuli increase in correlation in the L range was much  
1403 weaker in the vocoding and chorus noise situations.

1404 **B.** Scattergrams showing the variation of the maximal correlation ( $\Delta R_{maxE-PSTH}$ ) in the low AM  
1405 range as a function of the variation of MI ( $\Delta MI$ ) in the -10 dB SNR condition compared to the  
1406 original condition in each structure.

1407 **C.** Mean changes ( $\Delta MI$ , in percentage) of mutual information in sANF, CN, CNIC MG<sub>v</sub>, A1 and  
1408 VRB as a function of the variation ( $\Delta R_{Env}$ , in percentage) of the acoustic similarity in low AM  
1409 range relative to the original condition. Each dot represents neuronal data ( $\Delta MI$ ) in sANF (in dark  
1410 red), CN (in black), CNIC (in green), MG<sub>v</sub> (in orange), A1 (in blue) and VRB (in purple). From left  
1411 to right, all degraded acoustic conditions were organized according to the acoustic distance of the  
1412 envelopes ( $R_{Env}$ ) between the four whistles quantified on Figure 7A (+10 dB SNR - Chorus N., Voc  
1413 38, Voc20, +10 dB SNR - Stationary N., 0 dB SNR - Chorus N., Voc10, -10 dB SNR - Chorus N., 0  
1414 dB SNR - Stationary N., -10 dB SNR - Stationary N.). Linear fits were generated for the different  
1415 structures across all degraded conditions (color lines). For the sake of clarity, we did not use an  
1416 orthonormal coordinate system.

1417 **D.** Percentage of correct responses obtained during the four last sessions for each condition. The  
1418 dark thick line corresponds to the mean ( $\pm$ STD) values obtained for all mice. The individual  
1419 performances of each mouse (n=9) are presented by the grey thin lines. The last four sessions of  
1420 discrimination in the original conditions are represented followed by the discrimination in the three  
1421 conditions in stationary noise (+10, 0 and -10 dB SNR), followed by the discrimination in the three  
1422 conditions in chorus noise (+10, 0 and -10 dB SNR). The chance level is represented by the red  
1423 dashed line. The drops in performance were observed for the 0 dB and the -10 dB SNR in the  
1424 stationary noise. The inset shows that the decrease in behavioral performance (average across  
1425 sessions and animals) was strongly related to the reduction in the differences between the two  
1426 temporal envelopes (W1 and W3) in the low AM range.

1427

1428

1429

### 1430 **Abstract figure legend**

1431 *Methods:* We simulated auditory nerve fiber (sANF) responses and recorded the neuronal activity in  
1432 five auditory structures (from cochlear nucleus to secondary auditory cortex) in response to four  
1433 vocalizations presented in quiet and in two types of noise (a stationary and a chorus noise at three  
1434 SNRs: +10, 0 and -10 dB). In addition, we tested whether behaving animals can discriminate  
1435 between whistles when engaged in a Go/No-Go task involving the discrimination between two of  
1436 the four whistles used in our electrophysiological studies (W1 and W3). Licks to the S+ were  
1437 rewarded by a drop of water and licks to the S- were punished by a 5-second time-out period.

1438 *Results:* Subcortical and cortical auditory neurons track the slow changes of the temporal envelope  
1439 (<20Hz), with a high degree of fidelity in the original (positively correlated with the neuronal  
1440 discrimination) and degraded conditions. Our results demonstrate that the between-stimulus  
1441 envelope similarity, which increases in noise, negatively correlates both with the neuronal  
1442 discrimination and the behavioral performance.

1443

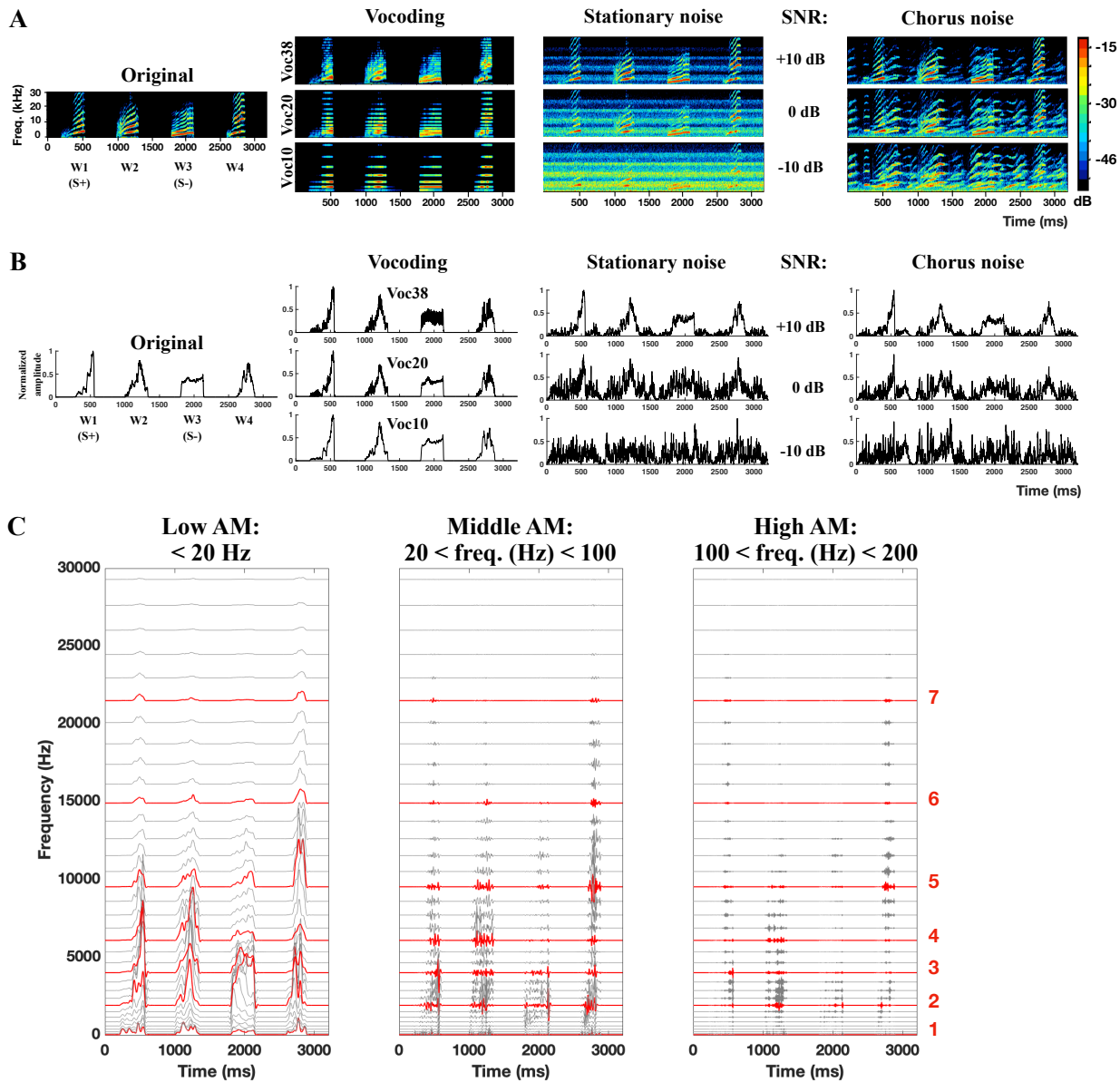


Figure 1.

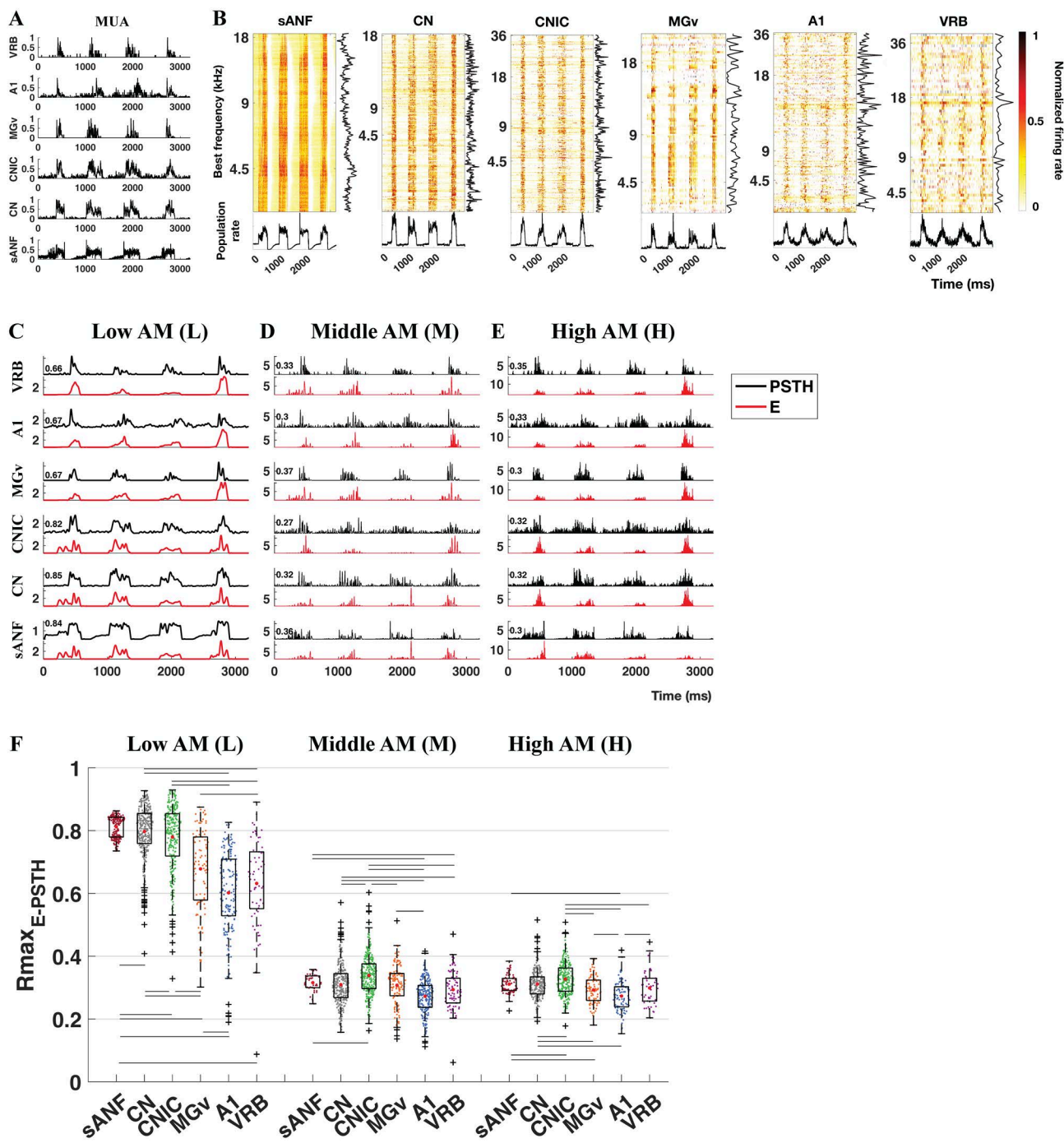
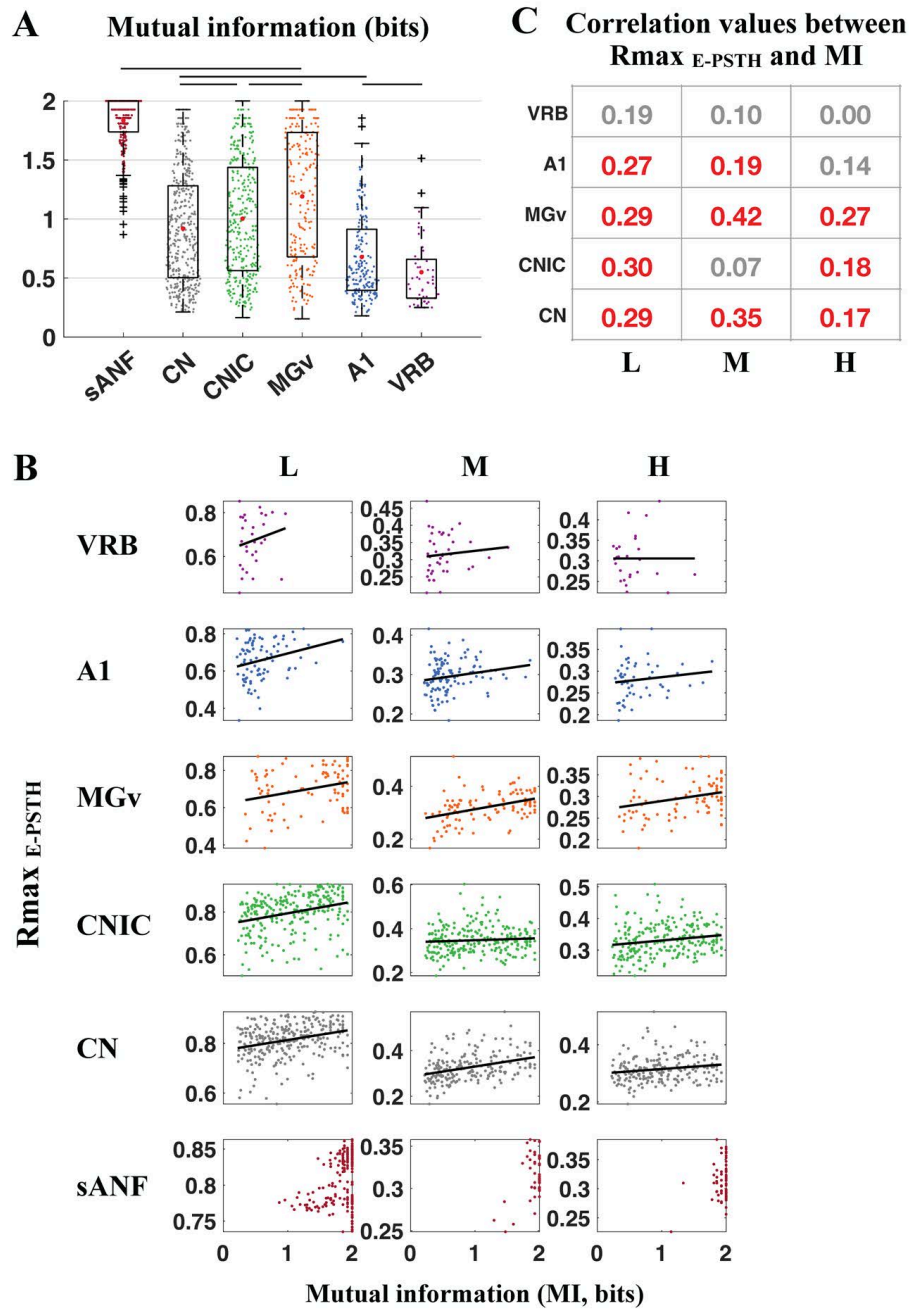


Figure 2.



**Figure 3.**



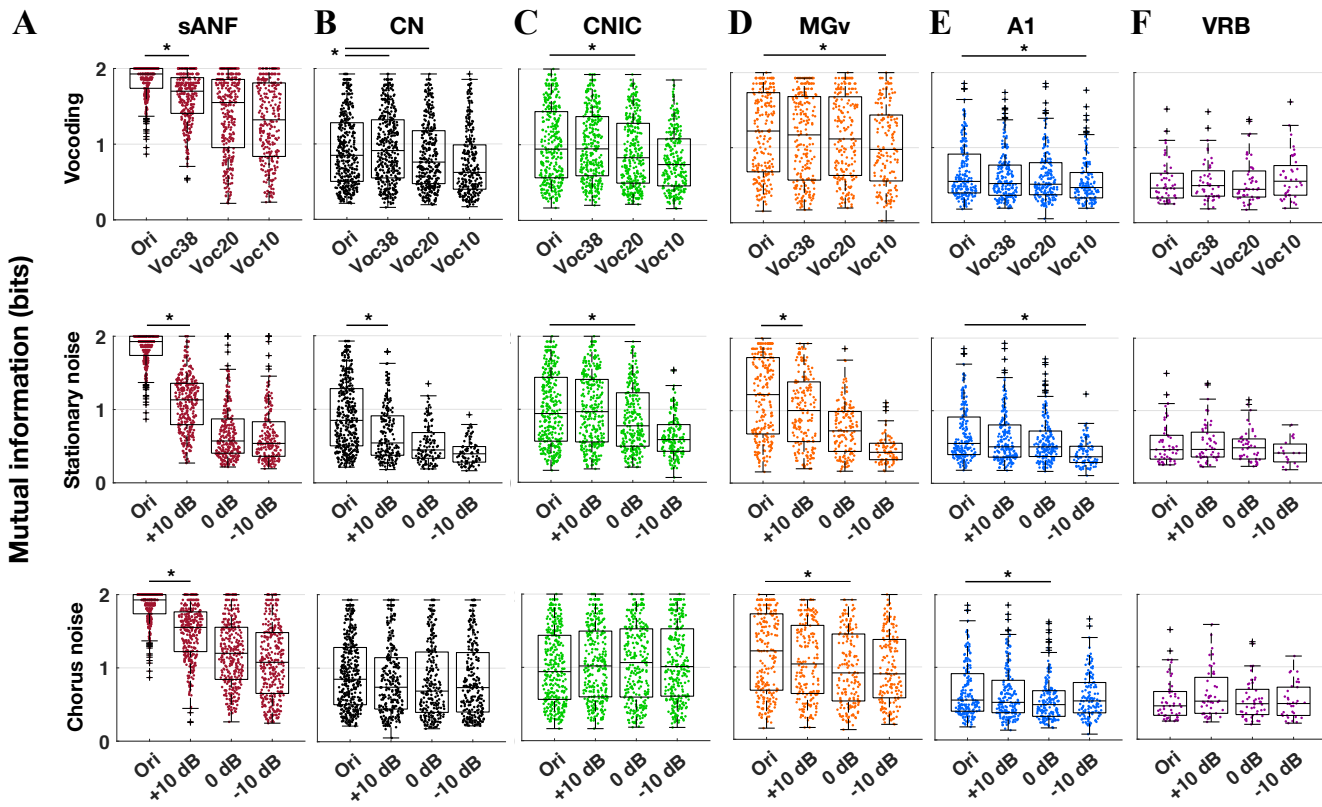


Figure 4.

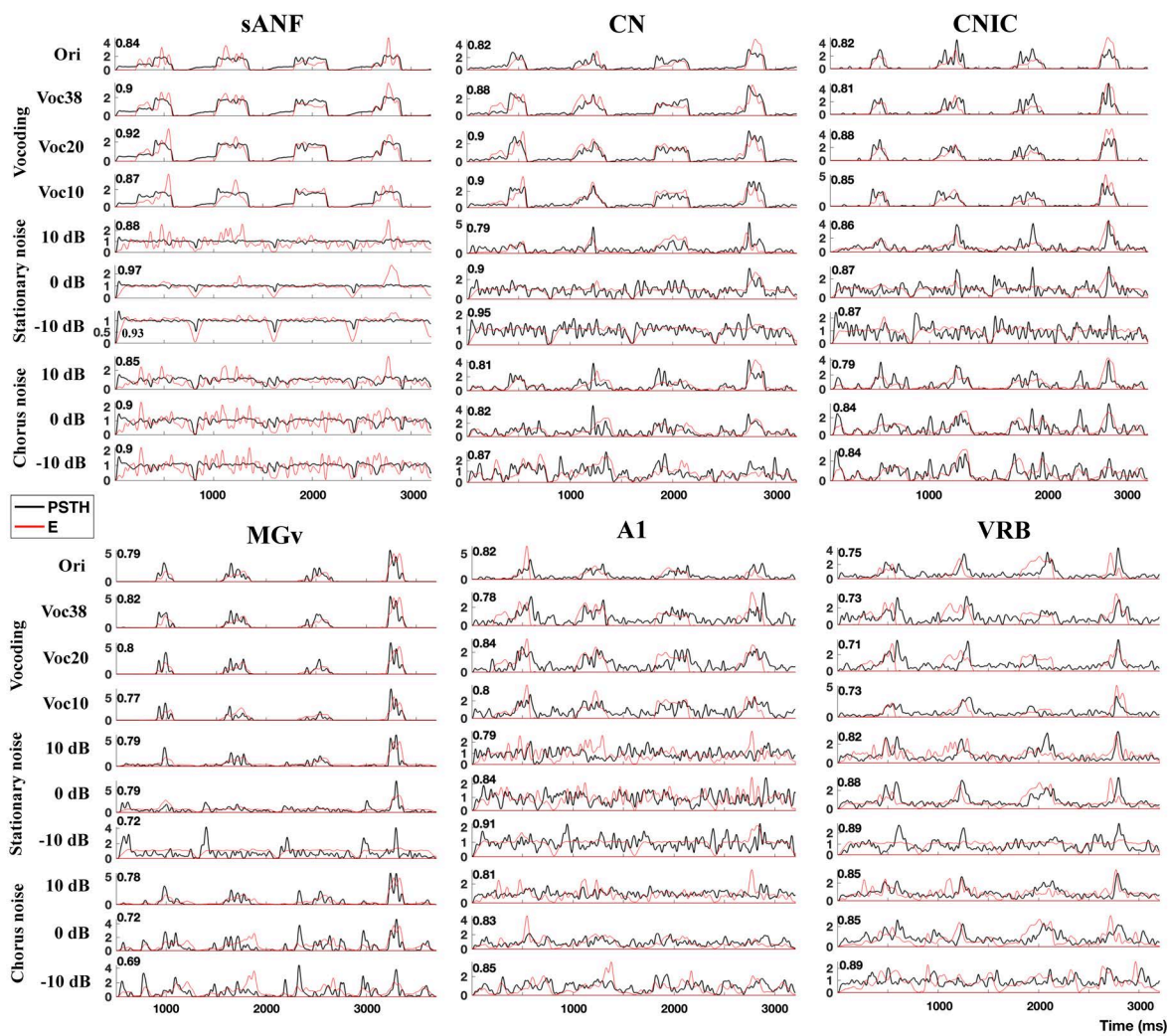


Figure 5.

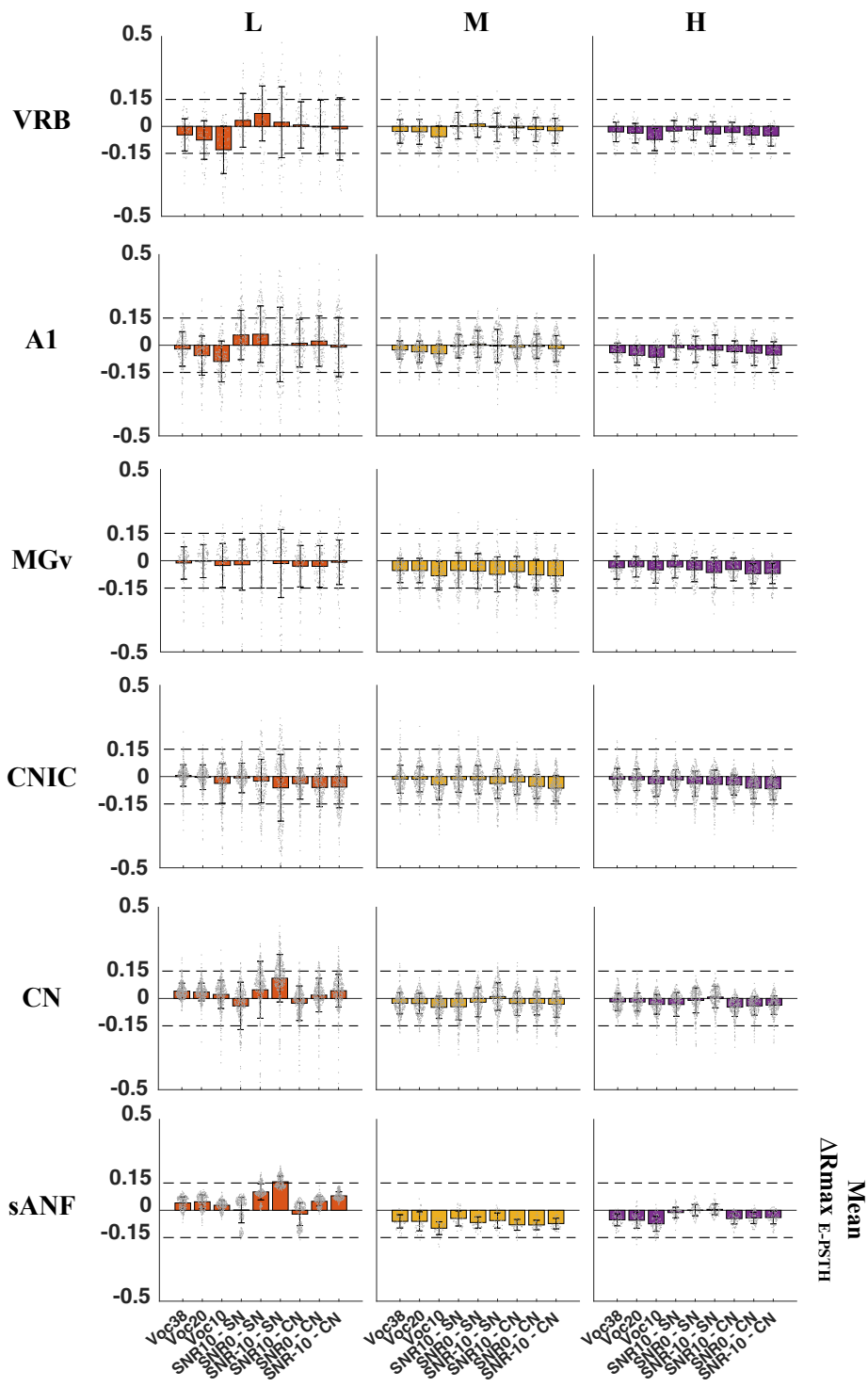


Figure 6.

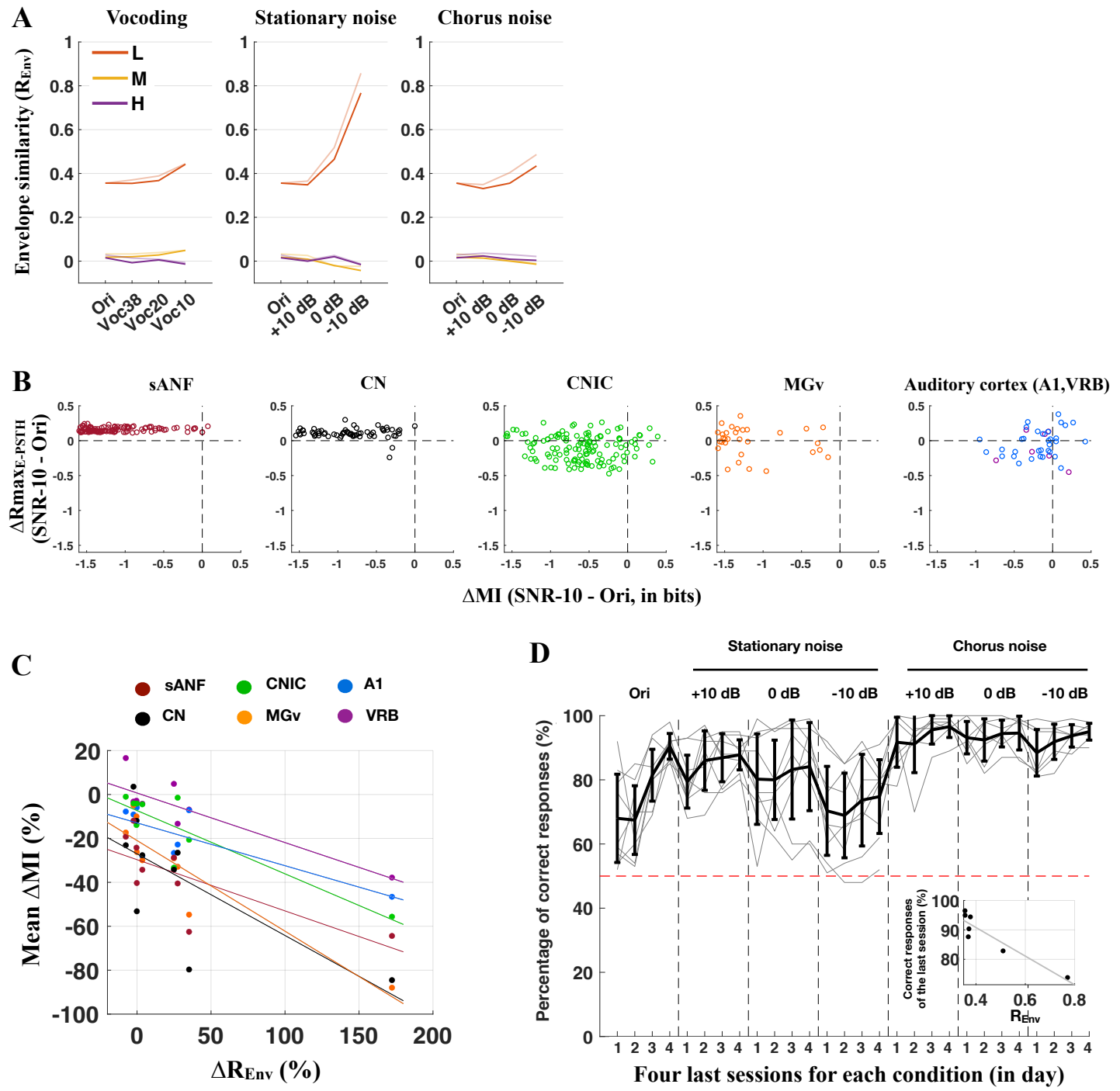
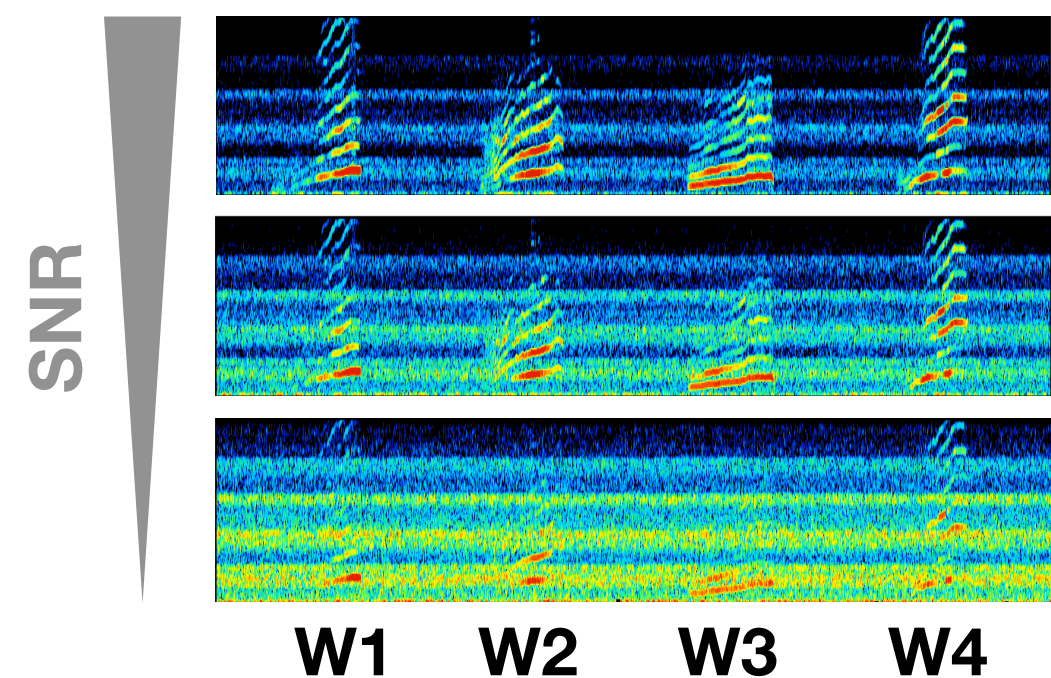


Figure 7.

## Methods

### Vocalizations in noise



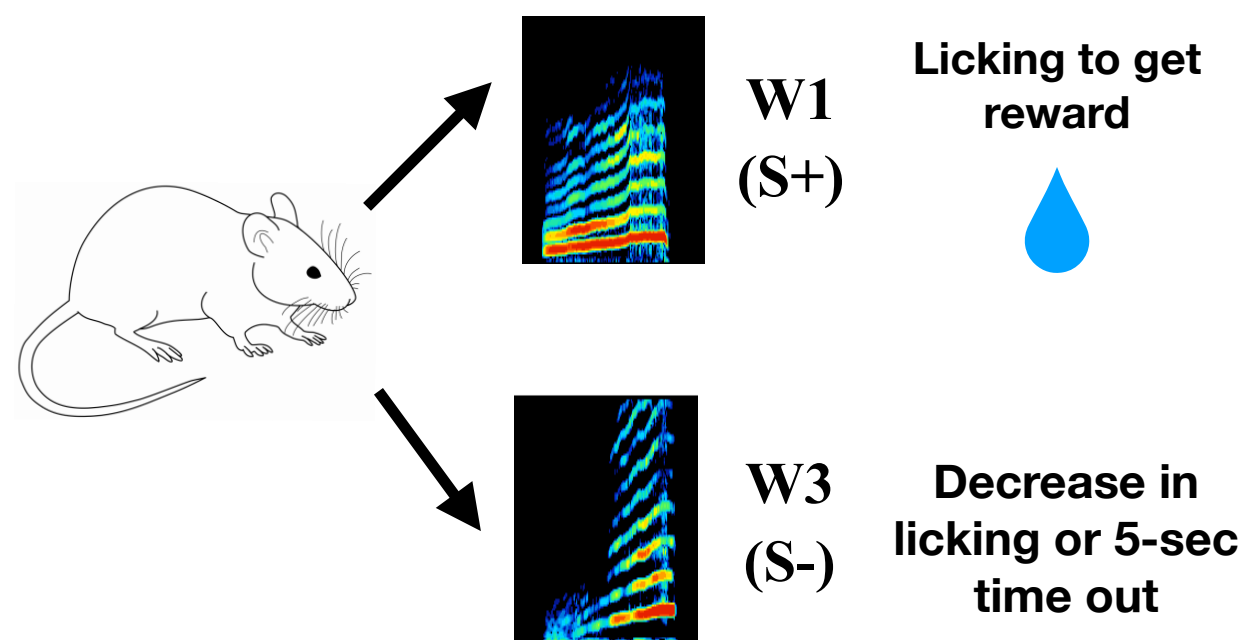
Extracellular recordings from the whole auditory system

MUA



+

Simulations of auditory nerve fibers



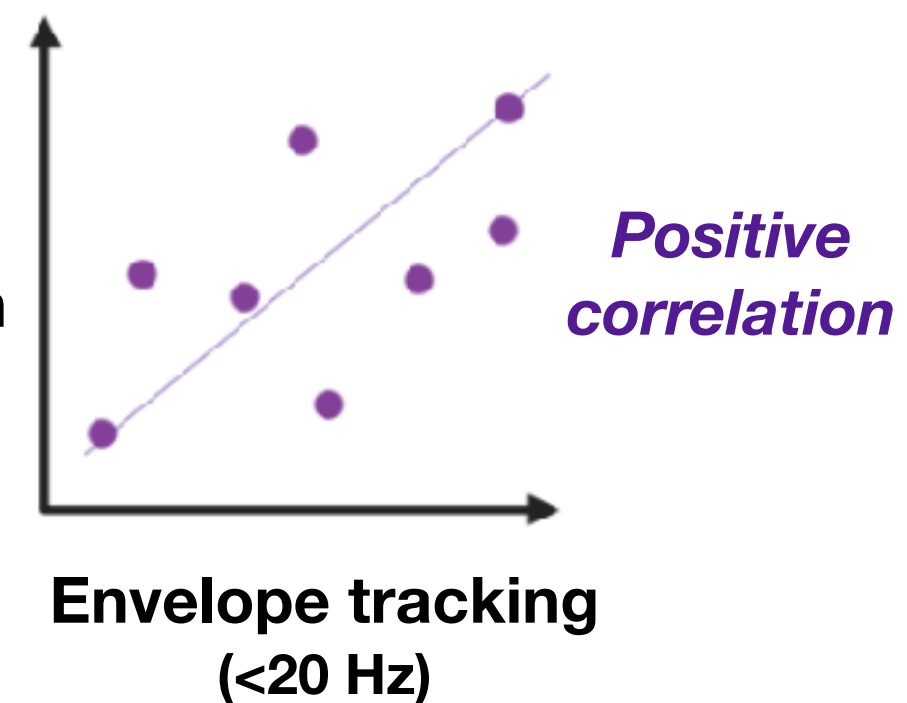
Go/No-Go sound discrimination task

## Results



*In quiet*

Neuronal discrimination



*Noise addition*

Envelope tracking abilities remain stable

Behavioral and neuronal discrimination

