



**HAL**  
open science

# Cortical and subcortical neurons discriminate sounds in noise on the sole basis of acoustic amplitude modulations

Samira Souffi, Christian Lorenzi, Chloé Huetz, Jean-Marc Edeline

## ► To cite this version:

Samira Souffi, Christian Lorenzi, Chloé Huetz, Jean-Marc Edeline. Cortical and subcortical neurons discriminate sounds in noise on the sole basis of acoustic amplitude modulations. 2019. hal-02111910

**HAL Id: hal-02111910**

**<https://hal.science/hal-02111910>**

Preprint submitted on 10 May 2019

**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.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28

## **Cortical and subcortical neurons discriminate sounds in noise on the sole basis of acoustic amplitude modulations**

Souffi S.<sup>1,2</sup>, Lorenzi C.<sup>3</sup>, Huetz C.<sup>1,2</sup>, Edeline J-M\*<sup>1,2</sup>.

Paris-Saclay Institute of Neurosciences (Neuro-PSI), Department Cognition and Behavior

<sup>1</sup> CNRS UMR 9197, <sup>2</sup> Université Paris-Sud, Bâtiment 446, 91405 Orsay cedex, France

<sup>3</sup> Laboratoire des systèmes perceptifs, UMR CNRS 8248, Département d'Etudes Cognitives, Ecole normale supérieure, Université Paris Sciences & Lettres, Paris, France.

Number of Figures: 5

Number of Supplementary figures: 5

Number of Tables: 1

Abbreviated title : Cortical and subcortical sound discrimination in noise

*Corresponding Author :*

Jean-Marc Edeline

Paris-Saclay Institute of Neuroscience (Neuro-PSI)

UMR CNRS 9197 Université Paris-Sud, Bâtiment 446,

91405 Orsay cedex, France

email: [jean-marc.edeline@u-psud.fr](mailto:jean-marc.edeline@u-psud.fr)

29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

## Abstract

Humans and animals maintain accurate sound discrimination in the presence of loud sources of background noise. It is commonly assumed that this ability relies on the robustness of auditory cortex responses. However, no attempt has been made to characterize neural discrimination of sounds masked by noise at each stage of the auditory system and disentangle the sub-effects of noise, namely the distortion of temporal cues conveyed by modulations in instantaneous amplitude and frequency, and the introduction of randomness (stochastic fluctuations in amplitude). Here, we measured neural discrimination between communication sounds masked by steady noise in the cochlear nucleus, inferior colliculus, auditory thalamus, primary and secondary auditory cortex at several signal-to-noise ratios. Sound discrimination by neuronal populations markedly decreased in each auditory structure, but collicular and thalamic populations showed better performance than cortical populations at each signal-to-noise ratio. Comparison with neural responses to tone-vocoded sounds revealed that the reduction in neural discrimination caused by noise was mainly driven by the attenuation of slow amplitude modulation cues, with the exception of the cochlear nucleus that showed a dramatic drop in discrimination caused by the randomness of noise. These results shed new light on the specific contributions of subcortical structures to robust sound encoding, and demonstrate that neural discrimination in the presence of background noise is mainly determined by the distortion of the slow temporal cues conveyed by communication sounds.

### Acknowledgments:

CL and JME were supported by grants from the French Agence Nationale de la Recherche (ANR) (ANR-14-CE30-0019-01). CL was also supported by grants ANR-11-0001-02 PSL and ANR-10-LABX-0087. SS was supported by the Fondation pour la Recherche Médicale (FRM) grant number ECO20160736099.

We thank warmly Léo Varnet for help in analyzing the modulation spectra of the acoustic stimuli and Roger Mundry for detailed and relevant comments on statistical analyses. We thank Nihaad Paraouty for training us to cochlear-nucleus surgery. We also wish to thank Mélanie Dumont, Aurélie Bonilla and Céline Dubois for taking care of the guinea-pig colony. We warmly thank Dr. Dexter Irvine and Dr. Jonathan Fritz for helpful comments on an earlier draft of this paper.

66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100

## Introduction

Understanding the neural mechanisms used by the central auditory system to extract and represent relevant information for discriminating communication sounds in a variety of acoustic environments is a major goal of auditory neurosciences. This enterprise is motivated by the repeated observation that humans and animals successfully maintain high discrimination performance for speech and behaviorally salient calls when the latter are embedded into loud sources of background noise produced by environmental medium (such as tropical forests, underwater or urban environments)<sup>1-6</sup>. Previous studies have assumed that this perceptual robustness mainly relies on the capacity of cortical neurons to extract invariant features<sup>5,7-10</sup>. For example, in the cortical field L (the analogous of primary auditory cortex (A1) in bird), the percentage of correct neuronal discrimination between zebra-finch songs embedded in different types of acoustic maskers decreases proportionally to the target-to-masker ratio and parallels behavioral performance<sup>5</sup>. Similarly, in a secondary auditory area, neurons generate background-invariant representation of vocalizations at signal-to-noise ratios that match behavioral recognition thresholds<sup>10</sup>. More recently, between-vowels discrimination performance of neuronal populations located in A1 was found to resist to a large range of acoustic alterations (including changes in fundamental frequency, spatial location, or level) and was similar to behavioral performance<sup>9</sup>. The goal of the present study was to identify the auditory brain structures responsible for these robust neural computations and clarify the effects of background noise on the neural representation of communication sounds, knowing that external noise has three disruptive sub-effects on communication sounds<sup>11,12</sup>: noise attenuates the power of their amplitude modulation cues (AM, also called “temporal-envelope”)<sup>13-15</sup>, corrupts their frequency modulation cues (FM, also called “temporal fine structure”)<sup>16,17</sup>, and introduces stochastic fluctuations in level, that is statistical variability of the AM power<sup>14</sup>. To address these issues, we investigated whether the ability of populations of auditory neurons to discriminate between communication sounds belonging to the same category (e.g. the alarm call in guinea pig) and masked by external noise *increases or decreases* along the auditory pathway from the first auditory relay (the cochlear nucleus) up to the primary and secondary cortical areas. An increased ability may result from the specialization of cortical responses for detecting crucial vocalization features<sup>10,18,19</sup>, whereas a decreased ability may result from the loss of spectro-temporal details promoting the identification of auditory objects<sup>20, 21</sup>. For the first time, the discrimination performance of neuronal populations recorded along the auditory

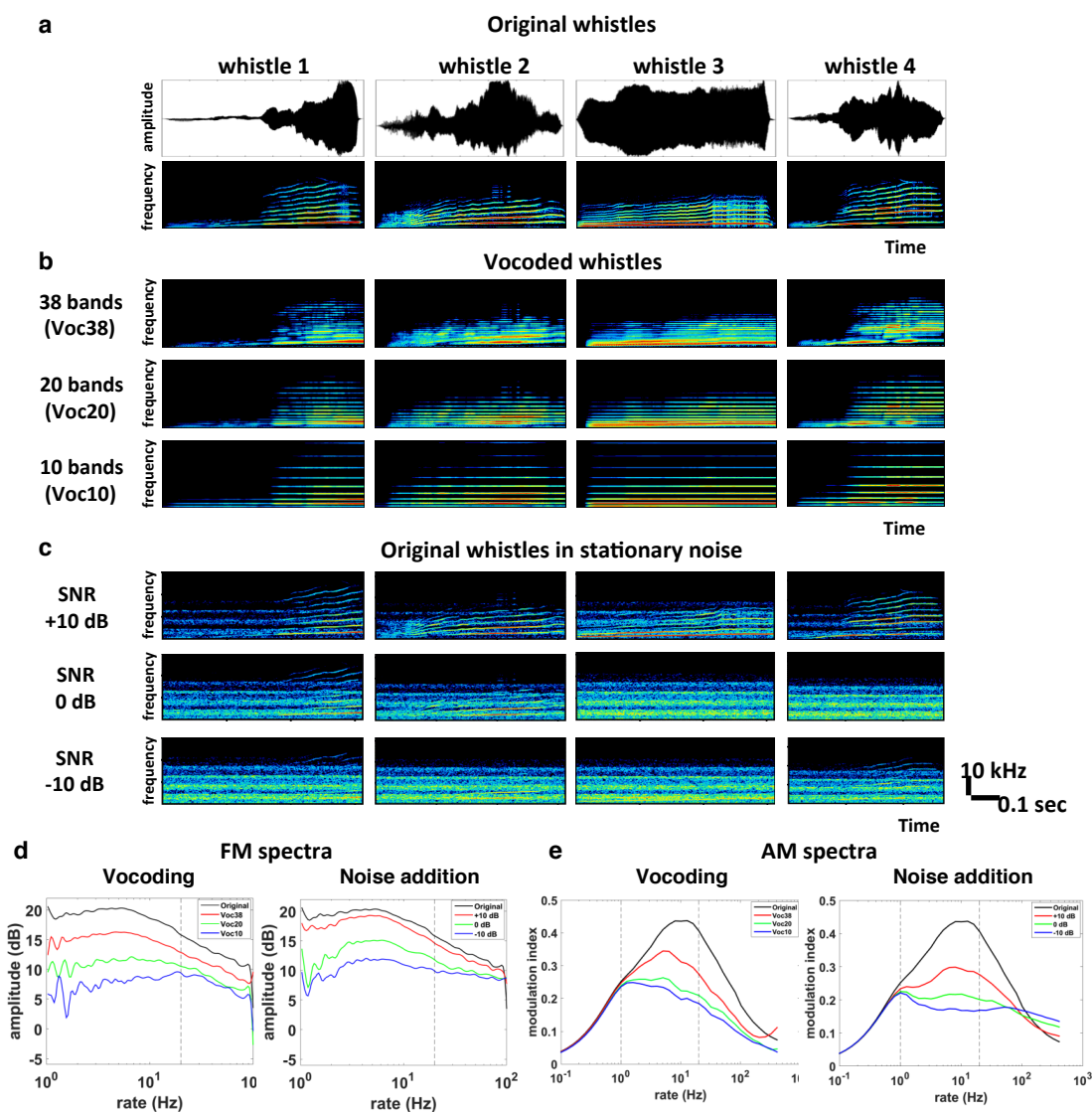
101 pathway, from the cochlear nucleus up to a secondary auditory cortex, was assessed for four  
102 utterances of the same vocalization presented against a stationary broadband noise using three  
103 signal-to-noise ratios (SNRs). The results were compared to the effects of an artificial signal-  
104 processing scheme (a tone vocoder) that progressively degraded acoustic AM and FM cues  
105 (within 38 to 10 frequency bands) without introducing any stochastic fluctuation as in the case  
106 of background noise. AM and FM spectra of communication sounds were computed at the  
107 output of simulated cochlear filterbank for each acoustic alteration allowing us to identify  
108 masking noise and vocoder conditions matched in terms of amount of modulation reduction.  
109 We then correlated these reductions of temporal modulation cues with the discriminative  
110 neuronal performance recorded in each structure. We demonstrate that, with the noticeable  
111 exception of the cochlear nucleus, the larger the reduction of AM cues (the first sub-effect of  
112 background noise), the larger the decrease in discriminative abilities in cortical and  
113 subcortical structures. Corruption of FM cues (the second sub-effect of noise) had little  
114 observable effects if any on neural discrimination. Introduction of stochastic fluctuations (the  
115 third sub-effect of noise) impacted neural discrimination in the cochlear nucleus only. In  
116 addition, this study revealed that, for each acoustic distortion tested here, the highest level of  
117 discrimination was found in subcortical structures, either at the collicular level (in masking-  
118 noise conditions) or at the thalamic level (in vocoder conditions).

119

120

## Results

121 From a database of 2334 recordings collected in the different auditory structures, two criteria  
 122 were used to include neuronal recordings in our analyses. A recording had to show a  
 123 significant STRF (see Methods) and an evoked firing rate significantly above spontaneous  
 124 firing rate (200 ms before each original vocalization) for at least one of the four original  
 125 vocalizations. Applying these two criteria led to the inclusion of 499 recordings in CN, 386  
 126 recordings in CNIC, 262 recordings in MGv, 354 recordings in A1 and 95 recordings in VRB  
 127 (see supplementary Table 1). In the following sections, the neuronal responses to the original  
 128 vocalizations (see Fig. 1a) presented in quiet are compared across brain structures and the  
 129 discriminative abilities are described at the individual and population level. The neuronal  
 130 discriminative abilities tested at the cortical and subcortical level with tone vocoded  
 131 vocalizations (Fig. 1b) and vocalizations presented against different levels of masking noise  
 132 (Fig. 1c) are described and compared next.

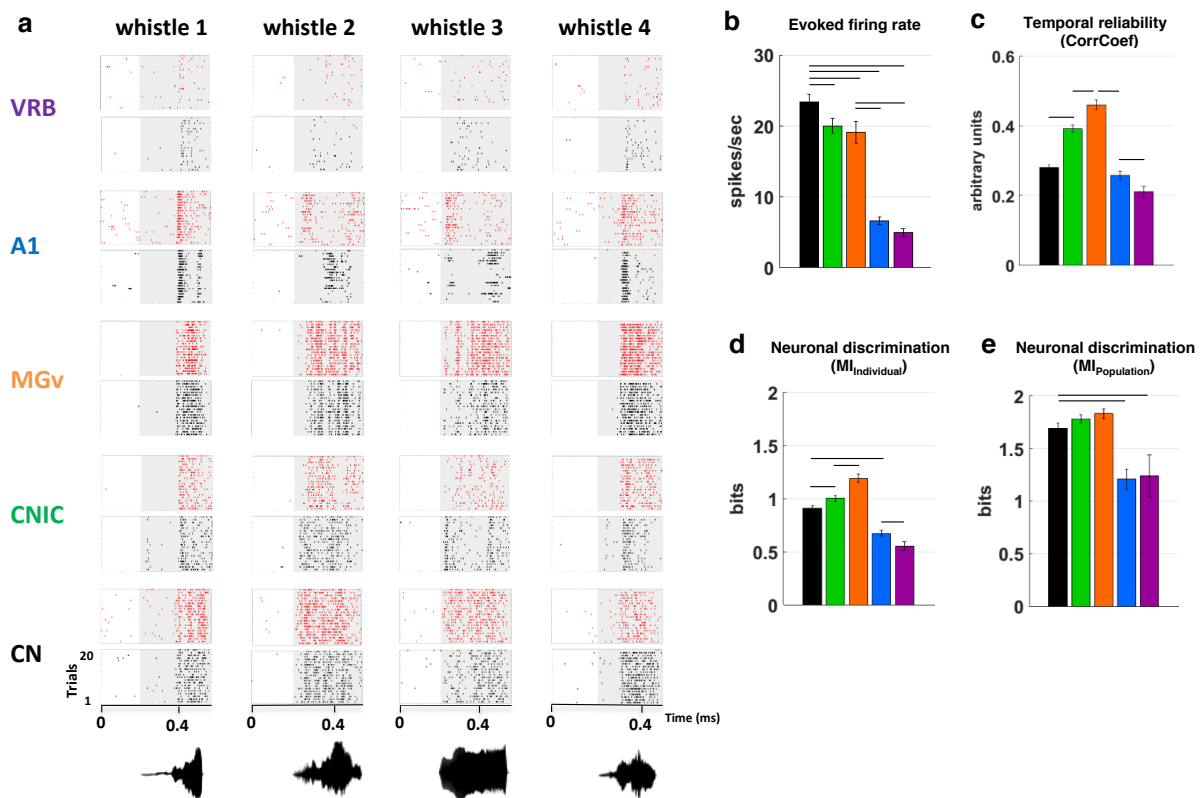


133

134 **Figure 1. Acoustic stimuli and averaged modulation spectra.** **a.** Waveforms (*top*) and spectrograms (*bottom*) of the four  
 135 original whistles used in this study. **b.** From *top to bottom*, spectrograms of the four vocoded whistles using 38, 20 and 10  
 136 frequency bands. **c.** From *top to bottom*, spectrograms of the four original whistles embedded in stationary noise at three  
 137 SNRs (+10, 0 and -10 dB) and spectrograms of the stationary noise only (*Noise only*). **d.** Vocoding and noise effects on  
 138 frequency-modulation (FM) spectra. The two plots represent the averaged modulation spectra of the four original  
 139 vocalizations (*in black*), vocoded vocalizations (*Voc38, Voc20 and Voc10: red, green and blue respectively, left panel*) and  
 140 vocalizations in stationary noise at three SNRs (+10, 0 and -10 dB : *red, green and blue respectively, right panel*) **e.**  
 141 Vocoding and noise effects on amplitude-modulation (AM) spectra. The two plots represent the averaged modulation spectra  
 142 of the four original vocalizations (*in black*), vocoded vocalizations (*Voc38, Voc20 and Voc10: red, green and blue*  
 143 *respectively, left panel*) and vocalizations in stationary noise at three SNRs (+10, 0 and -10 dB : *red, green and blue*  
 144 *respectively, right panel*). Vertical black dashed lines on AM and FM spectra correspond to the frequency range (1-20 Hz)  
 145 selected for the data analysis.  
 146

### 147 ***Discrimination of the original vocalizations in quiet culminates at the subcortical level***

148 Figure 2a displays neuronal responses of two simultaneous recordings obtained at five levels  
 149 of the auditory pathway (CN, CNIC, MGv, AI and VRB). The neuronal responses were strong  
 150 and sustained in the three subcortical structures whereas they were more phasic in AI and  
 151 more diffuse in VRB. For most of the recordings, temporal patterns of responses were clearly  
 152 reproducible from trial-to-trial, but they differed from one vocalization to another both at the  
 153 cortical and subcortical level.



154 **Figure 2. Subcortical neurons discriminate better the original vocalizations than cortical neurons.** **a.** From *bottom to*  
 155 *top*, neuronal responses were recorded in CN, CNIC, MGv, A1 and VRB simultaneously under 16 electrodes but only two  
 156



157 are represented here, with alternated black and red colors. Each dot represents the emission of an action potential and each  
158 line corresponds to the neuronal discharges to one of four original whistles. The grey part of rasters corresponds to evoked  
159 activity. The waveforms of the four original whistles are displayed under the rasters. **b-e**. The panels show **(b)** the evoked  
160 firing rate (spikes/sec), **(c)** the temporal reliability quantified by the CorrCoef value (arbitrary units), **(d)** the neuronal  
161 discrimination assessed by the mutual information (MI) computed at the level of the individual recording ( $MI_{\text{Individual}}$ , bits)  
162 and **(e)** at the level of neuronal population ( $MI_{\text{Population}}$ , bits) with populations of 9 simultaneous recordings obtained with the  
163 four original vocalizations in CN (*in black*), CNIC (*in green*), MGv (*in orange*), A1 (*in blue*) and VRB (*in purple*). In each  
164 structure, error bars represent the SD of the mean values and black lines represent significant differences between the mean  
165 values (unpaired *t* test,  $p < 0.05$ ). Note that the evoked firing rate decreases from the CN to VRB but both the temporal  
166 reliability (CorrCoef) and the discriminative ability (MI) values reach a maximal value in MGv. Note also that at the  
167 population level, all the subcortical structures discriminate better the original vocalizations than cortical areas.

168

169 Quantifications of evoked responses to original vocalizations are presented on Figures 2b-e  
170 for each auditory structure. These analyses clearly pointed out large differences between the  
171 mean values of evoked firing rate, CorrCoef and MI quantified at the cortical vs. at the  
172 subcortical level. First, the evoked firing rate was significantly higher in the subcortical  
173 structures than in the cortex (unpaired t-test, lowest  $p$  value  $p < 0.001$ ). It was also higher in  
174 CN than in the other subcortical structures (Fig. 2b). Second, the CorrCoef values were  
175 significantly higher in the subcortical structures than in AI and VRB (Fig. 2c), indicating that  
176 the trial-to-trial reliability of evoked responses was stronger at the subcortical than at the  
177 cortical level, reaching its maximum in the CNIC and MGv. CorrCoef value significantly  
178 increased from CN to CNIC (unpaired t-test,  $p < 0.01$ ), and then from CNIC to MGv ( $p < 0.01$ ;  
179 Fig. 2c). CorrCoef decreased significantly between the two cortical areas and CNIC and MGv  
180 ( $p < 0.01$ ); it also decreased relative to CN ( $p = 0.09$ ). The CorrCoef value was also significantly  
181 lower in VRB than in AI ( $p = 0.035$ ). Third, the  $MI_{\text{Individual}}$  values found at the subcortical level  
182 were significantly higher than at the cortical level (unpaired t-test, highest  $p < 0.001$  between  
183 the cortex and the other structures; Fig. 2d). At the subcortical level, the  $MI_{\text{Individual}}$  value was  
184 significantly higher in MGv than in CNIC and CN (unpaired t-test,  $p < 0.01$ ). The  $MI_{\text{Individual}}$   
185 value was also significantly lower in VRB than in AI ( $p = 0.037$ ). The highest mean  
186  $MI_{\text{Individual}}$  value was found in MGv, suggesting that, on average, thalamic neurons  
187 discriminate better the four original whistles than the other auditory structures. Note that this  
188 high  $MI_{\text{Individual}}$  value is related to the higher temporal trial-to-trial reliability (indexed by the  
189 CorrCoef value) obtained at the thalamic level (see Fig. 2c).

190 The distributions of  $MI_{\text{Individual}}$  value were plotted as a function of temporal precision for each  
191 structure (see supplementary Fig. 2) to investigate whether the higher mean  $MI_{\text{Individual}}$  values  
192 found in subcortical responses result from the fact that neurons in these structures



193 systematically conveyed more information about the stimuli than cortical neurons. With a  
194 temporal resolution of 8 ms, the proportion of neurons having a  $MI_{\text{Individual}}$  curve reaching a  
195 value of 1.5 bits (indicating that at least 3 stimuli can be discriminated) was similar in CN and  
196 CNIC (22% and 21% respectively, chi-square test  $<1$ ;  $p=0.99$ ); it was significantly higher in  
197 MGv (39%,  $p=0.017$  and  $p=0.04$ ) and lower in AI (3.5%,  $p=0.001$  in both cases) and in VRB  
198 (2%,  $p=0.001$ ).

199 Finally, MI was also computed based on the temporal patterns obtained from two to sixteen  
200 simultaneous recordings to determine whether the discriminative abilities of neural networks  
201 confirm the results obtained at the individual (i.e., single unit) level.  $MI_{\text{Population}}$  quantifies how  
202 well the four whistles can be discriminated based on temporal patterns expressed by  
203 populations distributed on the tonotopic map. Figure 2e presents the  $MI_{\text{Population}}$  computed  
204 from 9 simultaneous recordings for the five structures under investigation: This figure  
205 confirms that neural populations in subcortical structures discriminate the four original  
206 whistles better than the cortical populations (unpaired t-test, highest p value  $p<0.002$  between  
207 CN and VRB) without any statistical difference between the three subcortical structures. An  
208 examination of the evolution of the  $MI_{\text{Population}}$  as a function of the number of simultaneous  
209 recordings in the different structures revealed that the growth functions rapidly reached high  
210 values in all subcortical structures, whereas there were only a few of such curves in AI and  
211 VRB whatever the number of recordings considered (see supplementary Fig. 3).

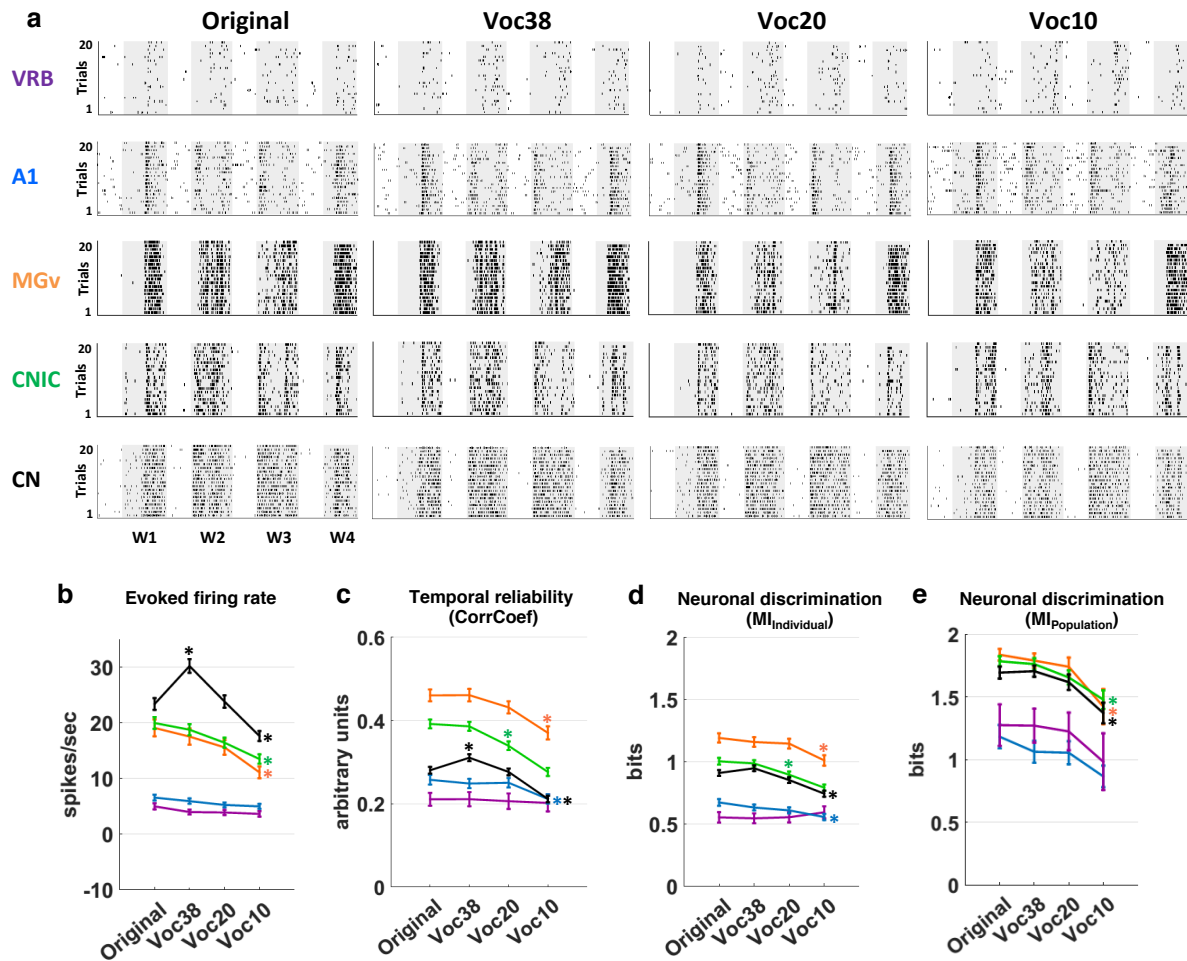
212

213

#### 214 ***Modest effects of tone vocoding***

215

216 Figure 3a displays rasters of recordings obtained in the five structures in response to the  
217 original and tone vocoded vocalizations. As illustrated here, at all levels, neurons still  
218 responded to the vocoded stimuli even for 10 frequency-band vocoded stimuli. However, the  
219 firing rate was decreased in each structure, and so was the precise organization of neuronal  
220 responses, especially with the 10 frequency-band vocoded stimuli.



221  
 222 **Figure 3. Vocoding slightly alters neuronal responses at each stage of the auditory system.** **a.** From left to right, raster  
 223 plots of responses to the four original whistles (*Original*) and their vocoded versions generated using either 38, 20 or 10  
 224 frequency bands (*Voc38*, *Voc20* and *Voc10*). From bottom to top, neuronal responses were recorded in CN, CNIC, MGv, A1  
 225 and VRB. **b-e.** The mean values ( $\pm$ SEM) represent **(b)** the evoked firing rate (spikes/sec), **(c)** the temporal reliability  
 226 represented by the CorrCoef value (arbitrary units), **(d)** the neuronal discrimination assessed by the mutual information (MI)  
 227 computed at the level of the individual recordings ( $MI_{\text{Individual}}$ , bits) and **(e)** at the level of neuronal population ( $MI_{\text{Population}}$ ,  
 228 bits) with populations of 9 simultaneous recordings obtained with original (*Original*) and vocoded vocalizations (*Voc38*,  
 229 *Voc20* and *Voc10*) in CN (in black), CNIC (in green), MGv (in orange), A1 (in blue) and VRB (in purple) (one-way ANOVA,  
 230 \* $P < 0.05$ ). At the population level, the discriminative abilities significantly decreased only for 10 frequency bands in  
 231 subcortical structures and did not decrease in cortical areas.

232  
 233 Figures 3b-e summarize vocoding effects on the four parameters quantifying neuronal  
 234 responses. Apart from an initial increase in firing rate observed only in CN with the 38-band  
 235 vocoded stimuli, the effects on evoked firing rate were modest in each structure (Fig. 2b): A  
 236 significant decrease in evoked firing rate between the responses to the original and the 10-  
 237 band vocoded vocalizations was only found at the subcortical level (for all subcortical  
 238 structures, ANOVA test:  $p < 0.001$ ,  $F_{\text{CN}(3,1995)} = 22.6$ ;  $F_{\text{CNIC}(3,1543)} = 8.85$ ;  $F_{\text{MGv}(3,1047)} = 6.55$ ),  
 239 whereas there was no decrease in either AI or VRB. Vocoding also decreased the mean

240 CorrCoef values in each structure except in VRB (Fig. 3c): this decrease was significant with  
241 the 10-band vocoded vocalizations in CN, MGv and in AI (ANOVA test, highest p value,  
242  $p < 0.02$ ,  $F_{CN(3,1930)} = 26.48$ ,  $F_{MGv(3,889)} = 7.7$ ,  $F_{AI(3,1125)} = 3.42$ ). The decrease in CorrCoef value  
243 was already significant with 20-band vocoded vocalizations in the CNIC ( $p < 0.001$ ,  
244  $F_{(3,1391)} = 26.19$ ).

245 Similarly, tone vocoding decreased the  $MI_{Individual}$  values in each structure except in VRB  
246 (Fig. 3d). Here too, the decrease was significant with the 10-band vocoded vocalizations in  
247 CN, MGv and AI (ANOVA test, highest p value,  $p < 0.02$ ,  $F_{CN(3,1445)} = 12.23$ ,  $F_{MGv(3,810)} = 3.75$ ,  
248  $F_{AI(3,720)} = 3.59$ ) and it was already significant with 20-band vocoded vocalizations in the  
249 CNIC ( $p < 0.001$ ,  $F_{(3,1231)} = 13.17$ ). At the population level, there was a striking difference  
250 between the subcortical and cortical structures (Fig. 3e): compared with the values obtained  
251 with original vocalizations, the  $MI_{Population}$  values computed with the 10-band vocoded  
252 vocalizations were significantly lower in the subcortical structures (ANOVA test, highest p  
253 value,  $p < 0.005$ ,  $F_{CN(3,127)} = 6.46$ ,  $F_{MGv(3,67)} = 4.62$ ,  $F_{CNIC(3,115)} = 6.28$ ) but not at the cortical level.  
254 The evolution of  $MI_{Population}$  as a function of the number of simultaneous recordings (see  
255 supplementary Fig. 4a) indicated that in each subcortical structure, the curves rapidly reached  
256 high  $MI_{Population}$  values (close to the maximal value of 2) in each vocoding conditions, whereas  
257 there were only a few of such curves in AI and VRB whatever the vocoding condition.

258 In conclusion, for the five auditory structures under study, the neuronal responses to 10-band  
259 vocoded vocalizations were slightly weaker, temporally less accurate and less discriminative  
260 than the responses to the original vocalizations. Nonetheless, subcortical neurons still  
261 maintained the highest ability to discriminate between tone vocoded vocalizations, both at the  
262 level of individual recordings and at the population level.

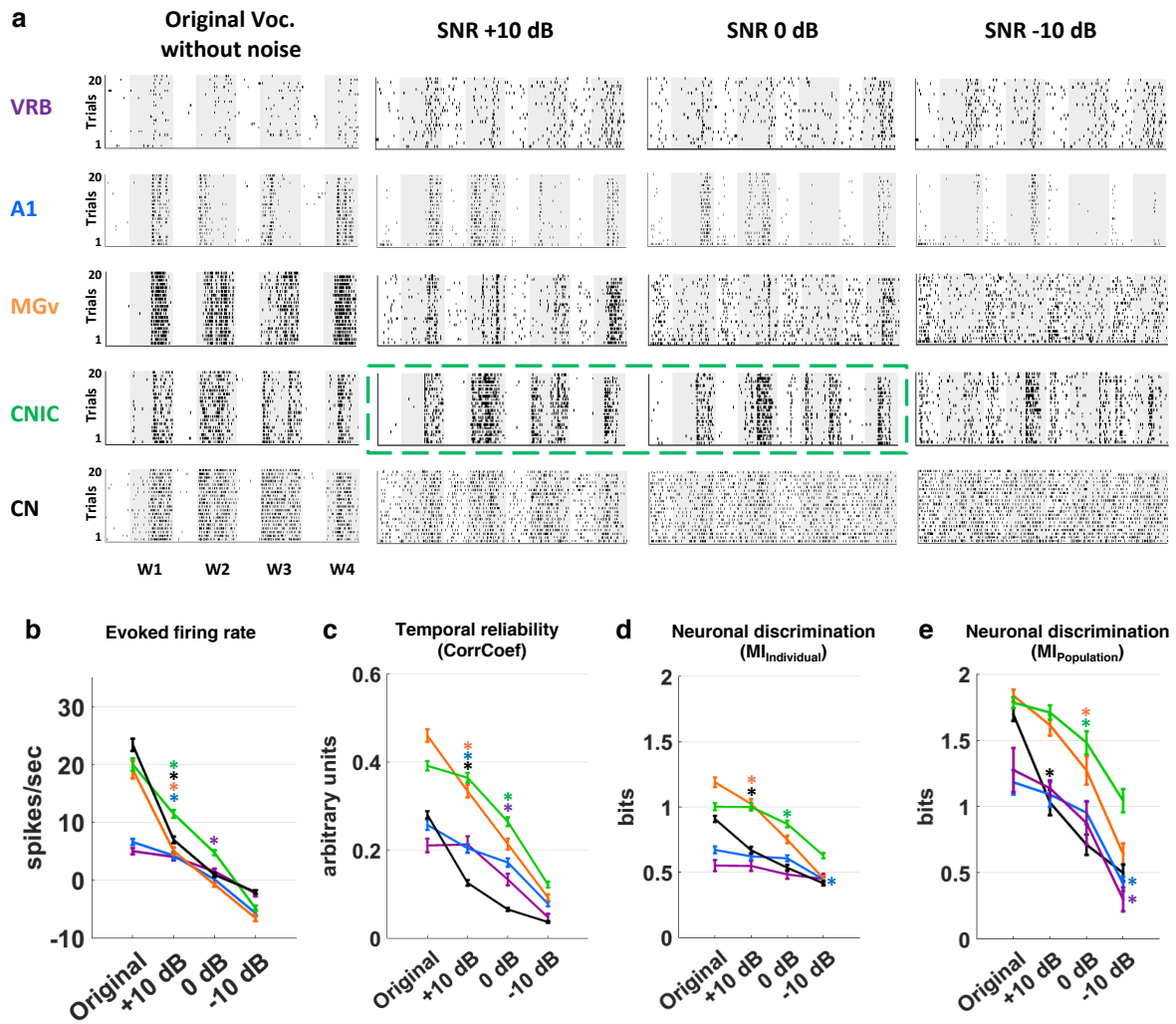
263

### 264 ***Pronounced effects of masking noise on neuronal discrimination***

265

266 The rasters presented in figure 4a show the effects produced by presenting the original  
267 vocalizations against a stationary noise at three SNRs (+10, 0 and -10 dB). Masking noise  
268 attenuated neuronal responses at each level of the auditory system. However, auditory  
269 structures were differentially affected by noise. In these rasters, the responses in the CNIC did  
270 not change up to a 0 dB SNR, decreasing only at a -10 dB SNR. This was not the case in the  
271 other auditory structures where the responses decreased either at a +10 dB SNR (MGv and  
272 CN) or at a 0 dB SNR (AI and VRB).

273



274  
 275 **Figure 4. Noise strongly reduces neuronal responses in all structures but to a lesser extent in the central nucleus of the**  
 276 **inferior colliculus.** **a.** From left to right, raster plots of responses of four original whistles (*Original*) and their noisy versions  
 277 in stationary noise at three SNRs: +10, 0 and -10 dB. From bottom to top, neuronal responses were recorded in CN, CNIC,  
 278 MGv, A1 and VRB. The grey area corresponds to the evoked activity for each vocalization. The green dashed lines show a  
 279 typical example of CNIC neuronal responses that are resistant to the noise addition. **b-e.** The mean values ( $\pm$ SEM) represent  
 280 **(b)** the evoked firing rate (spikes/sec), **(c)** the temporal reliability represented by the CorrCoef value (arbitrary units), **(d)** the  
 281 neuronal discrimination assessed by the mutual information (MI) computed at the level of the individual recordings  
 282 ( $MI_{\text{Individual}}$ , bits) and **(e)** at the level of neuronal population ( $MI_{\text{Population}}$ , bits) with populations of 9 simultaneous recordings  
 283 obtained with original and vocalizations in stationary noise at three SNRs (+10, 0 and -10 dB SPL) in CN (*in black*), CNIC  
 284 (*in green*), MGv (*in orange*), A1 (*in blue*) and VRB (*in purple*) (one-way ANOVA,  $*P < 0.05$ ). Note that at the population  
 285 level, the discriminative abilities significantly decreased in all structures when SNR decreased, with the CNIC populations  
 286 still able to discriminate 2 out of 4 stimuli ( $MI_{\text{Population}}$  value  $> 1$ ).

287  
 288 Figures 4b-e summarize the effects of masking noise on the different parameters quantifying  
 289 neuronal responses. Masking noise significantly reduced the evoked firing rate in each  
 290 auditory structure as early as the +10 dB SNR (Fig. 4b, ANOVA test:  $p < 0.001$ ,  
 291  $F_{\text{CN}(3,1995)}=309.33$ ,  $F_{\text{CNIC}(3,1543)}=220.64$ ,  $F_{\text{MGv}(3,1047)}=155.07$ ,  $F_{\text{A1}(3,1415)}=96.27$ ), except in VRB.

292 Masking noise strongly reduced the CorrCoef values in CN and MGv at the highest (+10 dB)  
293 SNR tested here (Fig. 4c; ANOVA test,  $p < 0.001$ ,  $F_{CN(3,1884)} = 382.22$ ,  $F_{MGv(3,791)} = 155.82$ )  
294 whereas in the CNIC, this reduction was significant only at the 0 dB SNR (ANOVA test,  
295  $p < 0.001$ ,  $F_{(3,1357)} = 154.12$ ). At the cortical level, the CorrCoef values were significantly  
296 reduced in AI at the +10 dB SNR and in VRB at the 0 dB SNR (ANOVA test,  $p < 0.001$ ,  
297  $F_{AI(3,1093)} = 60.83$ ,  $F_{VRB(3,335)} = 29.56$ ). At the cortical level, noise significantly reduced the mean  
298  $MI_{Individual}$  value in AI at the -10 dB SNR (ANOVA test,  $p < 0.001$ ,  $F_{(3,649)} = 9.49$ ) whereas the  
299 mean  $MI_{Individual}$  value in VRB remained unchanged. At the subcortical level, noise reduced  
300 the  $MI_{Individual}$  values but again, there was a marked difference between the CNIC and the  
301 other subcortical structures: the mean  $MI_{Individual}$  value in CN and MGv was significantly  
302 reduced at the +10 dB SNR (Fig. 4d; ANOVA test,  $p < 0.001$ ,  $F_{CN(3,819)} = 56.75$ ,  
303  $F_{MGv(3,621)} = 63.61$ ), whereas the  $MI_{Individual}$  value in the CNIC was only significantly reduced at  
304 the SNR of 0 dB (ANOVA test,  $p < 0.001$ ,  $F_{(3,1078)} = 32.08$ ). Note, however, that at least 20% of  
305 the CN recordings maintained  $MI_{Individual}$  values above 1 bit, suggesting that a sub-population  
306 of CN neurons still sent information about the vocalization identity at higher brainstem  
307 centers (see Supplementary Fig. 5). This specific sub-population of CN neurons did not  
308 display parameters of their STRFs quantification that differ from the neurons exhibiting  
309  $MI_{Individual}$  values below 1 bit at the +10 dB SNR.

310 In noise conditions,  $MI_{Population}$  also allowed to characterize the effects of masking noise on the  
311 network discriminative abilities (Fig. 4e). At the cortical level, there was a significant  
312 reduction of  $MI_{Population}$  values only at the -10 dB SNR (ANOVA test,  $p < 0.001$ ,  
313  $F_{AI(3,111)} = 16.63$ ,  $F_{VRB(3,23)} = 11.41$ ) whereas there was a significant decrease in CN as early as  
314 the +10 dB SNR (ANOVA test,  $p < 0.001$ ,  $F_{(3,127)} = 51.49$ ). In MGv and CNIC, neuronal  
315 populations displayed the highest discriminative abilities although the decrease in  $MI_{Population}$   
316 value was significant at the 0 dB SNR (ANOVA test,  $p < 0.001$ ,  $F_{MGv(3,67)} = 41.59$ ,  
317  $F_{CNIC(3,115)} = 22.59$ ).

318 The evolution of the  $MI_{Population}$  as a function of the number of simultaneous recordings in the  
319 different structures (see Supplementary Fig. 4b) revealed that whatever the number of neurons  
320 considered, noise effects were similar: the population curves in CNIC and MGv grew up  
321 relatively rapidly and reached higher values than the curves obtained in CN and in the two  
322 cortical areas whatever the SNR.

323 To summarize, masking noise reduced similarly firing rate in each structure but impacted  
324 differently the neurons' discriminative abilities. Although cortical neurons were the most  
325 resistant to changes in noise level, the thalamic and collicular neurons maintained higher MI

326 values, with the CNIC neurons displaying the highest discriminative abilities both at the  
327 individual and population level in the most challenging condition (i.e., at the -10 dB SNR).

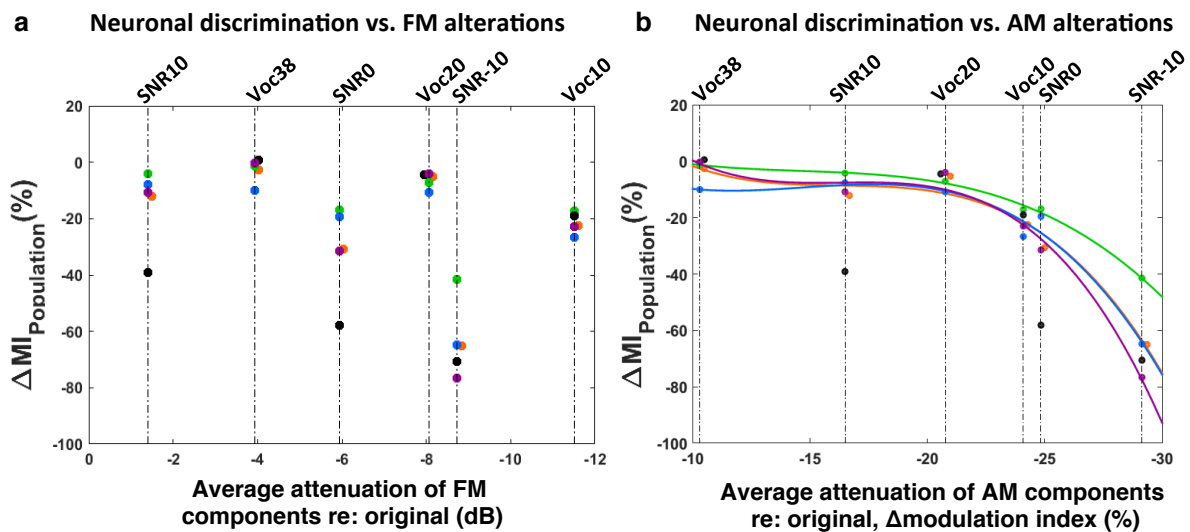
328

### 329 *Reduction of amplitude modulation cues explains changes in neuronal discrimination*

330

331 Tone vocoding degraded the spectro-temporal structure (i.e., it reduced both AM and FM  
332 cues) of original vocalizations in a deterministic way. Masking noise produced spectro-  
333 temporal degradations of the same nature, but it also introduced stochastic fluctuations in AM  
334 power. For each experimental condition (vocoding and noise), degradations of the spectro-  
335 temporal structure of vocalizations were quantified through the computation of AM and FM  
336 spectra at the output of a simulated cochlear filterbank. Not surprisingly, figure 1d (left panel)  
337 shows that tone vocoding corrupted drastically FM cues in each vocoding condition, and that  
338 this effect was already substantial in the 38-band vocoding condition. Importantly, this  
339 degradation was much stronger in vocoding conditions than in masking noise conditions (Fig.  
340 1d). Figure 1e reveals that both tone vocoding and masking noise also attenuated the AM cues  
341 conveyed by vocalizations. There were only small AM degradations in the 38-band vocoding  
342 and in +10 dB SNR conditions; whereas important AM degradations were observed for the 0  
343 dB SNR and the -10 dB SNR conditions. As for tone vocoding, this finding is consistent with  
344 the conclusions of previous modelling studies demonstrating the interplay between so-called  
345 temporal-envelope and TFS cues at the output of cochlear filters in response to vocoded  
346 sounds<sup>16, 22</sup>. Figure 5 relates these degradations of FM (Fig. 5a) and AM (Fig. 5b) cues to  
347 neural discrimination ( $MI_{\text{Population}}$ ) in the five brain structures for each experimental condition.  
348 More precisely, for all adverse conditions, figure 5 shows the changes in  $MI_{\text{Population}}$  for each  
349 auditory structure as a function of the attenuation of FM and AM cues (computed from  
350 modulation spectra for modulation rates between 1 and 20 Hz). Figure 5a reveals that an  
351 important attenuation of FM components caused by the 20-band vocoder was not associated  
352 with significant changes in neural discrimination in each structure, whereas a comparable  
353 attenuation caused by noise at -10 dB SNR resulted in a large drop in neural discrimination in  
354 each structure. Even more pronounced FM degradations caused by the 10-band vocoder  
355 condition produced smaller changes in  $MI_{\text{Population}}$  than the -10 dB SNR condition. This  
356 suggests that neural discrimination of vocalizations in the presence of background noise is not  
357 determined by the distortion of FM cues. A different pattern of results is obtained when  
358 changes in  $MI_{\text{Population}}$  for each auditory structure are represented as a function of degradations  
359 in AM cues. First, in all structures other than the CN,  $MI_{\text{Population}}$  is barely affected as long as

360 the reduction of the AM index ( $\Delta$ modulation index) remains lower than 25%; beyond this  
 361 limit, the  $MI_{Population}$  is markedly reduced (i.e., at -10 dB SNR). The straightforward  
 362 conclusion is that the reduction of AM cues is a key factor controlling the decrease in  
 363  $MI_{Population}$  at the cortical and subcortical levels. Second, in the cochlear nucleus, the impact  
 364 on the  $MI_{Population}$  is much larger in the noise conditions (at 10 dB SNR and 0 dB SNR) than in  
 365 the vocoding conditions, suggesting that the alteration of AM cues is not the only parameter  
 366 driving the  $MI_{Population}$  at the most peripheral level. Noise and vocoding have in common to  
 367 reduce the slow AM cues and corrupt the FM cues of vocalizations, but only background  
 368 noise introduces randomness, that is stochastic fluctuations in AM power. Therefore, in the  
 369 cochlear nucleus, the most likely factor responsible for the additional decrease in  $MI_{Population}$  is  
 370 noise stochasticity. Third, in the condition causing the most severe alterations of AM cues,  
 371 namely at -10 dB SNR, the  $MI_{Population}$  in CNIC is less impacted than in the other structures.  
 372 Consistent with previous studies<sup>23,24</sup>, this finding suggests that inferior colliculus neurons  
 373 adapt to the noise statistics while responding to the acoustic cues distinguishing between the  
 374 four target stimuli.



375  
 376 **Figure 5. Reduction of AM cues determine the neuronal discriminative abilities at the subcortical and cortical levels.**  
 377 **a.** Percentage of  $\Delta MI_{Population}$  as a function of degradation of FM components computed for each structure from means  
 378  $MI_{Population}$  or FM values obtained in all adverse conditions (noise and vocoded conditions) minus mean values in the original  
 379 condition. Each dot represents neuronal data ( $\Delta MI_{Population}$ ) in CN (in black), CNIC (in green), MGv (in orange), A1 (in blue)  
 380 and VRB (in purple). **b.** Percentage of  $\Delta MI_{Population}$  as a function of  $\Delta$ modulation index computed for each structure from  
 381 mean  $MI_{Population}$  or mean modulation-index values obtained in all adverse conditions and mean values in the original  
 382 condition. Each dot represents neuronal data ( $\Delta MI_{Population}$ ) in CN (in black), CNIC (in green), MGv (in orange), A1 (in blue)  
 383 and VRB (in purple). Polynomial curves fitting all acoustic conditions have been generated for each auditory structure (color  
 384 lines) except for the cochlear nucleus. Note that equivalent AM degradations ( $\Delta$ modulation index) induced equivalent



385 decreases in  $MI_{\text{Population}}$  except for the CN. Note also that there is a limit of 25% of AM reduction beyond which the  
386  $\Delta MI_{\text{Population}}$  strongly decreases in cortical and subcortical structures.  
387

## Discussion

388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421

Here, we demonstrate that the ability to discriminate between communication sounds is not increasing or decreasing monotonically along the auditory system: the neuronal discriminative abilities did strongly differ across auditory structures, and subcortical neurons in inferior colliculus and thalamus displayed higher discriminative abilities than cochlear nucleus and cortical neurons, both at the individual and population levels in each acoustic condition. Background noise markedly reduced the neuronal discriminative abilities in all auditory structures with larger effects in the cochlear nucleus. Amongst the three disruptive sub-effects of background noise identified in previous psychophysical investigations<sup>12</sup>, fidelity in the transmission of slow ( $< 20$  Hz) amplitude modulation information proved to be the main factor determining the neural discrimination abilities in noise at the cortical and subcortical levels. The effects of randomness were found to be limited to the most peripheral structure of the central auditory system, namely the cochlear nucleus.

### **The capacity to encode amplitude-modulation cues explains the better discrimination of the original stimuli by subcortical neurons**

To the best of our knowledge, this is the first time that a direct comparison of neural responses to the same natural stimuli has been made along the ascending auditory system. We computed  $MI_{\text{Population}}$  in each structure from the cochlear nucleus to the secondary auditory cortex, and showed that on average subcortical populations discriminate the original vocalizations better than cortical populations. A much larger number of neurons exhibited high values of MI in the subcortical structures whatever the temporal precision considered (see supplementary figure 2); as a consequence, the growth of  $MI_{\text{Population}}$  as a function of the number of recordings included in the population increased more rapidly in the subcortical structures (see supplementary figure 3). The higher temporal reliability of subcortical neurons (higher CorrCoef values) probably allows them to follow more precisely the stimulus temporal envelope and encode more accurately the between-stimuli differences, both at the individual and population level.

To a large extent, our results corroborate those of Chechick et al. (2006)<sup>25</sup> as we provide evidence that the neuronal discriminative abilities between communication sounds is higher in subcortical than in cortical structures. These authors showed that the MGB and AI responses contain 2-to-4 fold less information than the responses of IC neurons. Here, the neuronal

422 discriminative ability of the ventral division of the auditory thalamus (MGv) was closer than  
423 the ones displayed by the other subcortical structures. A potential explanation is that Chechick  
424 et al. (2006)<sup>25</sup> recorded from all divisions of the auditory thalamus, including the medial and  
425 dorsal divisions of the auditory thalamus, whereas our thalamic recordings were limited to the  
426 MGv and exhibited tonic responses to vocalizations similar to those observed in the CNIC  
427 and in the CN (see Fig. 2a and 3a). The stimulus sets also differ, as we used four utterances of  
428 the same category (the Whistle, an alarm call)<sup>26</sup>, whereas Chechick et al. (2006)<sup>25</sup> used three  
429 birds chirps and variants of these stimuli such as the stimuli's echoes or the background  
430 noises (15 stimuli in total), leading potentially to an easier classification between groups of  
431 stimuli compared to the present situation.

432 In our study, the original stimuli clearly differed in terms of AM patterns (i.e., their so-called  
433 “temporal envelope”) and, as a consequence, the most efficient way to discriminate them is  
434 probably to follow the time course of AM cues. It is well known that when progressing from  
435 the lower to the upper stages of the auditory system, the neurons' ability to follow AM  
436 changes considerably<sup>27,28</sup>. Brainstem neurons phase-lock on the sounds' AM pattern for AM  
437 rates up to hundreds of Hertz<sup>29,30</sup>, whereas thalamic neurons do so for a few tens of Hertz  
438 only<sup>31,32</sup> and cortical neurons for even lower rates<sup>33-35</sup>. As a consequence, subcortical neurons  
439 are able to follow faster changes in the AM patterns of the original vocalizations, in addition  
440 to being more sensitive to spectro-temporal details. This likely explains why subcortical  
441 neurons better discriminate the original stimuli both at the individual and population level.

442

#### 443 **Alterations of the slowest amplitude modulation cues explain changes in cortical and** 444 **subcortical discrimination**

445

446 Our main hypothesis is that the remarkable resistance of neural responses to the 0 dB SNR  
447 and the 10-band vocoded conditions results from the fact that AM cues were still sufficiently  
448 preserved. Previous studies using vocoded vocalizations reported little response changes at  
449 both the cortical and subcortical levels. In several species, cortical responses were not  
450 drastically reduced even when the number of frequency bands was reduced to two  
451 (marsomet<sup>36</sup>, gerbil<sup>37</sup>, rat<sup>38</sup>, guinea pig<sup>39</sup>). Most of the studies describing the effects of  
452 background noise on neuronal responses have been performed at the AI level, and many of  
453 them have pointed out the relationships between the noise impact on the cortical and  
454 behavioral discrimination performance. For example, in bird field L (homologous to AI),

455 neuronal responses to song motifs were strongly reduced by three types of masking noises,  
456 and the neural discriminative ability was progressively reduced when the SNR decreased, in  
457 parallel with the behavioral performance<sup>5</sup>. Our VRB results are reminiscent of those obtained  
458 in the bird homologue of a secondary area (area NCM) where feed-forward inhibition, which  
459 potentially contributes to reduce the evoked discharges of pyramidal cells, allows the  
460 emergence of invariant neural representations of target songs in noise conditions<sup>10</sup>.

461 In mammals, the discriminative abilities of AI responses to speech sounds presented in quiet  
462 or against background noise closely match behavioral performance<sup>40</sup>. As in here, Shetake et  
463 al. (2011)<sup>40</sup> did not find significant reduction in neural discrimination using an index of  
464 neuronal population performance (similar to MI) at a +12-dB SNR. Recent results revealed  
465 that responses of cortical neurons to calls could be classified in four classes - named robust,  
466 balanced, insensitive and brittle - when these calls were embedded in broadband white or  
467 babble noises<sup>8</sup>. In fact, the results of Bar-Yosef and Nelken (2007)<sup>41</sup> in the cat primary  
468 auditory cortex have already shown that some neurons are more sensitive to the noise  
469 background than to the actual target stimulus. Here, we observed that the  $MI_{\text{Individual}}$  and  
470  $MI_{\text{Population}}$  were only significantly reduced at a SNR of 0 dB, which indicates that on average,  
471 AI neurons were quite resistant to noise. This resistance is even higher in VRB where  
472  $MI_{\text{Individual}}$  did not significant decrease and where  $MI_{\text{Population}}$  significantly decreased only at a  
473 SNR of -10 dB. In the only study performed at the subcortical level, responses of IC neurons  
474 were found to be resistant to drastic spectral degradations<sup>42</sup>. Here, we show that both at the  
475 individual and population level, the temporal reliability and discriminative ability of neurons  
476 in three subcortical auditory structures (CN, CNIC and MGv) are slightly but significantly  
477 reduced in the 10-band vocoded condition.

478 If our main hypothesis is valid, in situations where AM cues are attenuated either by vocoding  
479 or by noise, the neuronal discrimination based upon AM cues should be largely reduced. We  
480 showed that, in both conditions, the discriminative abilities decreased in each auditory  
481 structure when the AM index ( $\Delta$ modulation index) was degraded by more than 25%.  
482 Therefore, these results indicate that accurate representation of slow AM cues is necessary at  
483 each level of the auditory system to discriminate efficiently communication sounds. More  
484 importantly, it appears that there is a limit to AM degradation beyond which the  
485 discriminative abilities decrease at the cortical and subcortical level. Thus, whatever the  
486 acoustic distortion used to degrade the amplitude modulations, equivalent AM alterations  
487 reduced the neuronal discrimination abilities to a similar extent. This is exactly what we  
488 observed for the Voc10 and 0 dB SNR conditions where degradations in the AM spectrum

489 were comparable and produced the same decrease in neuronal discrimination in each  
490 structure, except in the cochlear nucleus. In this structure, discriminative abilities were more  
491 sensitive to noise than to vocoding, suggesting that the stochastic fluctuations introduced by  
492 noise impact the responses in cochlear nucleus, but not in the upper stages of the auditory  
493 system.

494 Only one previous study directly compared the impact of vocoding and masking noise on  
495 cortical responses to vocalizations<sup>36</sup>. This study shares several characteristics with our study.  
496 First, auditory cortex neurons were found to be robust to spectral degradations since there was  
497 little change in evoked firing rate, even in response to 2-band vocoded vocalizations. Second,  
498 broadband white noise reduced neuronal responses at 0 dB SNR. Third, temporal-envelope  
499 degradations strongly reduced the evoked firing rate and the neural synchronization to the  
500 vocalization envelope. Importantly, bandpass filtering the vocalizations between 2-30 Hz did  
501 not reduce firing rate and neural synchronization to the vocalization envelope. This is in total  
502 agreement with our results: when the AM index ( $\Delta$ modulation index) - computed between 1  
503 and 20 Hz – revealed modest AM alterations, there was little effect on the neuronal  
504 discrimination, but when the AM alterations were larger than 25%, the neuronal responses  
505 and neuronal discrimination were reduced (Fig. 5b). Thus, our results confirm that at the  
506 cortical level, the key factors constraining auditory discrimination are the slowest (< 20 Hz)  
507 AM cues and, importantly, extend this conclusion to subcortical structures.

508 Direct comparison between the consequences of acoustic degradations in different auditory  
509 structures using the same set of stimuli, anaesthetic agent and methods to quantify neural  
510 discrimination is the more straightforward way for dissecting where invariant representations  
511 are generated. When measuring how different levels of noise alter neuronal coding in the  
512 auditory system, it was found that the neural representation of natural sounds becomes  
513 progressively independent of the level of background noise from the auditory nerve till the IC  
514 and AI<sup>23</sup>. It was proposed that at the population level, this tolerance to background noise  
515 results from an adaptation to the noise statistics, which is much more pronounced at the  
516 cortical than at the subcortical level<sup>23</sup>. In agreement with this study, we found that populations  
517 of cortical neurons (AI and VRB) were more resistant to noise than subcortical ones.  
518 However, we did not observe a monotonic evolution of resistance to noise in the auditory  
519 system: at the subcortical level, the discrimination abilities of CN neuronal populations  
520 drastically dropped for a +10 dB SNR and those of thalamic ones largely decreased at a -10  
521 dB SNR, whereas populations of CNIC cells maintained relatively good discriminative

522 abilities, suggesting that they were the more resistant to noise, even in the most adverse  
523 conditions. In the IC, previous work showed that background noise changes the shape of the  
524 temporal modulation transfer function of individual neurons from bandpass to lowpass<sup>24</sup>. The  
525 CNIC is a massive hub receiving probably the highest diversity of inhibitory and excitatory  
526 inputs<sup>43,44</sup> and potentially the large diversity of these inputs allows this structure to extract  
527 crucial temporal information about the stimulus' temporal envelope, even at relatively low  
528 SNR.

529

### 530 **General conclusions**

531 The present study led to two major findings with regard to the main factors influencing  
532 auditory processing in noise, and the respective contributions of auditory structures to robust  
533 sound coding in noise.

534 Comparison of neural data collected in response to noisy versus vocoded vocalizations  
535 clarified a long-lasting debate<sup>11;12,16,45</sup>: from a neural perspective, the *main* effect of  
536 (notionally) steady background noise on complex-sound discrimination corresponds to the  
537 attenuation of the gross AM (i.e., “temporal envelope”) cues conveyed by sounds. Corruption  
538 of FM cues and introduction of stochastic fluctuations in AM power have little influence if  
539 any on neural discrimination in noise (with the noticeable exception of the cochlear nucleus  
540 showing strong sensitivity to stochasticity). This is in accordance with objective measures  
541 currently used by audio engineers to predict the intelligibility and perceived quality of speech  
542 masked by noise<sup>13, 45</sup>.

543 Inconsistent with our initial expectations, the ability of auditory neurons to discriminate  
544 between communication sounds masked by external noise neither increased nor decreased  
545 along the auditory pathway from the first auditory relay up to the primary and secondary  
546 cortical areas, but culminated at the collicular and thalamic levels. In humans, speech sounds  
547 (such as phonemes) showing similar acoustic properties trigger similar responses and are  
548 represented as a single category in the superior temporal gyrus<sup>4</sup>. Here, the use of vocalizations  
549 belonging to the same category of the communication repertoire of guinea pigs, i.e.  
550 “whistles”, may explain both the relatively poor discriminative abilities of cortical neurons  
551 compared to subcortical ones and the robustness of cortical responses to vocoding and  
552 background noise. In these two challenging situations, our results reveal that cortical neurons  
553 resist more than the other auditory structures, potentially because cortical neurons do not code  
554 for the spectro-temporal details of the stimuli but rather respond to more abstract stimulus  
555 properties<sup>21</sup> carried by the gross spectro-temporal envelope patterns.

556 In a recent study, auditory cortex responses collected in behaving ferrets were found to be  
557 sufficiently robust to preserve vowel identity across a large range of acoustic transformations,  
558 such as changes in fundamental frequency, sound location or level<sup>9</sup>. It is notable that earlier  
559 studies from the same laboratory performed in anaesthetized conditions<sup>46,47</sup> have reached very  
560 similar conclusions for vowels varying in fundamental frequency and virtual acoustic  
561 location, indicating that the general principles allowing neuronal discrimination are  
562 observable across anesthetized and behavioral states.

563 Our results extend these findings by showing that downstream from AI, neurons in a  
564 secondary auditory area (VRB) are even more resistant to spectral degradations than in AI.  
565 This is in line with the results of Carruthers et al. (2015)<sup>7</sup> showing that in the secondary  
566 auditory cortex (the SRAF area), neuronal populations code invariant representations of  
567 conspecific vocalizations despite important spectro-temporal degradations. As already  
568 proposed<sup>21</sup> by Chechick and Nelken (2012)<sup>21</sup>, auditory cortex neurons extract abstract  
569 auditory entities rather than detailed spectro-temporal features. This suggests that even when  
570 the four vocalizations were vocoded, the mere fact that gross spectro-temporal envelope cues  
571 were preserved in a limited number of frequency bands was sufficient for auditory cortex  
572 neurons to classify these stimuli as belonging to the alarm-call category.

573 Neurons in subcortical structures were able to discriminate better the stimuli than the neurons  
574 in cortical areas when temporal envelope cues were degraded even in the most severe  
575 condition (>25% of alterations in the -10 dB SNR condition). Indeed, the most accurate  
576 representations of target stimuli (and of their differences) were found at the collicular and  
577 thalamic levels, not at the cortical level. In each condition, the identification of an auditory  
578 object necessarily involves both subcortical and cortical processing. We therefore suggest that  
579 in challenging conditions, cortical representations co-exist with more detailed representations  
580 of the stimuli in one, or several, subcortical structures. In fact, in the most adverse noise  
581 condition (at a -10 dB SNR) where temporal envelope cues were strongly degraded, CNIC  
582 neurons still exhibited discrimination abilities whereas the other subcortical and cortical  
583 neurons did not (a value of 1 for the  $MI_{\text{Population}}$  indicates that 2 out of 4 stimuli could still be  
584 discriminated). Further studies are required to determine to what extent these incomplete  
585 subcortical representations influence auditory abilities in animals and humans.

586



587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632

## References

1. Bidelman, G. M., Davis, M. K. & Pridgen, M. H. Brainstem-cortical functional connectivity for speech is differentially challenged by noise and reverberation. *Hearing Research* **367**, 149–160 (2018).
2. Teng, S., Sommer, V. R., Pantazis, D. & Oliva, A. Hearing Scenes: A Neuromagnetic Signature of Auditory Source and Reverberant Space Separation. *eNeuro* **4**, (2017).
3. Fuglsang, S. A., Dau, T. & Hjortkjær, J. Noise-robust cortical tracking of attended speech in real-world acoustic scenes. *Neuroimage* **156**, 435–444 (2017).
4. Mesgarani, N., David, S. V., Fritz, J. B. & Shamma, S. A. Mechanisms of noise robust representation of speech in primary auditory cortex. *Proceedings of the National Academy of Sciences* **111**, 6792–6797 (2014).
5. Narayan, R. *et al.* Cortical interference effects in the cocktail party problem. *Nature Neuroscience* **10**, 1601–1607 (2007).
6. Beetz, M. J., García-Rosales, F., Kössl, M. & Hechavarría, J. C. Robustness of cortical and subcortical processing in the presence of natural masking sounds. *Scientific Reports* **8**, (2018).
7. Carruthers, I. M. *et al.* Emergence of invariant representation of vocalizations in the auditory cortex. *Journal of Neurophysiology* **114**, 2726–2740 (2015).
8. Ni, R., Bender, D. A., Shanechi, A. M., Gamble, J. R. & Barbour, D. L. Contextual effects of noise on vocalization encoding in primary auditory cortex. *Journal of Neurophysiology* **117**, 713–727 (2017).
9. Town SM, Wood KC, Bizley JK. Sound identity is represented robustly in auditory cortex during perceptual constancy. *Nat Commun.* 9(1):4786 (2018).
10. Schneider, D. M. & Woolley, S. M. N. Sparse and Background-Invariant Coding of Vocalizations in Auditory Scenes. *Neuron* **79**, 141–152 (2013).
11. Noordhoek, I. M., and Drullman, R. Effect of reducing temporal intensity modulations on sentence intelligibility. *J. Acoust. Soc. Am.* **101**, 498–502 (1997).
12. Dubbelboer, F. & Houtgast, T. A detailed study on the effects of noise on speech intelligibility. *The Journal of the Acoustical Society of America* **122**, 2865 (2007).
13. Houtgast T, Steeneken H. A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. *The Journal of the Acoustical Society of America.* **77** (3): 1069–1077 (1985).
14. Ewert SD, Dau T. Characterizing frequency selectivity for envelope fluctuations. *J Acoust Soc Am.* 108(3 Pt 1):1181-96 (2000).
15. Biberger T, Ewert SD. The role of short-time intensity and envelope power for speech intelligibility and psychoacoustic masking. *J Acoust Soc Am.* 142(2):1098 (2017);
16. Shamma S, Lorenzi C. On the balance of envelope and temporal fine structure in the encoding of speech in the early auditory system. *J Acoust Soc Am* 133(5):2818–2833 (2013).
17. Varnet, L., Ortiz-Barajas, M. C., Erra, R. G., Gervain, J. & Lorenzi, C. A cross-linguistic study of speech modulation spectra. *The Journal of the Acoustical Society of America* **142**, 1976–1989 (2017).
18. Wang, X., Merzenich, M. M., Beitel, R. & Schreiner, C. E. Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. *Journal of Neurophysiology* **74**, 2685–2706 (1995).
19. Wang, X. & Kadia, S. C. Differential Representation of Species-Specific Primate Vocalizations in the Auditory Cortices of Marmoset and Cat. *Journal of Neurophysiology* **86**, 2616–2620 (2001).

- 633 20. Nelken I, Bar-Yosef O. Neurons and objects: the case of auditory cortex. *Front Neurosci.*  
634 2(1):107-13. (2008)
- 635 21. Chechick G, Nelken I. Auditory abstraction from spectro- temporal features to coding auditory  
636 entities. *Proc Natl Acad Sci U S A* 109(46):18968–18973 (2012).
- 637 22. Kates, J. M. Spectro-temporal envelope changes caused by temporal fine structure modification.  
638 *The Journal of the Acoustical Society of America* **129**, 3981–3990 (2011).
- 639 23. Rabinowitz, N. C., Willmore, B. D. B., King, A. J. & Schnupp, J. W. H. Constructing Noise-  
640 Invariant Representations of Sound in the Auditory Pathway. *PLoS Biology* **11**, e1001710 (2013).
- 641 24. Lesica, N. A. & Grothe, B. Efficient Temporal Processing of Naturalistic Sounds. *PLoS ONE* **3**,  
642 e1655 (2008).
- 643 25. Chechik, G. *et al.* Reduction of Information Redundancy in the Ascending Auditory Pathway.  
644 *Neuron* **51**, 359–368 (2006).
- 645 26. Beryman JC. Guinea-Pig Vocalizations - Their Structure, Causation and Function. *Z Tierpsychol.*  
646 41(1):80–106 (1976).
- 647 27. Joris, P. X., Schreiner, C. E. & Rees, A. Neural processing of amplitude-modulated sounds.  
648 *Physiol. Rev.* **84**, 541–577 (2004).
- 649 28. Escabí, M. A. & Read, H. L. Neural Mechanisms for Spectral Analysis in the Auditory Midbrain,  
650 Thalamus, and Cortex. in *International Review of Neurobiology* **70**, 207–252 (2005).
- 651 29. Frisina, R. D., Smith, R. L. & Chamberlain, S. C. Encoding of amplitude modulation in the gerbil  
652 cochlear nucleus: I. A hierarchy of enhancement. *Hear. Res.* **44**, 99–122 (1990).
- 653 30. Rhode, W. S. & Greenberg, S. Encoding of amplitude modulation in the cochlear nucleus of the  
654 cat. *J. Neurophysiol.* **71**, 1797–1825 (1994).
- 655 31. Creutzfeldt, O., Hellweg, F.-C. & Schreiner, C. Thalamocortical transformation of responses to  
656 complex auditory stimuli. *Experimental Brain Research* **39**, (1980).
- 657 32. Preuss, A. & Müller-Preuss, P. Processing of amplitude modulated sounds in the medial  
658 geniculate body of squirrel monkeys. *Exp Brain Res* **79**, 207–211 (1990).
- 659 33. Gaese, B. H. & Ostwald, J. Temporal coding of amplitude and frequency modulation in the rat  
660 auditory cortex. *Eur. J. Neurosci.* **7**, 438–450 (1995).
- 661 34. Liang, L., Lu, T. & Wang, X. Neural representations of sinusoidal amplitude and frequency  
662 modulations in the primary auditory cortex of awake primates. *J. Neurophysiol.* **87**, 2237–2261  
663 (2002).
- 664 35. Schreiner, C. E. & Urbas, J. V. Representation of amplitude modulation in the auditory cortex of  
665 the cat. II. Comparison between cortical fields. *Hear. Res.* **32**, 49–63 (1988).
- 666 36. Nagarajan, S. S. *et al.* Representation of Spectral and Temporal Envelope of Twitter Vocalizations  
667 in Common Marmoset Primary Auditory Cortex. *Journal of Neurophysiology* **87**, 1723–1737  
668 (2002).
- 669 37. Ter-Mikaelian, M., Semple, M. N. & Sanes, D. H. Effects of spectral and temporal disruption on  
670 cortical encoding of gerbil vocalizations. *Journal of Neurophysiology* **110**, 1190–1204 (2013).
- 671 38. Ranasinghe, K. G., Vrana, W. A., Matney, C. J. & Kilgard, M. P. Neural Mechanisms Supporting  
672 Robust Discrimination of Spectrally and Temporally Degraded Speech. *Journal of the Association*  
673 *for Research in Otolaryngology* **13**, 527–542 (2012).
- 674 39. Aushana, Y., Souffi, S., Edeline, J.-M., Lorenzi, C. & Huetz, C. Robust Neuronal Discrimination  
675 in Primary Auditory Cortex Despite Degradations of Spectro-temporal Acoustic Details:  
676 Comparison Between Guinea Pigs with Normal Hearing and Mild Age-Related Hearing Loss. *J.*  
677 *Assoc. Res. Otolaryngol.* **19**, 163–180 (2018).
- 678 40. Shetake, J. A. *et al.* Cortical activity patterns predict robust speech discrimination ability in noise.  
679 *Eur. J. Neurosci.* **34**, 1823–1838 (2011).

- 680 41. Bar-Yosef, O. & Nelken, I. The effects of background noise on the neural responses to natural  
681 sounds in cat primary auditory cortex. *Front Comput Neurosci* **1**, 3 (2007).
- 682 42. Ranasinghe, K. G., Vrana, W. A., Matney, C. J. & Kilgard, M. P. Increasing diversity of neural  
683 responses to speech sounds across the central auditory pathway. *Neuroscience* **252**, 80–97 (2013).
- 684 43. Malmierca, M. S. The Inferior Colliculus: A Center for Convergence of Ascending and  
685 Descending Auditory Information. *Neuroembryology and Aging* **3**, 215–229 (2004).
- 686 44. Ayala, Y. A., Pérez-González, D. & Malmierca, M. S. Stimulus-specific adaptation in the inferior  
687 colliculus: The role of excitatory, inhibitory and modulatory inputs. *Biological Psychology* **116**,  
688 10–22 (2016).
- 689 45. Kates, J. M. & Arehart, K. H. The Hearing-Aid Speech Perception Index (HASPI). *Speech*  
690 *Communication* **65**, 75–93 (2014).
- 691 46. Bizley, J. K., Walker, K. M. M., Silverman, B. W., King, A. J. & Schnupp, J. W. H.  
692 Interdependent Encoding of Pitch, Timbre, and Spatial Location in Auditory Cortex. *Journal of*  
693 *Neuroscience* **29**, 2064–2075 (2009).
- 694 47. Walker, K. M. M., Bizley, J. K., King, A. J. & Schnupp, J. W. H. Multiplexed and Robust  
695 Representations of Sound Features in Auditory Cortex. *Journal of Neuroscience* **31**, 14565–14576  
696 (2011).
- 697 48. Gourévitch, B., Doisy, T., Avillac, M. & Edeline, J.-M. Follow-up of latency and threshold shifts  
698 of auditory brainstem responses after single and interrupted acoustic trauma in guinea pig. *Brain*  
699 *Res.* **1304**, 66–79 (2009).
- 700 49. Gourévitch, B. & Edeline, J.-M. Age-related changes in the guinea pig auditory cortex:  
701 relationship with brainstem changes and comparison with tone-induced hearing loss. *Eur. J.*  
702 *Neurosci.* **34**, 1953–1965 (2011).
- 703 50. Redies, H., Brandner, S. & Creutzfeldt, O. D. Anatomy of the auditory thalamocortical system of  
704 the guinea pig. *The Journal of Comparative Neurology* **282**, 489–511 (1989).
- 705 51. Anderson, L. A., Wallace, M. N. & Palmer, A. R. Identification of subdivisions in the medial  
706 geniculate body of the guinea pig. *Hearing Research* **228**, 156–167 (2007).
- 707 52. Wallace, M. N., Anderson, L. A. & Palmer, A. R. Phase-Locked Responses to Pure Tones in the  
708 Auditory Thalamus. *Journal of Neurophysiology* **98**, 1941–1952 (2007).
- 709 53. Parouty, N., Stasiak, A., Lorenzi, C., Varnet, L. & Winter, I. M. Dual Coding of Frequency  
710 Modulation in the Ventral Cochlear Nucleus. *J. Neurosci.* **38**, 4123–4137 (2018).
- 711 54. Edeline, J. M. & Weinberger, N. M. Receptive field plasticity in the auditory cortex during  
712 frequency discrimination training: selective retuning independent of task difficulty. *Behav.*  
713 *Neurosci.* **107**, 82–103 (1993).
- 714 55. Manunta, Y. & Edeline, J. M. Effects of noradrenaline on frequency tuning of auditory cortex  
715 neurons during wakefulness and slow-wave sleep. *Eur. J. Neurosci.* **11**, 2134–2150 (1999).
- 716 56. Edeline, J. M., Dutrieux, G., Manunta, Y. & Hennevin, E. Diversity of receptive field changes in  
717 auditory cortex during natural sleep. *Eur. J. Neurosci.* **14**, 1865–1880 (2001).
- 718 57. Wallace, M. N., Rutkowski, R. G. & Palmer, A. R. Identification and localisation of auditory areas  
719 in guinea pig cortex. *Experimental Brain Research* **132**, 445–456 (2000).
- 720 58. Grimsley, J. M. S., Shanbhag, S. J., Palmer, A. R. & Wallace, M. N. Processing of  
721 Communication Calls in Guinea Pig Auditory Cortex. *PLoS ONE* **7**, e51646 (2012).
- 722 59. Rutkowski, R. G., Shackleton, T. M., Schnupp, J. W. H., Wallace, M. N. & Palmer, A. R.  
723 Spectrotemporal Receptive Field Properties of Single Units in the Primary, Dorsocaudal and  
724 Ventrorostral Auditory Cortex of the Guinea Pig. *Audiology and Neurotology* **7**, 214–227 (2002).
- 725 60. Gaucher, Q., Edeline, J.-M. & Gourévitch, B. How different are the local field potentials and  
726 spiking activities? Insights from multi-electrodes arrays. *J. Physiol. Paris* **106**, 93–103 (2012).

- 727 61. Occelli, F., Suied, C., Pressnitzer, D., Edeline, J.-M. & Gourévitch, B. A Neural Substrate for  
728 Rapid Timbre Recognition? Neural and Behavioral Discrimination of Very Brief Acoustic  
729 Vowels. *Cereb. Cortex* **26**, 2483–2496 (2016).
- 730 62. Gnansia, D., Péan, V., Meyer, B. & Lorenzi, C. Effects of spectral smearing and temporal fine  
731 structure degradation on speech masking release. *J. Acoust. Soc. Am.* **125**, 4023–4033 (2009).
- 732 63. Gnansia, D., Pressnitzer, D., Péan, V., Meyer, B. & Lorenzi, C. Intelligibility of interrupted and  
733 interleaved speech for normal-hearing listeners and cochlear implantees. *Hearing Research* **265**,  
734 46–53 (2010).
- 735 64. Stone, M. A., Füllgrabe, C., Mackinnon, R. C. & Moore, B. C. J. The importance for speech  
736 intelligibility of random fluctuations in ‘steady’ background noise. *J. Acoust. Soc. Am.* **130**, 2874–  
737 2881 (2011).
- 738 65. Patterson, R. D. A pulse ribbon model of monaural phase perception. *J. Acoust. Soc. Am.* **82**,  
739 1560–1586 (1987).
- 740 66. Sayles, M. & Winter, I. M. Equivalent-rectangular bandwidth of single units in the anaesthetized  
741 guinea-pig ventral cochlear nucleus. *Hear. Res.* **262**, 26–33 (2010).
- 742 67. Hohmann, V. Frequency analysis and synthesis using a Gammatone filterbank. *Acust Acta Acust*  
743 **88**:433–442 (2002)
- 744 68. Wallace, M. N. & Palmer, A. R. Laminar differences in the response properties of cells in the  
745 primary auditory cortex. *Exp Brain Res* **184**, 179–191 (2008).
- 746 69. Gaucher, Q., Huetz, C., Gourévitch, B. & Edeline, J.-M. Cortical inhibition reduces information  
747 redundancy at presentation of communication sounds in the primary auditory cortex. *J. Neurosci.*  
748 **33**, 10713–10728 (2013a).
- 749 70. Huetz, C., Philibert, B. & Edeline, J.-M. A spike-timing code for discriminating conspecific  
750 vocalizations in the thalamocortical system of anesthetized and awake guinea pigs. *J. Neurosci.*  
751 **29**, 334–350 (2009).
- 752 71. Gaucher, Q. & Edeline, J.-M. Stimulus-specific effects of noradrenaline in auditory cortex:  
753 implications for the discrimination of communication sounds. *J. Physiol. (Lond.)* **593**, 1003–1020  
754 (2015).
- 755 72. Schnupp, J. W. H., Hall, T. M., Kokelaar, R. F. & Ahmed, B. Plasticity of temporal pattern codes  
756 for vocalization stimuli in primary auditory cortex. *J. Neurosci.* **26**, 4785–4795 (2006).
- 757 73. Shannon, C. E. (1948) A mathematical theory of communication. *Bell Syst Tech J* 27(3):379–423.  
758

759

## Materials and Methods

### 760 **Subjects**

761 These experiments were performed under the national license A-91-557 (project 2014-25,  
762 authorization 05202.02) and using the procedures N° 32-2011 and 34-2012 validated by the  
763 Ethic committee N°59 (CEEA Paris Centre et Sud). All surgical procedures were performed  
764 in accordance with the guidelines established by the European Communities Council  
765 Directive (2010/63/EU Council Directive Decree).

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

772 Two to three days before each experiment, the animal's pure-tone audiogram was determined  
773 by testing auditory brainstem responses (ABR) under isoflurane anaesthesia (2.5 %) as  
774 described in Gourévitch et al. (2009)<sup>48</sup>. The ABR was obtained by differential recordings  
775 between two subdermal electrodes (SC25-NeuroService) placed at the vertex and behind the  
776 mastoid bone. A software (RTLlab, Echodia, Clermont-Ferrand, France) allowed averaging  
777 500 responses during the presentation of nine pure-tone frequencies (between 0.5 and 32 kHz)  
778 delivered by a speaker (Knowles Electronics) placed in the animal right ear. The auditory  
779 threshold of each ABR was the lowest intensity where a small ABR wave can still be detected  
780 (usually wave III). For each frequency, the threshold was determined by gradually decreasing  
781 the sound intensity (from 80 dB down to -10 dB SPL). All animals used in this study had  
782 normal pure-tone audiograms<sup>39,48,49</sup>.

### 783 **Surgical procedures**

784 All animals were anesthetized by an initial injection of urethane (1.2 g/kg, i.p.) supplemented  
785 by additional doses of urethane (0.5 g/kg, i.p.) when reflex movements were observed after  
786 pinching the hind paw (usually 2-4 times during the recording session). A single dose of  
787 atropine sulphate (0.06mg/kg, s.c.) was given to reduce bronchial secretions and a small dose  
788 of buprenorphine was administrated (0.05mg/kg, s.c.) as urethane has no antalgic properties.

789 After placing the animal in a stereotaxic frame, a craniotomy was performed and a local  
790 anesthetic (Xylocain 2%) was liberally injected in the wound.

791 For auditory cortex recordings (area A1 and VRB), a craniotomy was performed above the  
792 left temporal cortex. The opening was 8 mm wide starting at the intersection point between  
793 parietal and temporal bones and 8-10 mm height. The dura above the auditory cortex was  
794 removed under binocular control and the cerebrospinal fluid was drained through the cisterna  
795 to prevent the occurrence of oedema.

796 For the recordings in MGB, a craniotomy was performed the most posterior part of the MGB  
797 (8mm posterior to Bregma) to reach the left auditory thalamus at a location where the MGB is  
798 mainly composed of its ventral, tonotopic, part<sup>50-52</sup>.

799 For IC recordings, a craniotomy was performed above the IC and portions of the cortex were  
800 aspirated to expose the surface of the left IC. For CN recordings, after opening the skull above  
801 the right cerebellum, portions of the cerebellum were aspirated to expose the surface of the  
802 right CN<sup>53</sup>.

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

## 808 **Recording procedures**

809 Data were from multi-unit recordings collected in 5 auditory structures, the non-primary  
810 cortical area VRB, the primary cortical area A1, the medial geniculate body (MGB), the  
811 inferior colliculus (IC) and the cochlear nucleus (CN). Cortical extracellular recordings were  
812 obtained from arrays of 16 tungsten electrodes ( $\varnothing$ : 33  $\mu\text{m}$ , <1 M $\Omega$ ) composed of two rows of 8  
813 electrodes separated by 1000  $\mu\text{m}$  (350  $\mu\text{m}$  between electrodes of the same row). A silver wire,  
814 used as ground, was inserted between the temporal bone and the dura matter on the  
815 contralateral side. The location of the primary auditory cortex was estimated based on the  
816 pattern of vasculature observed in previous studies<sup>54-57</sup>. The non-primary cortical area VRB  
817 was located ventral to A1 and distinguished because of its long latencies to pure tones<sup>58,59</sup>. For  
818 each experiment, the position of the electrode array was set in such a way that the two rows of  
819 eight electrodes sample neurons responding from low to high frequency when progressing in

820 the rostro-caudal direction [see examples of tonotopic gradients recorded with such arrays in  
821 figure 1 of Gaucher et al. (2012)<sup>60</sup> and in figure 6A of Occelli et al. (2016)<sup>61</sup>.

822 All the remaining extracellular recordings (in MGB, IC and CN) were obtained using 16  
823 channel multi-electrode arrays (NeuroNexus) composed of one shank (10 mm) of 16  
824 electrodes spaced by 110  $\mu\text{m}$  and with conductive site areas of  $177\mu\text{m}^2$ . The electrodes were  
825 advanced vertically (for MGB and IC) or with a  $40^\circ$  angle (for CN) until evoked responses to  
826 pure tones could be detected on at least 10 electrodes.

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

834 The raw signal was amplified 10,000 times (TDT Medusa). It was then processed by an RX5  
835 multichannel data acquisition system (TDT). The signal collected from each electrode was  
836 filtered (610-10000 Hz) to extract multi-unit activity (MUA). The trigger level was set for  
837 each electrode to select the largest action potentials from the signal. On-line and off-line  
838 examination of the waveforms suggests that the MUA collected here was made of action  
839 potentials generated by 2 to 6 neurons in the vicinity of the electrode.

#### 840 **Acoustic stimuli**

841 Acoustic stimuli were generated using MatLab, transferred to a RP2.1-based sound delivery  
842 system (TDT) and sent to a Fostex speaker (FE87E). The speaker was placed at 2 cm from the  
843 guinea pig's right ear, a distance at which the speaker produced a flat spectrum ( $\pm 3$  dB)  
844 between 140 Hz and 36 kHz. Calibration of the speaker was made using noise and pure tones  
845 recorded by a Bruel & Kjaer microphone 4133 coupled to a preamplifier B&K 2169 and a  
846 digital recorder Marantz PMD671.

847 Spectro-temporal receptive fields (STRFs) were first determined using 97 or 129 pure-tones  
848 frequencies scaled with a gamma function, covering six (0.14-9 kHz or 0.28-18 kHz or 0.56-  
849 36 kHz) or eight (0.14-36 kHz) octaves respectively, and presented at 75 dB SPL. At a given  
850 level, each frequency was repeated eight times at a rate of 2.35 Hz in pseudorandom order.



851 The duration of these tones over half-peak amplitude was 15 ms and the total duration of the  
852 tone was 50 ms, so there was no overlap between tones.

853 A set of four conspecific vocalizations was used to assess the neuronal responses to  
854 communication sounds. These vocalizations were recorded from animals of our colony. Pairs  
855 of animals were placed in the acoustic chamber and their vocalizations were recorded by a  
856 Bruel & Kjaer microphone 4133 coupled to a preamplifier B&K 2169 and a digital recorder  
857 Marantz PMD671. A large set of whistle calls was loaded in the Audition software (Adobe  
858 Audition 3) and four representative examples of whistle were selected. As shown in figure 1a  
859 (lower panels), despite the fact the maximal energy of the four selected whistle was in the  
860 same frequency range (typically between 4 and 26 kHz), these calls displayed slight  
861 differences in their spectrograms. In addition, their temporal (amplitude) envelopes clearly  
862 differed as shown by their waveforms (Fig. 1a, upper panels). The four selected whistles were  
863 processed by three tone vocoders<sup>62,63</sup>. In the following figures, the unprocessed whistles will  
864 be referred to as the original versions, and the vocoded versions as Voc38 (Voc20, Voc10  
865 respectively) for the 38-band (20-band, 10-band, respectively) vocoded whistles. In contrast  
866 to previous studies that used noise-excited vocoders<sup>36-38</sup>, a tone vocoder was used here,  
867 because noise vocoders were found to introduce random (i.e., non-informative) intrinsic  
868 temporal-envelope fluctuations distorting the crucial spectro-temporal modulation features of  
869 communication sounds<sup>16, 22,64</sup>.

870 Figure 1b displays the spectrograms of the 38-band vocoded (first row), the 20-band vocoded  
871 (second row) and the 10-band vocoded (third row) of the four whistles. The three vocoders  
872 differed only in terms of the number of frequency bands (i.e., analysis filters) used to  
873 decompose the whistles (38, 20 or 10 bands). The 38-band vocoding process is briefly  
874 described below, but the same principles apply to the 20-band or the 10-band vocoders. Each  
875 digitized signal was passed through a bank of 38 fourth-order Gammatone filters<sup>65</sup> with center  
876 frequencies uniformly spaced along a guinea-pig adapted ERB (Equivalent Rectangular  
877 Bandwidth) scale<sup>66</sup> ranging from 50 to 35505 Hz. In each frequency band, the temporal  
878 envelope was extracted using full-wave rectification and lowpass filtering at 64 Hz with a  
879 zero-phase, sixth-order Butterworth filter. The resulting envelopes were used to amplitude  
880 modulate sine-wave carriers with frequencies at the center frequency of the Gammatone  
881 filters, and with random starting phase. Impulse responses were peak-aligned for the envelope  
882 (using a group delay of 16 ms) and the acoustic temporal fine structure across frequency  
883 channels<sup>67</sup>. The modulated signals were finally weighted and summed over the 38 frequency  
884 bands. The weighting compensated for imperfect superposition of the bands' impulse

885 responses at the desired group delay. The weights were optimized numerically to achieve a  
886 flat frequency response. Amplitude-modulation (AM) spectra were computed for the original  
887 and vocoded versions of each vocalization and the averaged modulation spectra are displayed  
888 in figure 1d (left panel). AM spectra were computed<sup>17</sup> by decomposing each sound using a  
889 bank of 50 (spanning the range 0.1-50 kHz), 1 ERB-wide Gammatone filters and then  
890 analyzing the temporal envelope in each frequency band through a bank of AM filters (1/3-  
891 octave wide first-order Butterworth bandpass filters overlapping at -3 dB, with center  
892 frequencies between 0.1 Hz and 410 Hz). The root-mean-square amplitude of the filtered  
893 output was multiplied by a factor of 1.414. For each AM filter, a modulation index was  
894 calculated by dividing the output by the mean amplitude of the AM component for the  
895 vocalization sample in a given Gammatone filter. Finally, the 50 AM spectra were averaged  
896 to generate a single AM spectrum per stimulus. All vocalizations were presented at 75dB  
897 SPL.

898 The four whistles were also presented in a frozen stationary background noise ranging from  
899 100-30000 Hz. The first three rows of figure 1c display the spectrograms of the four whistles  
900 in the stationary noise with a SNR of +10 dB SPL, 0 dB SPL, -10 dB SPL and the last row  
901 shows the masking noise only. The alterations of the AM spectra produced by masking noise  
902 on the original vocalizations are displayed in figure 1d (right panel). To quantify the  
903 alterations of the temporal envelope induced either by the noise addition or by the vocoding,  
904 we calculated the mean percentage of variation of the modulation index ( $\Delta$ modulation index,  
905 computed from 1-20 Hz) between each acoustic condition (noise or vocoded condition) and  
906 the original condition (Fig. 5b).

## 907 **Experimental protocol**

908 As inserting an array of 16 electrodes in a brain structure almost systematically induces a  
909 deformation of this structure, a 30-minutes recovering time lapse was allowed for the  
910 structure to return to its initial shape, then the array was slowly lowered. Tests based on  
911 measures of spectro-temporal receptive fields (STRFs) were used to assess the quality of our  
912 recordings and to adjust electrodes' depth. For auditory cortex recordings (AI and VRB), the  
913 recording depth was 500-1000  $\mu$ m, which corresponds to layer III and the upper part of layer  
914 IV according to Wallace and Palmer (2008)<sup>68</sup>. For thalamic recordings, the NeuroNexus probe  
915 was lowered of about 7mm below pia before the first responses to pure tones were detected.

916 When a clear frequency tuning was obtained for at least 10 of the 16 electrodes, the stability  
917 of the tuning was assessed: we required that the recorded neurons displayed at least three  
918 successive similar STRFs (each lasting 6 minutes) before starting the protocol. When the  
919 stability was satisfactory, the protocol was started by presenting the acoustic stimuli in the  
920 following order: We first presented the 4 whistles at 75 dB SPL in their natural versions,  
921 followed by the vocoded versions with 38, 20 and 10 bands. The same set of original whistles  
922 was then presented in the stationary noise presented at 65, 75 and 85 dB SPL. In each case,  
923 each vocalization was repeated 20 times. Presentation of this entire stimulus set lasted 45  
924 minutes. The protocol was re-started either after moving the electrode arrays on the cortical  
925 map or after lowering the electrode at least by 300  $\mu\text{m}$  for subcortical structures.

## 926 **Data analysis**

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

928 The STRFs derived from MUA were obtained by constructing post-stimulus time histograms  
929 for each frequency with 1 ms time bins. The firing rate evoked by each frequency was  
930 quantified by summing all the action potentials from the tone onset up to 100 ms after this  
931 onset. Thus, STRFs are matrices of 100 bins in abscissa (time) multiplied by 97 or 129 bins in  
932 ordinate (frequency). All STRFs were smoothed with a uniform 5x5 bin window.

933 For each STRF, the Best Frequency (BF) was defined as the frequency at which the highest  
934 firing rate was recorded. Peaks of significant response were automatically identified using the  
935 following procedure: A positive peak in the MU-based STRF was defined as a contour of  
936 firing rate above the average level of the baseline activity (estimated from the ten first  
937 milliseconds of STRFs at all intensity levels) plus six times the standard deviation of the  
938 baseline activity.

### 939 *Quantification of responses evoked by vocalizations*

940 The responses to vocalizations were quantified using two parameters: (i) The firing rate of the  
941 evoked response, which corresponds to the total number of action potentials occurring during  
942 the presentation of the stimulus minus spontaneous activity; (ii) the spike-timing reliability  
943 coefficient (CorrCoef) which quantifies the trial-to-trial reliability of the response. This index  
944 was computed for each vocalization: it corresponds to the normalized covariance between  
945 each pair of spike trains recorded at presentation of this vocalization and was calculated as  
946 follows:

947

$$\text{CorrCoef} = \frac{1}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{\sigma_{x_i x_j}}{\sigma_{x_i} \sigma_{x_j}}$$

948 where N is the number of trials and  $\sigma_{x_i x_j}$  is the normalized covariance at zero lag between  
949 spike trains  $x_i$  and  $x_j$  where i and j are the trial numbers. Spike trains  $x_i$  and  $x_j$  were previously  
950 convolved with a 10-msec width Gaussian window. Based upon computer simulations, we  
951 have previously shown that this CorrCoef index is not a function of the neurons' firing rate<sup>69</sup>.  
952 We have computed the CorrCoef index with a Gaussian window ranging from 1 to 50 ms to  
953 determine if the selection of a particular value for the Gaussian window influences the  
954 difference in CorrCoef mean values obtained in the different auditory structures. As a general  
955 rule, the largest the Gaussian window, the largest the CorrCoef value whatever the structure  
956 was. Based upon the responses to the original vocalizations, supplementary Figure S1A shows  
957 that the relative ranking between auditory structures remained unchanged whatever the size of  
958 the Gaussian window was. Therefore, we kept the value of 10 ms for the Gaussian window  
959 (dashed line in figure S1) as it was used in several previous studies<sup>39, 69-71</sup>.

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

961 The method developed by Schnupp et al. (2006)<sup>72</sup> was used to quantify the amount of  
962 information (Shannon 1948) contained in the responses to vocalizations obtained with natural  
963 vocoded and noise stimuli. This method allows quantifying how well the vocalization's  
964 identity can be inferred from neuronal responses. Here, "neuronal responses" refer either to (i)  
965 the spike trains obtained from a small group of neurons below one electrode (for the  
966 computation of the individual Mutual Information,  $MI_{\text{Individual}}$ ), or to (ii) a concatenation of  
967 spike trains simultaneously recorded under several electrodes (for the computation of the  
968 population,  $MI_{\text{Population}}$ ). In both cases, the following computation steps were the same.  
969 Neuronal responses were represented using different time scales ranging from the duration of  
970 the whole response (firing rate) to a 1-ms precision (precise temporal patterns), which allows  
971 analyzing how much the spike timing contributes to the information. As this method is  
972 exhaustively described in Schnupp et al. (2006)<sup>72</sup> and in Gaucher et al. (2013a)<sup>69</sup>, we only  
973 present below the main principles.

974 The method relies on a pattern-recognition algorithm that is designed to "guess which  
975 stimulus evoked a particular response pattern"<sup>72</sup> by going through the following steps: From  
976 all the responses of a cortical site to the different stimuli, a single response (test pattern) is

977 extracted and represented as a PSTH with a given bin size (different sizes were considered as  
978 discussed further below). Then, a mean response pattern is computed from the remaining  
979 responses (training set) for each stimulus class. The test pattern is then assigned to the  
980 stimulus class of the closest mean response pattern. This operation is repeated for all the  
981 responses, generating a confusion matrix where each response is assigned to a given stimulus  
982 class. From this confusion matrix, the Mutual Information (MI) is given by Shannon's  
983 formula:  
984

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

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

987 In our case, we used responses to the 4 whistles and selected the first 264 ms of these  
988 responses to work on spike trains of exactly the same duration (the shortest whistle being  
989 280msec long). In a scenario where the responses do not carry information, the assignments  
990 of each response to a mean response pattern is equivalent to chance level (here 0.25 because  
991 we used 4 different stimuli and each stimulus was presented the same number of times) and  
992 the MI would be close to zero. In the opposite case, when responses are very different  
993 between stimulus classes and very similar within a stimulus class, the confusion matrix would  
994 be diagonal and the mutual information would tend to  $\log_2(4) = 2$  bits.

995 This algorithm was applied with different bin sizes ranging from 1 to 280 ms. Supplementary  
996 figure S1B presents the evolution of MI as a function of temporal precision ranging from 1 to  
997 40ms with the responses to the original vocalizations. At the cortical level, we previously  
998 showed that an optimal bin size for obtaining a maximal value of MI was on average 8ms<sup>69,72</sup>.  
999 However, it has never been demonstrated that the same bin size value was optimal at the  
1000 thalamic and brainstem levels. In our experimental conditions, and with our set of acoustic  
1001 stimuli, the 8-ms temporal precision was found to be optimal for all auditory structures  
1002 (dashed line in supplementary Fig. 1b). The MI value obtained for a temporal precision of  
1003 8ms was subsequently used in our analysis.

1004 The MI estimates are subject to non-negligible positive sampling biases. Therefore, as in  
1005 Schnupp et al. (2006)<sup>72</sup>, we estimated the expected size of this bias by calculating MI values  
1006 for “shuffled” data, in which the response patterns were randomly reassigned to stimulus  
1007 classes. The shuffling was repeated 100 times, resulting in 100 MI estimates of the bias

1008 ( $MI_{\text{bias}}$ ). These  $MI_{\text{bias}}$  estimates are then used as estimators for the computation of the  
1009 statistical significance of the MI estimate for the real (unshuffled) datasets: the real estimate is  
1010 considered as significant if its value is statistically different from the distribution of  $MI_{\text{bias}}$   
1011 shuffled estimates. Significant MI estimates were computed for MI calculated from neuronal  
1012 responses under one electrode.

1013 The information carried by a group of recording was estimated by the population MI  
1014 ( $MI_{\text{Population}}$ ), using the same method described above: responses of several simultaneous  
1015 recordings were grouped together and considered as a single pattern. To assess the influence  
1016 of the group size of simultaneous recordings on the information carried by that group  
1017 ( $MI_{\text{Population}}$ ), the number of sites used for computing  $MI_{\text{Population}}$  varied from 2 to the maximal  
1018 possible size (which is equal to 16 minus the non-responsive sites). As the number of possible  
1019 combinations could be extremely large ( $C_n^k$ , where  $k$  is the group size and  $n$  the number of  
1020 responsive sites in a recording session), a threshold was fixed to save computation time: when  
1021 the number of possible combinations exceeded one hundred, 100 combinations were  
1022 randomly chosen, and the mean of all combinations was taken as the  $MI_{\text{Population}}$  for this group  
1023 size.

#### 1024 *Statistics*

1025 To assess the significance of the multiple comparisons (Vocoding process: four levels;  
1026 Masking noise conditions: three levels; Brain structure: five levels), we used an analysis of  
1027 variance (ANOVA) for multiple factors to analyze the whole data set. Follow-up tests were  
1028 corrected for multiple comparisons using Bonferroni corrections and were considered as  
1029 significant if their  $p$  value was below 0.05. All data are presented as mean values  $\pm$  standard  
1030 error (s.e.m.).

1031

1032

## Figure legends

1033  
1034

1035 **Figure 1. Acoustic stimuli and averaged modulation spectra.** **a.** Waveforms (*top*) and  
1036 spectrograms (*bottom*) of the four original whistles used in this study. **b.** *From top to bottom,*  
1037 spectrograms of the four vocoded whistles using 38, 20 and 10 frequency bands. **c.** *From top*  
1038 *to bottom,* spectrograms of the four original whistles embedded in stationary noise at three  
1039 SNRs (+10, 0 and -10 dB) and spectrograms of the stationary noise only (*Noise only*). **d.**  
1040 Vocoding and noise effects on frequency-modulation (FM) spectra. The two plots represent  
1041 the averaged modulation spectra of the four original vocalizations (*in black*), vocoded  
1042 vocalizations (*Voc38, Voc20 and Voc10: red, green and blue respectively, left panel*) and  
1043 vocalizations in stationary noise at three SNRs (+10, 0 and -10 dB : *red, green and blue*  
1044 *respectively, right panel*) **e.** Vocoding and noise effects on amplitude-modulation (AM)  
1045 spectra. The two plots represent the averaged modulation spectra of the four original  
1046 vocalizations (*in black*), vocoded vocalizations (*Voc38, Voc20 and Voc10: red, green and*  
1047 *blue respectively, left panel*) and vocalizations in stationary noise at three SNRs (+10, 0 and -  
1048 *10 dB : red, green and blue respectively, right panel*). Vertical black dashed lines on AM and  
1049 FM spectra correspond to the frequency range (1-20 Hz) selected for the data analysis.

1050

1051 **Figure 2. Subcortical neurons discriminate better the original vocalizations than cortical**  
1052 **neurons.** **a.** *From bottom to top,* neuronal responses were recorded in CN, CNIC, MGv, A1  
1053 and VRB simultaneously under 16 electrodes but only two are represented here, with  
1054 alternated black and red colors. Each dot represents the emission of an action potential and  
1055 each line corresponds to the neuronal discharges to one of four original whistles. The grey  
1056 part of rasters corresponds to evoked activity. The waveforms of the four original whistles are  
1057 displayed under the rasters. **b-e.** The panels show **(b)** the evoked firing rate (spikes/sec), **(c)**  
1058 the temporal reliability quantified by the CorrCoef value (arbitrary units), **(d)** the neuronal  
1059 discrimination assessed by the mutual information (MI) computed at the level of the  
1060 individual recording ( $MI_{\text{Individual}}$ , bits) and **(e)** at the level of neuronal population ( $MI_{\text{Population}}$ ,  
1061 bits) with populations of 9 simultaneous recordings obtained with the four original  
1062 vocalizations in CN (*in black*), CNIC (*in green*), MGv (*in orange*), A1 (*in blue*) and VRB (*in*  
1063 *purple*). In each structure, error bars represent the SD of the mean values and black lines  
1064 represent significant differences between the mean values (unpaired *t* test,  $p < 0.05$ ). Note that  
1065 the evoked firing rate decreases from the CN to VRB but both the temporal reliability  
1066 (CorrCoef) and the discriminative ability (MI) values reach a maximal value in MGv. Note  
1067 also that at the population level, all the subcortical structures discriminate better the original  
1068 vocalizations than cortical areas.

1069

1070 **Figure 3. Vocoding slightly alters neuronal responses at each stage of the auditory**  
1071 **system.** **a.** *From left to right,* raster plots of responses to the four original whistles (*Original*)  
1072 and their vocoded versions generated using either 38, 20 or 10 frequency bands (*Voc38,*  
1073 *Voc20 and Voc10*). *From bottom to top,* neuronal responses were recorded in CN, CNIC,  
1074 MGv, A1 and VRB. **b-e.** The mean values ( $\pm$ SEM) represent **(b)** the evoked firing rate

1075 (spikes/sec), (c) the temporal reliability represented by the CorrCoef value (arbitrary units),  
1076 (d) the neuronal discrimination assessed by the mutual information (MI) computed at the  
1077 level of the individual recordings ( $MI_{\text{Individual}}$ , bits) and (e) at the level of neuronal population  
1078 ( $MI_{\text{Population}}$ , bits) with populations of 9 simultaneous recordings obtained with original  
1079 (*Original*) and vocoded vocalizations (*Voc38*, *Voc20* and *Voc10*) in CN (*in black*), CNIC (*in*  
1080 *green*), MGv (*in orange*), A1 (*in blue*) and VRB (*in purple*) (*one-way ANOVA*, \*P < 0.05). At  
1081 the population level, the discriminative abilities significantly decreased only for 10 frequency  
1082 bands in subcortical structures and did not decrease in cortical areas.

1083

1084 **Figure 4. Noise strongly reduces neuronal responses in all structures but to a lesser**  
1085 **extent in the central nucleus of the inferior colliculus. a.** *From left to right*, raster plots of  
1086 responses of four original whistles (*Original*) and their noisy versions in stationary noise at  
1087 three SNRs: +10, 0 and -10 dB. *From bottom to top*, neuronal responses were recorded in CN,  
1088 CNIC, MGv, A1 and VRB. The grey area corresponds to the evoked activity for each  
1089 vocalization. The green dashed lines show a typical example of CNIC neuronal responses that  
1090 are resistant to the noise addition. **b-e.** The mean values ( $\pm$ SEM) represent (b) the evoked  
1091 firing rate (spikes/sec), (c) the temporal reliability represented by the CorrCoef value  
1092 (arbitrary units), (d) the neuronal discrimination assessed by the mutual information (MI)  
1093 computed at the level of the individual recordings ( $MI_{\text{Individual}}$ , bits) and (e) at the level of  
1094 neuronal population ( $MI_{\text{Population}}$ , bits) with populations of 9 simultaneous recordings obtained  
1095 with original and vocalizations in stationary noise at three SNRs (+10, 0 and -10 dB SPL) in  
1096 CN (*in black*), CNIC (*in green*), MGv (*in orange*), A1 (*in blue*) and VRB (*in purple*) (*one-*  
1097 *way ANOVA*, \*P < 0.05). Note that at the population level, the discriminative abilities  
1098 significantly decreased in all structures when SNR decreased, with the CNIC populations still  
1099 able to discriminate 2 out of 4 stimuli ( $MI_{\text{Population}}$  value >1).

1100

1101 **Figure 5. Reduction of AM cues determine the neuronal discriminative abilities at the**  
1102 **subcortical and cortical levels. a.** Percentage of  $\Delta MI_{\text{Population}}$  as a function of degradation of  
1103 FM components computed for each structure from means  $MI_{\text{Population}}$  or FM values obtained in  
1104 all adverse conditions (noise and vocoded conditions) minus mean values in the original  
1105 condition. Each dot represents neuronal data ( $\Delta MI_{\text{Population}}$ ) in CN (*in black*), CNIC (*in green*),  
1106 MGv (*in orange*), A1 (*in blue*) and VRB (*in purple*). **b.** Percentage of  $\Delta MI_{\text{Population}}$  as a  
1107 function of  $\Delta$ modulation index computed for each structure from mean  $MI_{\text{Population}}$  or mean  
1108 modulation-index values obtained in all adverse conditions and mean values in the original  
1109 condition. Each dot represents neuronal data ( $\Delta MI_{\text{Population}}$ ) in CN (*in black*), CNIC (*in green*),  
1110 MGv (*in orange*), A1 (*in blue*) and VRB (*in purple*). Polynomial curves fitting all acoustic  
1111 conditions have been generated for each auditory structure (*color lines*) except for the  
1112 cochlear nucleus. Note that equivalent AM degradations ( $\Delta$ modulation index) induced  
1113 equivalent decreases in  $MI_{\text{Population}}$  except for the CN. Note also that there is a limit of 25% of  
1114 AM reduction beyond which the  $\Delta MI_{\text{Population}}$  strongly decreases in cortical and subcortical  
1115 structures.

1116



1117 **Supplementary figure 1. Evolution of the CorrCoef and MI mean values as a function of**  
1118 **temporal precision in each structure. a.** CorrCoef values were calculated from responses to  
1119 original vocalisations with a Gaussian window varying in width from 1 to 50 ms in CN (*in*  
1120 *black*), CNIC (*in green*), MGv (*in orange*), A1 (*in blue*) and VRB (*in purple*). In our study, a  
1121 10-ms width Gaussian window (*dashed black line*) was selected for the data analysis in each  
1122 structure. **b.** Mutual information (in *bits*) was calculated from neuronal responses to original  
1123 vocalizations with a bin size varying from 1 to 40 ms in CN (*in black*), CNIC (*in green*), MGv  
1124 (*in orange*), A1 (*in blue*) and VRB (*in purple*). In this study, the value of 8 ms was selected  
1125 for the data analysis because in each structure, the MI value was maximal (*dashed black line*).

1126

1127 **Supplementary figure 2. Large diversity of neuronal discrimination performance in**  
1128 **response to original vocalizations in each auditory structure.** Waterfall plots show the  
1129 mutual information (MI, *bits*) as a function of temporal resolution (1 to 256 ms) for the  
1130 selected recordings in CN, CNIC, MGv, A1 and VRB. In each structure, the units are ranked  
1131 by the MI value obtained with a bin size of 8ms (*dark rainbow colors*). Note that there was a  
1132 larger proportion of neurons with high values of MI (close from the maximal value of 2) in  
1133 MGv, CNIC and CN (*red curves*) compared to a much lower proportion in the cortical areas  
1134 AI and VRB.

1135

1136 **Supplementary figure 3. Population information quickly reaches high values with**  
1137 **simultaneous recordings at the subcortical but not cortical level.** For each auditory  
1138 structure, each thin line represents a particular case of simultaneous recordings with a  
1139 maximum number of electrodes varying from 2 to 12 or 16, and each thick line represents the  
1140 mean value of  $MI_{Population}$  in CN (*in black*), CNIC (*in green*), MGv (*in orange*), A1 (*in blue*)  
1141 and VRB (*in purple*). Note that the mean  $MI_{Population}$  value quickly reaches high values close  
1142 from the maximum value of 2 bits in the subcortical structures (CN, CNIC and MGv)  
1143 compared to the two cortical areas (A1 and VRB).

1144

1145

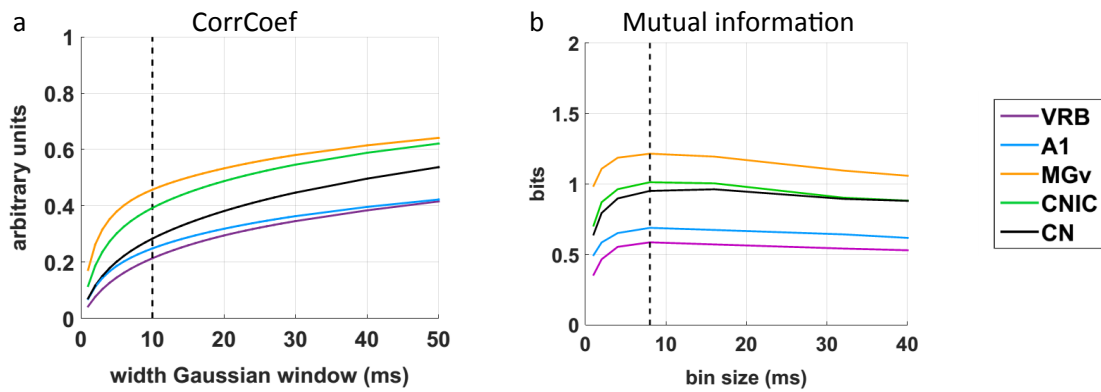
1146 **Supplementary figure 4. Effects of vocoding and noise on the  $MI_{Population}$  growth**  
1147 **functions in each auditory structure. a. Vocoding effects.** The graphics display the average  
1148 growth functions of the  $MI_{Population}$  for each structure in each vocoding condition (indicated by  
1149 a gradient colors). In each structure, the vocoding slightly reduced the  $MI_{Population}$  values. At  
1150 the cortical level, the reduction induced by vocoding was similar at 38 and 20 bands, then a  
1151 stronger reduction was observed at 10 bands. At the thalamic level, there was almost no  
1152 change in the growth function of the  $MI_{Population}$  with 38 and 20 bands vocalizations, but there  
1153 was a large decrease in  $MI_{Population}$  with the 10 bands vocoded stimuli. In the CNIC, there was  
1154 almost no change in the growth function of the  $MI_{Population}$  with 38-bands vocalizations, and a  
1155 progressive reduction of  $MI_{Population}$  with 20 and 10-bands vocalizations. A similar scenario  
1156 was observed at the CN level.

1157 **b. Noise effects.** The graphics display the effects noise on the growth functions of the  
1158  $MI_{\text{Population}}$  for each structure and at each SNR. In general, background noise largely altered  
1159 the growth functions of the  $MI_{\text{Population}}$  in each structure (but to a lesser extent in the CNIC). In  
1160 the cortex, SNR affected the growth functions of the  $MI_{\text{Population}}$  : the lower the SNR, the lower  
1161 the curves of the  $MI_{\text{Population}}$ . In the MGv, stationary noise progressively lowered the curves of  
1162 the  $MI_{\text{Population}}$ . In the CNIC, stationary noise induced SNR-dependent reduction in the  
1163  $MI_{\text{Population}}$  values, the reduction being modest at a +10 and 0 dB SNR but more important at a  
1164 -10 dB SNR. In the CN, stationary noise induced a stronger reduction of the  $MI_{\text{Population}}$  which  
1165 was clearly a function of SNR.

1166

1167 **Supplementary figure 5. A subpopulation of CN neurons maintains good neuronal**  
1168 **discrimination performance at a +10 dB SNR.** Waterfall plots show the mutual information  
1169 ( $MI_{\text{Individual}}$ , bits) as a function of temporal resolution (1 to 256 ms) for the CN recordings at a  
1170 +10 dB SNR. The recordings are ranked by the MI value obtained with a bin size of 8 ms  
1171 (*dark rainbow colors*). Note that at this particular SNR, 20% of the CN recordings (n=39)  
1172 maintained  $MI_{\text{Individual}}$  values above 1 bit, indicating that some CN neurons still send  
1173 information about the vocalization identity at higher brainstem centers such as the CNIC.

1174

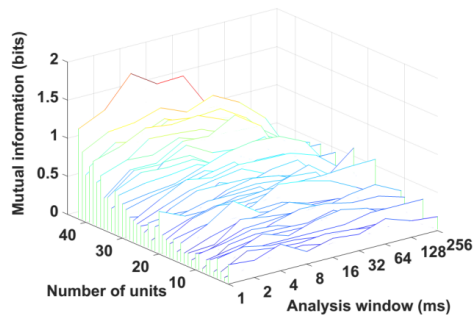


**Supplementary figure 1. Evolution of the CorrCoef and MI mean values as a function of temporal precision in each structure.** **a.** CorrCoef values were calculated from responses to original vocalisations with a Gaussian window varying in width from 1 to 50 ms in CN (*in black*), CNIC (*in green*), MGv (*in orange*), A1 (*in blue*) and VRB (*in purple*). In our study, a 10-ms width Gaussian window (*dashed black line*) was selected for the data analysis in each structure. **b.** Mutual information (in *bits*) was calculated from neuronal responses to original vocalizations with a bin size varying from 1 to 40 ms in CN (*in black*), CNIC (*in green*), MGv (*in orange*), A1 (*in blue*) and VRB (*in purple*). In this study, the value of 8 ms was selected for the data analysis because in each structure, the MI value was maximal (*dashed black line*).

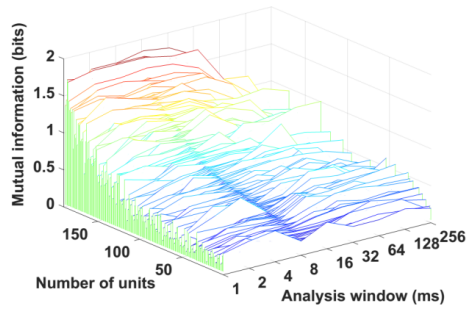
1175

1176

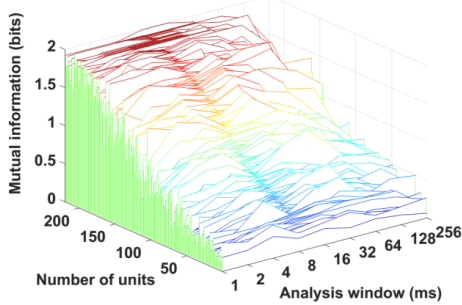
1177



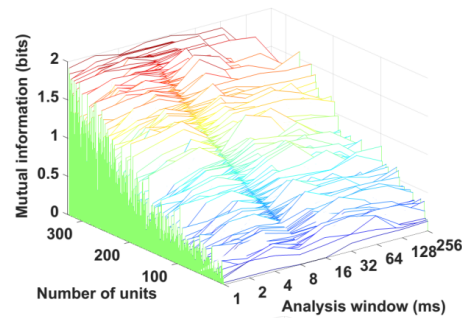
VRB



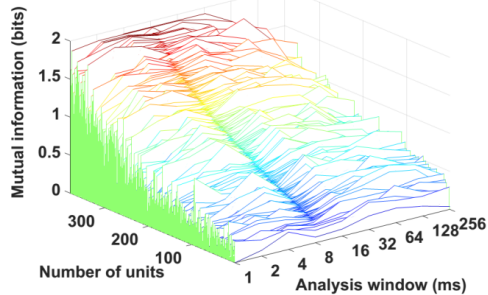
A1



MGv

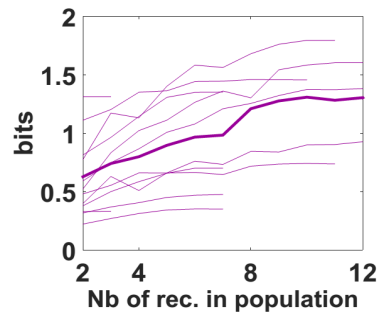


CNIC

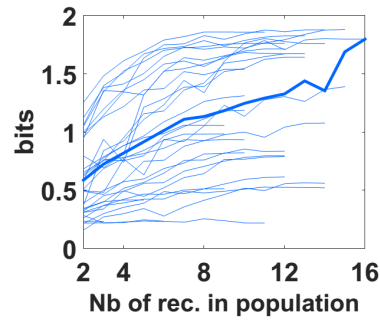


CN

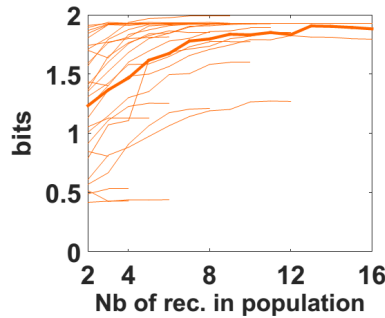
**Supplementary figure 2. Large diversity of neuronal discrimination performance in response to original vocalizations in each auditory structure.** Waterfall plots show the mutual information (MI, *bits*) as a function of temporal resolution (1 to 256 ms) for the selected recordings in CN, CNIC, MGv, A1 and VRB. In each structure, the units are ranked by the MI value obtained with a bin size of 8ms (*dark rainbow colors*). Note that there was a larger proportion of neurons with high values of MI (close from the maximal value of 2) in MGv, CNIC and CN (*red curves*) compared to a much lower proportion in the cortical areas AI and VRB.



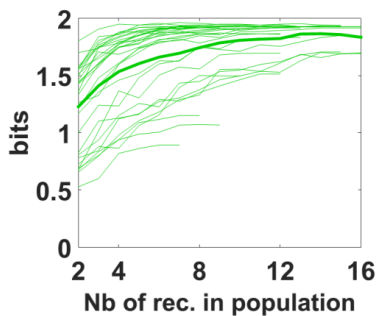
VRB



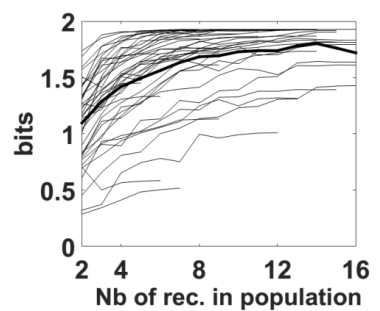
A1



MGv



CNIC



CN

**Supplementary figure 3. Population information quickly reaches high values with simultaneous recordings at the subcortical but not cortical level.** For each auditory structure, each thin line represents a particular case of simultaneous recordings with a maximum number of electrodes varying from 2 to 12 or 16, and each thick line represents the mean value of  $MI_{Population}$  in CN (*in black*), CNIC (*in green*), MGv (*in orange*), A1 (*in blue*) and VRB (*in purple*). Note that the mean  $MI_{Population}$  value quickly reaches high values close from the maximum value of 2 bits in the subcortical structures (CN, CNIC and MGv) compared to the two cortical areas (A1 and VRB).

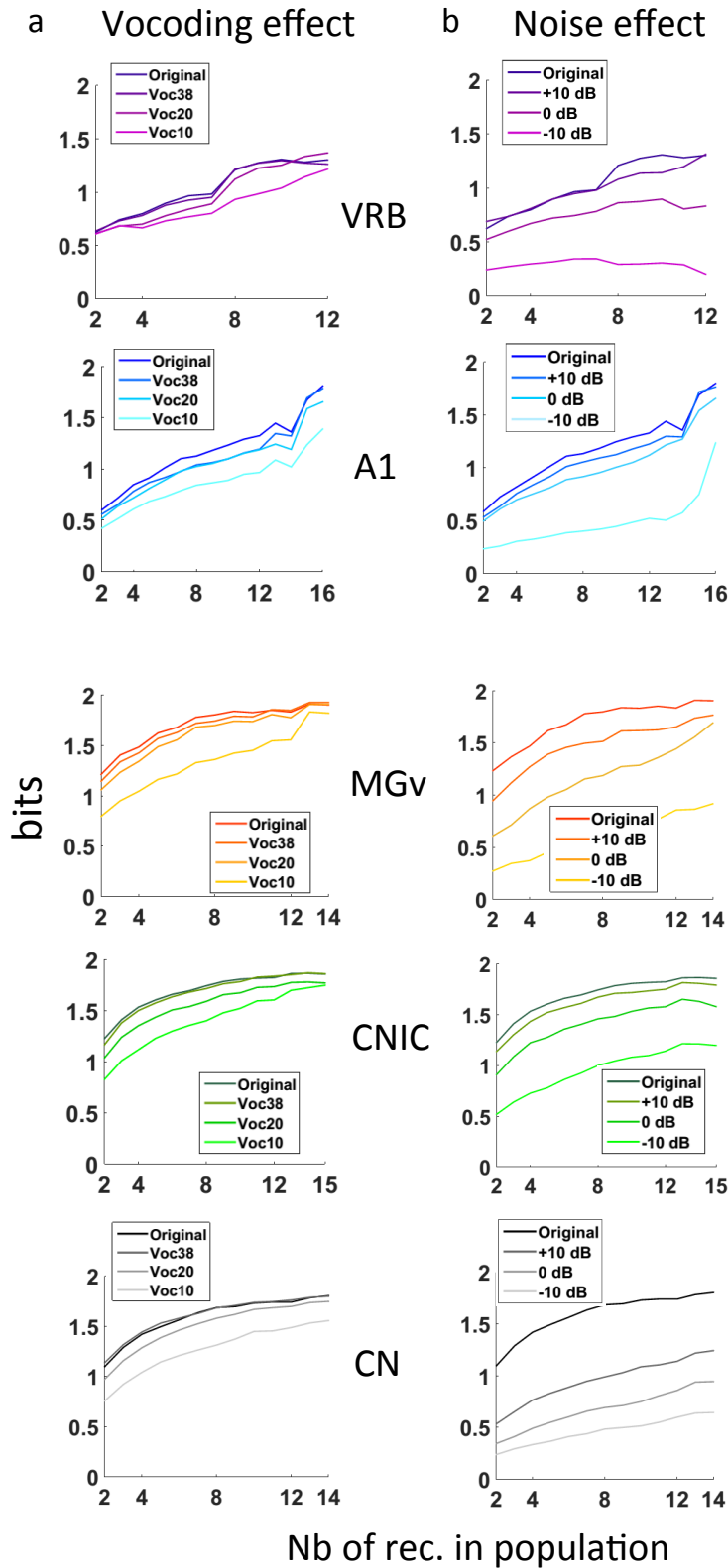
1184

1185

1186

1187

1188

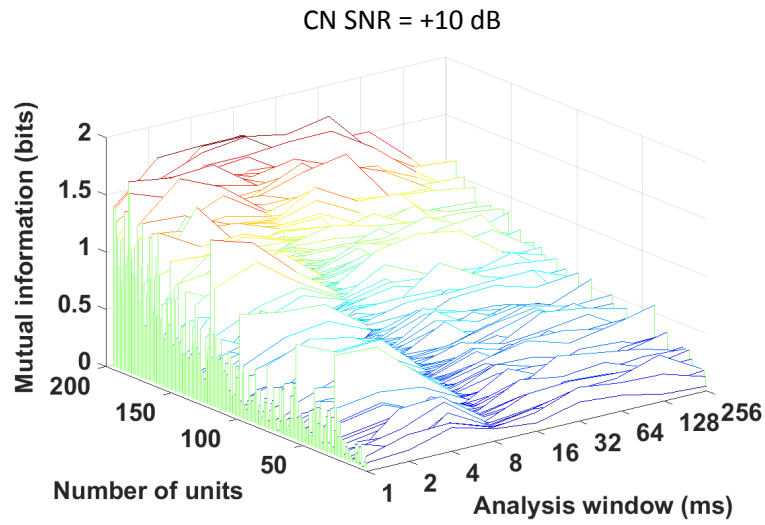


**Supplementary figure 4. Effects of vocoding and noise on the  $MI_{Population}$  growth functions in each auditory structure.**

**a. Vocoding effects.** The graphics display the average growth functions of the  $MI_{Population}$  for each structure in each vocoding condition (indicated by a gradient colors). In each structure, the vocoding slightly reduced the  $MI_{Population}$  values. At the cortical level, the reduction induced by vocoding was similar at 38 and 20 bands, then a stronger reduction was observed at 10 bands. At the thalamic level, there was almost no change in the growth function of the  $MI_{Population}$  with 38 and 20 bands vocalizations, but there was a large decrease in  $MI_{Population}$  with the 10 bands vocoded stimuli. In the CNIC, there was almost no change in the growth function of the  $MI_{Population}$  with 38-bands vocalizations, and a progressive reduction of  $MI_{Population}$  with 20 and 10-bands vocalizations. A similar scenario was observed at the CN level.

**b. Noise effects.** The graphics display the effects noise on the growth functions of the  $MI_{Population}$  for each structure and at each SNR. In general, background noise largely altered the growth functions of the  $MI_{Population}$  in each structure (but to a lesser extent in the CNIC). In the cortex, SNR affected the growth functions of the  $MI_{Population}$  : the lower the SNR, the lower the curves of the  $MI_{Population}$ . In the MGv, stationary noise progressively lowered the curves of the  $MI_{Population}$ . In the CNIC, stationary noise induced SNR-dependent reduction in the  $MI_{Population}$  values, the reduction being modest at a +10 and 0 dB SNR but more important at a -10 dB SNR. In the CN, stationary noise induced a stronger reduction of the  $MI_{Population}$  which was clearly a function of SNR.

1189  
1190  
1191  
1192



**Supplementary figure 5.** A subpopulation of CN neurons maintains good neuronal discrimination performance at a +10 dB SNR. Waterfall plots show the mutual information ( $MI_{\text{Individual}}$ , bits) as a function of temporal resolution (1 to 256 ms) for the CN recordings at a +10 dB SNR. The recordings are ranked by the MI value obtained with a bin size of 8 ms (*dark rainbow colors*). Note that at this particular SNR, 20% of the CN recordings ( $n=39$ ) maintained  $MI_{\text{Individual}}$  values above 1 bit, indicating that some CN neurons still send information about the vocalization identity at higher brainstem centers such as the CNIC.

1193  
 1194  
 1195  
 1196  
 1197

	CN	<i>Lemniscal pathway</i>			<i>Non-lemniscal pathway</i>
		CNIC	MGv	A1	VRB
Number of animals	<b>10</b>	<b>11</b>	<b>10</b>	<b>11</b>	<b>5</b>
Number tested	<b>672</b>	<b>478</b>	<b>448</b>	<b>544</b>	<b>192</b>
STRF only	<b>560</b>	<b>421</b>	<b>285</b>	<b>455</b>	<b>126</b>
Selection criteria : STRF & response to at least one vocalization	<b>499</b>	<b>386</b>	<b>262</b>	<b>354</b>	<b>95</b>
<b><i>STRF quantifications</i></b>					
BF range (kHz) : min-max	<b>0.18 - 18</b>	<b>0.34 - 36</b>	<b>0.33 - 33.01</b>	<b>0.14 - 36</b>	<b>0.67 - 36</b>
Mean bandwidth (octave)	<b>3.91</b>	<b>2.88</b>	<b>4.16</b>	<b>2.07</b>	<b>1.79</b>
Mean response duration (ms)	<b>26.83</b>	<b>35.37</b>	<b>17.31</b>	<b>43.73</b>	<b>44.83</b>
Response strength (spikes/sec)	<b>77.23</b>	<b>82.25</b>	<b>41.61</b>	<b>37.69</b>	<b>19.97</b>

**Supplementary Table 1.** Summary of the number of animals, number of selected recordings and STRFs quantifications in each structure.

1198

1199