



**HAL**  
open science

# Robustness to Noise in the Auditory System: A Distributed and Predictable Property

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

► **To cite this version:**

Samira Souffi, Christian Lorenzi, Chloé Huetz, Jean-Marc Edeline. Robustness to Noise in the Auditory System: A Distributed and Predictable Property. *eNeuro*, 2021, 8 (2), pp.0043-21. 10.1523/ENEURO.0043-21.2021 . hal-03154584

**HAL Id: hal-03154584**

**<https://hal.science/hal-03154584>**

Submitted on 1 Mar 2021

**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.

---

*Research Article: Confirmation | Sensory and Motor Systems*

## **Robustness to Noise in the Auditory System: A Distributed and Predictable Property**

<https://doi.org/10.1523/ENEURO.0043-21.2021>

**Cite as:** eNeuro 2021; 10.1523/ENEURO.0043-21.2021

Received: 29 January 2021

Revised: 17 February 2021

Accepted: 17 February 2021

---

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

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

Copyright © 2021 Souffi et al.

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

1  
2 **Robustness to noise in the auditory system:**  
3 **A distributed and predictable property**  
4

5 Souffi S.<sup>1,2</sup>, Lorenzi C.<sup>3</sup>, Huetz C.<sup>1,2</sup>, Edeline J.-M.\*<sup>1,2</sup>  
6

7 **Running title:** Noise-resistant neurons in the auditory system  
8

9 <sup>1</sup> Paris-Saclay Institute of Neuroscience (Neuro-PSI), Department Integrative and Computational Neu-  
10 roscience, UMR CNRS 9197.

11 <sup>2</sup> Université Paris-Sud, Bâtiment 446, 91405 Orsay cedex, France.

12 <sup>3</sup> Laboratoire des systèmes perceptifs, UMR CNRS 8248, Département d'Etudes Cognitives, Ecole  
13 Normale Supérieure, Université Paris Sciences & Lettres, Paris, France.  
14

15 \* *Corresponding Author:*

16 Jean-Marc Edeline

17 Paris-Saclay Institute of Neuroscience (Neuro-PSI)

18 UMR CNRS 9197 Université Paris-Sud, Bâtiment 446,

19 91405 Orsay cedex, France

20 email: jean-marc.edeline@u-psud.fr  
21

22  
23 Number of pages: 33

24 Number of Figures: 9

25 Number of Extended figures: 4

26 Number of Tables: 2

27 Number of words in the abstract: 222

28 Number of words in the significance statement: 119

29 Number of words in the introduction: 724

30 Number of words in the discussion: 2633  
31

32 **Conflict of interest statement**

33 The authors declare no competing financial interests.

34 **Acknowledgments**  
35

36 CL and JME were supported by grants from the French Agence Nationale de la Recherche (ANR)  
37 (ANR-14-CE30-0019-01). CL was also supported by grants ANR-11-0001-02 PSL and ANR-10-  
38 LABX-0087. SS was supported by the Fondation pour la Recherche Médicale (FRM) grant number  
39 ECO20160736099 and by the Entendre Foundation. We thank Nihaad Paraouty for training us to  
40 cochlear-nucleus surgery and Jennifer Linden for insightful comments on a previous version of this  
41 article. We also wish to thank Céline Dubois, Mélanie Dumont and Aurélie Bonilla, for taking care of  
42 the guinea-pig colony.  
43

44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78

## Abstract

Background noise strongly penalizes auditory perception of speech in humans or vocalizations in animals. Despite this, auditory neurons are still able to detect communications sounds against considerable levels of background noise. We collected neuronal recordings in cochlear nucleus, inferior colliculus, auditory thalamus, primary and secondary auditory cortex in response to vocalizations presented either against a stationary or a chorus noise in anesthetized guinea pigs at three signal-to-noise ratios (-10, 0 and 10 dB). We provide evidence that, at each level of the auditory system, five behaviors in noise exist within a continuum, from neurons with high-fidelity representations of the signal, mostly found in inferior colliculus and thalamus, to neurons with high-fidelity representations of the noise, mostly found in cochlear nucleus for the stationary noise and in similar proportions in each structure for the chorus noise. The two cortical areas displayed fewer robust responses than the inferior colliculus and thalamus. Furthermore, between 21 and 72% of the neurons (depending on the structure) switch categories from one background noise to another, even if the initial assignment of these neurons to a category was confirmed by a severe bootstrap procedure. Importantly, supervised learning pointed out that assigning a recording to one of the five categories can be predicted up to a maximum of 70% based on both the response to signal alone and noise alone.

## Significance statement

In daily situations, humans and animals are faced with various background noises in which they have to detect behaviorally salient signals. Noise resistance is often viewed as an emergent property of cortical networks, but only a few studies have characterized the relative contribution of cortical and subcortical neurons. Our results demonstrate that the neuronal resistance to noise is distributed along the auditory system with a more important fraction of robust neurons in subcortical structures compared to auditory cortex, and is relatively well predictable based on the responses to the signal alone and the noise alone. Our results also suggest that noise-invariant representations of communication sounds coexist with accurate noise representations, which are detected as early as the cochlear nucleus.

**Keywords:** auditory system; natural vocalizations; noise resistance; neuronal classification; noise-type sensitivity



## Introduction

79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114

In natural conditions, speech (in humans) and communication sounds (in animals) usually co-occur with many other competing acoustic signals. During evolution, the auditory system has developed strategies to extract these behaviorally important signals mixed up with substantial amounts of noise. Over the last decade, many studies performed on different species have reported that the responses of auditory cortex neurons are quite resistant to various types of noises, even at low SNR (Narayan et al., 2007; Schneider and Woolley, 2013; Rabinowitz et al., 2013, Mesgarani et al., 2014; Ni et al., 2017; Beetz et al., 2018). Several hypotheses have been formulated to account for the high performance of auditory cortex neurons. For example, it was proposed that noise tolerance is correlated with adaptation to the stimulus statistics, potentially more pronounced at the cortical than at the subcortical level (Rabinowitz et al., 2013). A dynamic model of synaptic depression was also suggested as a potential mechanism for robust speech representation in the auditory cortex (Mesgarani et al., 2014). Alternatively, a simple feedforward inhibition circuit was viewed as a mechanism to explain background-invariant responses detected in the secondary auditory cortex (Schneider and Woolley, 2013). A recent study (Ni et al., 2017) reported that auditory cortex neurons can be assigned to categories depending upon their robustness to noise. By testing the responses to conspecific vocalizations at different SNRs, this study described four types of response categories (robust, balanced, insensitive and brittle) in the marmoset primary auditory cortex, and pointed out that depending upon the background noise, two-thirds of A1 neurons exhibit different response classes (Ni et al., 2017). The present study aimed at determining whether the subcortical auditory structures display similar proportions of these four categories and whether the noise-type sensitivity is already present at the subcortical level. We used the same methodology as in Ni and colleagues (2017) to assign each recording to a given response class: the Extraction Index (EI, initially defined by Schneider and Woolley, 2013) was computed at three SNRs (+10, 0 and -10 dB) and an unsupervised clustering approach (the K-means algorithm) revealed groups of EI profiles in a given background noise. We performed this clustering approach in the cochlear nucleus, inferior colliculus, auditory thalamus, primary and secondary auditory cortex using two types of masking noise, a stationary or a chorus noise composed of a mixture of conspecific vocalizations. We renamed two neuronal behaviors in noise for recognizing equivalent roles to stimulus-like neuronal responses (i.e., the signal- and the masker-like responses) since in ethological conditions, both could play an important functional role. The categories range from signal-like responses (equivalent to the ‘robust’ neurons of Ni et al., 2017) showing a high-fidelity representation of the signal, to masker-like responses showing a high-fidelity representation of the noise (equivalent to the ‘brittle’ neurons of Ni et al., 2017), with two intermediary categories, one showing no preference either for the signal or for the noise named insensitive, and the other characterized by the highest sensitivity to the SNR named balanced. To minimize the intra-

115 category distances, we also added a new category, called signal-dominated, which corresponds to an  
116 attenuated version of the signal-like responses.

117 Here, we present evidence that the categories initially described by Ni and colleagues (2017) in the  
118 primary auditory cortex do exist at each stage of the auditory system, from cochlear nucleus to sec-  
119 ondary auditory cortex. From a continuum of EI values, we imposed a clustering and revealed that  
120 each category was represented at each relay of the auditory system in different proportions, depending  
121 on the auditory structure and the type of the masking noise. Signal-like and signal-dominated respons-  
122 es were in higher proportions in inferior colliculus and thalamus in both noises. Masker-like responses  
123 were found mostly in the cochlear nucleus in stationary noise but in similar proportions in each struc-  
124 ture in chorus noise. Interestingly, the proportion of balanced responses decreased as one ascends in  
125 the auditory system suggesting a decreased sensitivity to SNR at the cortical level. The noise-type  
126 sensitivity - that is the ability to switch category from a given background noise to another - exists at  
127 each level of the auditory system. Using a supervised learning approach with descriptors extracted  
128 from the responses to the original vocalizations alone (signal) and to the maskers alone, we provide  
129 evidence that the assignment to a given category is relatively well predicted in both types of noise.

130

131

## Materials and Methods

132 Most of the Methods are similar to those described in a previous study (Souffi et al., 2020).

133

### 134 **Subjects**

135 These experiments were performed under the national license A-91-557 (project 2014-25, authoriza-  
136 tion 05202.02) and using the procedures N° 32-2011 and 34-2012 validated by the Ethic committee  
137 N°59 (CEEA Paris Centre et Sud). All surgical procedures were performed in accordance with the  
138 guidelines established by the European Communities Council Directive (2010/63/EU Council Di-  
139 rective Decree).

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

146 Two days before the experiment, the animal's pure-tone audiogram was determined by testing audito-  
147 ry brainstem responses (ABR) under isoflurane anesthesia (2.5 %) as described in Gourévitch and  
148 colleagues (2009). A software (RTLab, Echodia, Clermont-Ferrand, France) allowed averaging 500  
149 responses during the presentation of nine pure-tone frequencies (between 0.5 and 32 kHz) delivered  
150 by a speaker (Knowles Electronics) placed in the animal right ear canal. The auditory threshold of  
151 each ABR was the lowest intensity where a small ABR wave can still be detected (usually wave III).  
152 For each frequency, the threshold was determined by gradually decreasing the sound intensity (from  
153 80 dB SPL down to -10 dB SPL). All animals used in this study had normal pure-tone audiograms  
154 (Gourévitch et al., 2009; Gourévitch and Edeline, 2011).

155

### 156 **Surgical procedures**

157 All animals were anesthetized by an initial injection of urethane (1.2 g/kg, i.p.) supplemented by addi-  
158 tional doses of urethane (0.5 g/kg, i.p.) when reflex movements were observed after pinching the hind  
159 paw (usually 2-4 times during the recording session). A single dose of atropine sulphate (0.06mg/kg,  
160 s.c.) was given to reduce bronchial secretions and a small dose of buprenorphine was administrated  
161 (0.05mg/kg, s.c.) as urethane has no antalgic properties. After placing the animal in a stereotaxic  
162 frame, a craniotomy was performed and a local anesthetic (Xylocain 2%) was liberally injected in the  
163 wound.

164 For auditory cortex recordings (areas A1 and VRB), a craniotomy was performed above the left tem-  
165 poral cortex. The dura above the auditory cortex was removed under binocular control and the cere-  
166 brospinal fluid was drained through the cisterna to prevent the occurrence of oedema. For the record-  
167 ings in MGB, a craniotomy was performed above the most posterior part of the MGB (8mm posterior  
168 to Bregma) to reach the left auditory thalamus at a location where the MGB is mainly composed of its  
169 ventral, tonotopic, part (Redies et al., 1989, Edeline et al., 1999; Anderson et al., 2007; Wallace et al.,  
170 2007). For IC recordings, a craniotomy was performed above the left IC and portions of the cortex  
171 were aspirated to expose the surface of the left IC (Malmierca et al., 1995, 1996; Rees et al., 1997).  
172 For CN recordings, after opening the skull above the left cerebellum, portions of the cerebellum were  
173 aspirated to expose the surface of the left CN (Paraouty et al., 2018).  
174 After all surgeries, a pedestal in dental acrylic cement was built to allow an atraumatic fixation of the  
175 animal's head during the recording session. The stereotaxic frame supporting the animal was placed in  
176 a sound-attenuating chamber (IAC, model AC1). At the end of the recording session, a lethal dose of  
177 Exagon (pentobarbital >200 mg/kg, i.p.) was administered to the animal.

#### 178 **Recording procedures**

179 Data from multi-unit recordings were collected in 5 auditory structures, the non-primary cortical area  
180 VRB, the primary cortical area A1, the medial geniculate body (MGB), the inferior colliculus (IC) and  
181 the cochlear nucleus (CN). In a given animal, neuronal recordings were only collected in one auditory  
182 structure.

183 Cortical extracellular recordings were obtained from arrays of 16 tungsten electrodes (TDT, Tuck-  
184 erDavis Technologies;  $\varnothing$ : 33  $\mu\text{m}$ , <1 M $\Omega$ ) composed of two rows of 8 electrodes separated by 1000  
185  $\mu\text{m}$  (350  $\mu\text{m}$  between electrodes of the same row). A silver wire, used as ground, was inserted be-  
186 tween the temporal bone and the dura matter on the contralateral side. The location of the primary  
187 auditory cortex was estimated based on the pattern of vasculature observed in previous studies (Wal-  
188 lace et al., 2000; Gaucher et al., 2013; Gaucher and Edeline, 2015). The non-primary cortical area  
189 VRB was located ventral to A1 and distinguished because of its long latencies to pure tones (Rutkow-  
190 ski et al., 2002; Grimsley et al., 2012). For each experiment, the position of the electrode array was set  
191 in such a way that the two rows of eight electrodes sample neurons responding from low to high fre-  
192 quency when progressing in the rostro-caudal direction [see examples in Figure 1 of Gaucher et al.,  
193 (2012) and in Figure 6A of Occelli et al., (2016)].

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

198 All thalamic recordings were from the ventral part of MGB (see above surgical procedures) and all  
199 displayed latencies < 9ms. At the collicular level, we distinguished the lemniscal and non-lemniscal  
200 divisions of IC based on depth and on the latencies of pure tone responses. We excluded the most su-  
201 perflicial recordings (until a depth of 1500 $\mu$ m) and those exhibiting latency  $\geq$  20ms in an attempt to  
202 select recordings from the central nucleus of IC (CNIC). At the level of the cochlear nucleus, the re-  
203 cordings were collected from both the dorsal (DCN) and ventral (VCN) divisions, but based on the  
204 recording depth, we estimate that the DCN recordings were more numerous.

205 The raw signal was amplified 10,000 times (TDT Medusa). It was then processed by an RX5 multi-  
206 channel data acquisition system (TDT). The signal collected from each electrode was filtered (610-  
207 10000 Hz) to extract multi-unit activity (MUA). The trigger level was set for each electrode to select  
208 the largest action potentials from the signal. On-line and off-line examination of the waveforms sug-  
209 gests that the MUA collected here was made of action potentials generated by a few neurons at the  
210 vicinity of the electrode. However, as we did not used tetrodes, the result of several clustering algo-  
211 rithms (Pouzat et al., 2004; Quiroga et al., 2004; Franke et al., 2015) based on spike waveform anal-  
212 yses were not reliable enough to isolate single units with good confidence. Although these are not  
213 direct proofs, the fact that the electrodes were of similar impedance (0.5-1M $\Omega$ ) and that the spike  
214 amplitudes had similar values (100-300 $\mu$ V) for the cortical and the subcortical recordings, were two  
215 indications suggesting that the cluster recordings obtained in each structure included a similar number  
216 of neurons. Even if a similar number of neurons were recorded in the different structures, we cannot  
217 discard the possibility that the homogeneity of the multi-unit recordings differs between structures. By  
218 collecting several hundreds of recordings in each structure, these potential differences in homogeneity  
219 should be attenuated in the present study.

#### 220 **Acoustic stimuli**

221 Acoustic stimuli were generated using MatLab, transferred to a RP2.1-based sound delivery system  
222 (TDT) and sent to a Fostex speaker (FE87E). The speaker was placed at 2 cm from the guinea pig's  
223 right ear (or left ear for CN recordings), a distance at which the speaker produced a flat spectrum ( $\pm$  3  
224 dB) between 140 Hz and 36 kHz. Calibration of the speaker was made using noise and pure tones re-  
225 corded by a Bruel and Kjaer microphone 4133 coupled to a preamplifier B & K 2169 and a digital re-  
226 corder Marantz PMD671. All the stimuli intensities were calculated as RMS.

227 The Time-Frequency Response Profiles (TFRP) were determined using 129 pure-tones frequencies  
228 covering eight octaves (0.14-36 kHz) and presented at 75 dB SPL. The tones had a gamma envelop  
229 given by  $\gamma(t) = \left(\frac{t}{4}\right)^2 e^{-\frac{t}{4}}$ , where t is time in ms. At a given level, each frequency was repeated eight  
230 times at a rate of 2.35 Hz in pseudorandom order. The duration of these tones over half-peak ampli-  
231 tude was 15 ms and the total duration of the tone was 50 ms, so there was no overlap between tones.

232 A set of four conspecific vocalizations was used to assess the neuronal responses to communication  
233 sounds. These vocalizations were recorded from animals of our colony. Pairs of animals were placed  
234 in the acoustic chamber and their vocalizations were recorded by a Bruel & Kjaer microphone 4133  
235 coupled to a preamplifier B&K 2169 and a digital recorder Marantz PMD671. A large set of whistle  
236 calls was loaded in the Audition software (Adobe Audition 3) and four representative examples of  
237 whistle were selected. As shown in figure 1A (lower panels), despite the fact the maximal energy of  
238 the four selected whistle was in the same frequency range (typically between 4 and 26 kHz), these  
239 calls displayed slight differences in their spectrograms. In addition, their temporal (amplitude) enve-  
240 lopes clearly differed as shown by their waveforms (Figure 1A, upper panels).

241 The four whistles were also presented in two frozen noises ranging from 10 to 24,000 Hz. To generate  
242 these noises, recordings were performed in the colony room where a large group of guinea pigs were  
243 housed (30-40; 2-4 animals/cage). Several 4-seconds of audio recordings were added up to generate  
244 the "chorus noise", which power spectrum was computed using the Fourier transform. This spectrum  
245 was then used to shape the spectrum of a white Gaussian noise. The resulting vocalization-shaped  
246 stationary noise therefore matched the "chorus-noise" audio spectrum, which explains why some fre-  
247 quency bands were over-represented in the vocalization-shaped stationary noise. Figures 1B-1C dis-  
248 play the spectrograms of the four whistles in the vocalization-shaped stationary noise (1B) and in the  
249 chorus noise (1C) with a SNR of +10 dB, 0 dB, -10 dB. The last spectrograms of these two figures  
250 represent the noises only.

### 251 **Experimental protocol**

252 As inserting an array of 16 electrodes in a brain structure almost systematically induces a deformation  
253 of this structure, a 30-minutes recovering time lapse was allowed for the structure to return to its ini-  
254 tial shape, then the array was slowly lowered. Tests based on measures of time-frequency response  
255 profiles (TFRPs) were used to assess the quality of our recordings and to adjust electrodes' depth. For  
256 auditory cortex recordings (A1 and VRB), the recording depth was 500-1000  $\mu\text{m}$ , which corresponds  
257 to layer III and the upper part of layer IV according to Wallace and Palmer (2008). For thalamic re-  
258 cordings, the NeuroNexus probe was lowered about 7mm below pia before the first responses to pure  
259 tones were detected.

260 When a clear frequency tuning was obtained for at least 10 of the 16 electrodes, the stability of the  
261 tuning was assessed: we required that the recorded neurons displayed at least three successive similar  
262 TFRPs (each lasting 6 minutes) before starting the protocol. When the stability was satisfactory, the  
263 protocol was started by presenting the acoustic stimuli in the following order: We first presented the  
264 four whistles at 75 dB SPL in their original versions (in quiet), then the chorus and the vocalization-  
265 shaped stationary noises were presented at 75 dB SPL followed by the masked vocalizations presented



266 against the chorus then against the vocalization-shaped stationary noise at 65, 75 and 85 dB SPL.  
 267 Thus, the level of the original vocalizations was kept constant (75 dB SPL), and the noise level was  
 268 increased (65, 75 and 85 dB SPL). In all cases, each vocalization was repeated 20 times. Presentation  
 269 of this entire stimulus set lasted 45 minutes. The protocol was re-started either after moving the elec-  
 270 trode arrays on the cortical map or after lowering the electrode at least by 300  $\mu\text{m}$  for subcortical  
 271 structures.

272

### 273 **Data analysis**

274 All the analyses were performed on MATLAB 2019 (MathWorks).

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

276 The TFRP were obtained by constructing post-stimulus time histograms for each frequency with 1 ms  
 277 time bins. The firing rate evoked by each frequency was quantified by summing all the action poten-  
 278 tials from the tone onset up to 100 ms after this onset. Thus, TFRP were matrices of 100 bins in ab-  
 279 scissa (time) multiplied by 129 bins in ordinate (frequency). All TFRPs were smoothed with a uni-  
 280 form 5x5 bin window.

281 For each TFRP, the Best Frequency (BF) was defined as the frequency at which the highest firing rate  
 282 was recorded. Peaks of significant response were automatically identified using the following proce-  
 283 dure: A positive peak in the TFRP was defined as a contour of firing rate above the average level of  
 284 the baseline activity plus six times the standard deviation of the baseline activity. Recordings without  
 285 significant peak of responses or with inhibitory responses were excluded from the data analyses.

286

### 287 **Quantification of responses evoked by original vocalizations and noises alone**

288 The responses to vocalizations were quantified using two parameters:

289 (i) The firing rate of the evoked response, which corresponds to the total number of action potentials  
 290 occurring during the presentation of the stimulus;

291 (ii) the trial-to-trial temporal reliability coefficient (CorrCoef) which quantifies the trial-to-trial relia-  
 292 bility of the response over the 20 repetitions of the same stimulus. This index was computed for each  
 293 vocalization: it corresponds to the normalized covariance between each pair of spike trains recorded at  
 294 presentation of this vocalization and was calculated as follows:

$$295 \text{CorrCoef} = \frac{1}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{\sigma_{x_i x_j}}{\sigma_{x_i} \sigma_{x_j}}$$

296 where N is the number of trials and  $\sigma_{x_i x_j}$  is the normalized covariance at zero lag between spike trains  
 297  $x_i$  and  $x_j$  where i and j are the trial numbers. Spike trains  $x_i$  and  $x_j$  were previously convolved with a

298 10-ms width Gaussian window. Based upon computer simulations, we have previously shown that  
 299 this CorrCoef index is not a function of the neurons' firing rate (Gaucher et al., 2013a).

300 We have computed the CorrCoef index with a Gaussian window ranging from 1 to 50 ms to determine  
 301 if the selection of a particular value for the Gaussian window influences the difference in CorrCoef  
 302 mean values obtained in the different auditory structures. Based upon the responses to the original  
 303 vocalizations, we observed that the relative ranking between auditory structures remained unchanged  
 304 whatever the size of the Gaussian window was. Therefore, we kept the value of 10 ms for the Gaussi-  
 305 an window as it was used in several previous studies (Huetz et al., 2009; Gaucher et al., 2013a; Gau-  
 306 cher and Edeline, 2015; Aushana et al., 2018; Souffi et al., 2020).

307

### 308 **Quantification of the Extraction Index**

309 To evaluate the influence of noise upon neural representation of vocalizations, we quantified the  
 310 amount of vocalization encoded by neurons at a particular SNR level by calculating the Extraction  
 311 Index (EI) adapted from a similar study in songbirds (Schneider and Woolley, 2013). This metric is  
 312 based on the repetition-averaged peristimulus time histogram (PSTH) of neural response with a time  
 313 bin of 4 ms. Different window bins of 1, 2, 4, 8, 16, and 32 ms were also evaluated, which yielded  
 314 qualitatively similar results. In this manuscript, we only report results based on 4 ms time bins. Only  
 315 the PSTH during the evoked activity is taken into account in this analysis.

316 EI was computed as follows:

$$EI = \frac{D_{n-snr} - D_{v-snr}}{D_{n-snr} + D_{v-snr}}$$

$$D_{n-snr} = 1 - \frac{\vec{P}_n \cdot \vec{P}_{snr}}{\|\vec{P}_n\| \|\vec{P}_{snr}\|}, \quad D_{v-snr} = 1 - \frac{\vec{P}_v \cdot \vec{P}_{snr}}{\|\vec{P}_v\| \|\vec{P}_{snr}\|}$$

320 where  $D_{n-snr}$  is the distance between  $PSTH_n$  of noise and  $PSTH_{snr}$  of vocalization at a particular SNR,  
 321 whereas  $D_{v-snr}$  is the distance between  $PSTH_v$  of pure vocalization and  $PSTH_{snr}$  of vocalization at a  
 322 particular SNR. EI is bounded between -1 and 1: a positive value indicates that the neural response to  
 323 noisy vocalization is more vocalization-like, and a negative value implies that the neural response is  
 324 more noise-like. The EI profile for each recording was determined by computing EI at every SNR  
 325 level. The normalized inner product was used to compute distance between  $n$  or  $v$  and  $snr$ , as shown in  
 326 equation above.

327 To probe the response patterns of each neuron, we further implemented an exploratory analysis based  
 328 on the calculated EI profiles as in Ni and colleagues (2017). By applying k-means clustering on the  
 329 blended EI profiles from both noise conditions separately, we obtained, from a continuum of EI val-  
 330 ues, subgroups of EI profiles, which divided the neuronal population into clusters according to the



331 similarity of their EI profiles. Similarity was quantified by Euclidean distance. The number of clusters  
332 was determined by the mean-squared error (MSE) of clustering, as in equation below,

$$MSE = \frac{1}{N} \sum_{i=1}^N (\overline{EIP}_{cluster-i} - EIP_i)^2$$

333  
334 where N is the number of neurons,  $EIP_i$  is the EI profile of a neuron, and  $\overline{EIP}_{cluster-i}$  is the mean EI profile  
335 of the cluster into which this neuron is categorized.

336 To determine a significance level for the Extraction Index of each neuron, we generated 100 false ran-  
337 dom spike trains which follow a Poisson law based on the firing rate values obtained for each stimulus  
338 (original and noisy vocalizations). For a given SNR and recording, we computed based on these false  
339 spike trains, 100  $EI_{Surrogate}$  values and fixed a significance level corresponding to the mean plus two  
340 standard deviations. Using this criterion, we selected only the recordings with at least one of the six EI  
341 values significantly higher than the  $EI_{Surrogate}$ .

342

#### 343 *Bootstrap procedure*

344 To estimate the variability of the EI index generated in assigning each recording to a particular cate-  
345 gory in particular noise, we used a bootstrap strategy for all the recordings, separately for the station-  
346 ary and the chorus noise. Even in anesthetized animals, auditory cortex responses can show some var-  
347 iability. We suspected that in a given type of noise, a recording could change category because of the  
348 response variability and/or because the border between two clusters was very close, independently to  
349 the change in noise type.

350 From the 20 trials obtained for each stimulus during the experiment, we resampled randomly 20 trials  
351 (allowing repetitions) keeping the total number of trials the same as in the experimental data. For each  
352 resampled group of 20 trials, we recalculated both the PSTHs and the Extraction Index at each SNR  
353 then the K-means algorithm was used to define five clusters as in the experimental data. For each re-  
354 cording, this procedure was performed 100 times. Then, we reallocated each resampled data in the  
355 closest cluster compared to the original centroids of the experimental data to measure the percentage  
356 of changed categories relative to the original clustering.

357

#### 358 **Classification using Linear Discriminant Analysis**

359

360 In order to investigate whether the assignment of a given recording to a particular category can be  
361 predicted from the response characteristics obtained with pure tones and/or with the original vocaliza-  
362 tions in quiet, and/or the noise responses alone, an automatic classification algorithm was applied.  
363 Linear Discriminant Analysis classifiers (LDA) were chosen among several linear supervised learning  
364 algorithms because of their slight higher classification performances when trained with all predictors

365 (Statistics and Machine Learning Toolbox™, MATLAB 2019). LDA is a statistical classifier that  
366 achieves a linear decision boundary based on the class scatter matrices. All classifiers used in this  
367 study were given identical parameters (same cost matrix, same cross-validation scheme), but were  
368 given different sets of predictors extracted from the data (as shown on Figures 6-7). A cost matrix was  
369 constructed in order to penalize the wrong assignment of all categories into the “insensitive” category  
370 that contained more recordings than the other categories (all costs were set to 1 except for the “insen-  
371 sitive” category where they were set to 2). Cross-validation was performed using a 5-fold validation  
372 scheme.

373 Fifteen LDA classifiers were built, trained and tested on the recordings considered as reliable with the  
374 bootstrap procedure (with a confidence interval  $\square$  95%) in both noises: 342 recordings were thus se-  
375 lected for the stationary and chorus noise (see Table 2, first line). For each recording and each type of  
376 noise (stationary or chorus), 12 predictors were available, extracted either from responses to pure  
377 tones, vocalizations, maskers or a combination of both signal and maskers. Each classifier has used a  
378 subset of predictors as shown in figure 8A.

379

#### 380 **Descriptors used for the classifier**

381

382 In total, 12 neuronal descriptors were extracted for each type of noise grouped in four types of de-  
383 scriptors: the TFRP, signal, masker and signal-to-masker ratio descriptors (Figures 6-7). Three de-  
384 scriptors were extracted from the responses to pure tones (TFRP): the best frequency (BF) firing rate  
385 (in spikes/sec), the bandwidth of tuning (in octave) and the response duration (in ms). From the re-  
386 sponses to the signal alone (original vocalizations) and the maskers alone, two descriptors were ex-  
387 tracted: the firing rate (in spikes/sec) and the temporal reliability (or CorrCoef).

388 Three other masker descriptors were computed to have an estimation of the firing rate short-term ad-  
389 aptation to the masker: first, we computed the ratio between the firing rate taken at the time the signal  
390 should have occurred and the initial firing rate during the first 200 ms of the masker (FRm300 and  
391 FRm200, Figures 6-7H). Second, we extracted the number of action potentials emitted during the first  
392 (Initial) and last (Final) 50 ms of the masker alone over a 564 ms-period (Figures 6-7I).

393 For the signal-to-masker ratios, the response to the maskers was extracted from the masker alone trials  
394 either from the initial firing rate (first 200 ms of the masker, FRm200, (1) in the Figures 6-7J) or from  
395 the firing rate taken at the time the signal should have occurred (i.e., for a mean duration of 300 ms,  
396 FRm300, (2) in the Figures 6-7K).

397

398

399

400 **Global quantification of category changes with mutual information**

401

402 The mutual information allowed quantifying how many recordings change category from stationary  
403 noise to chorus noise based upon either all recordings (Figure 5-1A) or only the reliable recordings  
404 (Figure 5-1B), independently of structure. For that, we built a matrix with five rows related to the five  
405 categories in stationary noise and five columns related to the becoming of each recording in chorus  
406 noise. From this matrix, the Mutual Information (MI) is given by Shannon's formula:

$$407 \quad MI = \sum_{x,y} p(x,y) \times \log_2 \left( \frac{p(x,y)}{p(x) \times p(y)} \right)$$

408 where x and y are the rows and columns of the confusion matrix.

409 In a scenario where the categorization based on the responses in stationary noise do not carry infor-  
410 mation on the categorization based on the responses in chorus noise, assignment of each recording to a  
411 category is equivalent to chance level (here 0.20 because there were 5 different categories) and the MI  
412 would be close to zero. In the opposite case, when responses in stationary noise always fall in the  
413 same category when recorded with chorus noise, the confusion matrix would be diagonal and the mu-  
414 tual information would tend to  $\log_2(5) = 2.3$  bits.

415

416 **Statistics**

417

418 To assess the significance of the multiple comparisons (masking noise conditions: three levels; audito-  
419 ry structure: five levels), we used an analysis of variance (ANOVA) for multiple factors to analyze the  
420 whole data set. Post-hoc pairwise tests were performed between the noisy conditions (paired t-tests)  
421 and between categories (Kruskal-Wallis tests). They were corrected for multiple comparisons using  
422 Bonferroni corrections and were considered as significant if their p-value was below 0.05.

423

424

## Results

425 From a database of 2334 multi-unit recordings collected in the five investigated auditory structures,  
426 several criteria were used to include each neuronal recording in our analyses (see Table 1). A record-  
427 ing had to show significant responses to pure tones (see Methods section) and a significant evoked  
428 firing rate relative to spontaneous firing rate (200 ms before each original vocalization) in response to  
429 at least one of the four original vocalizations (Figure 1A illustrates their waveforms and spectro-  
430 grams). These four vocalizations were presented in quiet and embedded either in a vocalization-  
431 shaped stationary noise (Figure 1B) or in a chorus noise (Figure 1C) using three SNRs (+10, 0 and -10  
432 dB). We selected recordings showing responses at the three SNRs both in stationary and chorus noise  
433 in order to derive systematically six Extraction Index (EI) values for each neuronal recording. The EI  
434 index quantifies to what extent the evoked response at a given SNR is similar to the response to vocal-  
435 izations in quiet or to noise alone. To determine a significance level of the EI value, we computed an  
436  $EI_{\text{Surrogate}}$  value for each recording (see Methods section) and included only the recordings with at least  
437 one of the six EI values significantly higher than the  $EI_{\text{Surrogate}}$ . Applying these criteria, we selected a  
438 total of 1267 recordings (selection type (b) in Table 1): 389 in the cochlear nucleus (CN), 339 in the  
439 central nucleus of the inferior colliculus (CNIC), 198 in the ventral division of the auditory thalamus  
440 (MGv), 261 in the primary auditory cortex (A1) and 80 in a secondary auditory cortical area (Vento-  
441 Rostral Belt, VRB).

442

### 443 **Chorus noise impacted more neuronal responses than stationary noise at each stage of the audi-** 444 **tory system**

445

446 Figure 2A shows rasters for recordings collected in the five auditory structures in response to the orig-  
447 inal (in quiet) and masked vocalizations embedded in stationary (top) and chorus (bottom) noise. In all  
448 structures, the neuronal responses evoked by the four whistles progressively vanished as the SNR de-  
449 creased from +10 to -10 dB. However, one can clearly notice that the recordings obtained in CNIC  
450 and MGv still display detectable responses at 0 dB SNR, even down to -10 dB for some vocalizations  
451 in CNIC.

452 To evaluate the neuronal resistance to noise, we quantified the Extraction Index (EI, see Methods sec-  
453 tion; Schneider and Woolley, 2013; Ni et al., 2017) of the 1267 recordings obtained in the five struc-  
454 tures. For each recording, the extraction index compares the PSTH obtained at a given SNR with the  
455 PSTHs obtained with the original vocalizations and with the noise alone: the higher the EI value  
456 (close to 1), the more the responses are signal-like (Figure 2B, left). Conversely, the lower the EI val-  
457 ue (close to -1), the more the responses are masker-like (Figure 2B, right). In both noises, the EI val-

458 ues were higher in the inferior colliculus and thalamus than in the cochlear nucleus and cortex, except  
459 in chorus noise at  $-10$  dB SNR, which strongly impacted all neuronal responses at each stage (Figure  
460 2C). In addition, the EI values obtained in chorus noise at 0 and  $-10$  dB SNR were significantly lower  
461 than those obtained in stationary noise in all structures except in the CN at 0 dB SNR (Figure 2C, one-  
462 way ANOVAs,  $p < 0.001$ ; with post-hoc paired t tests,  $p < 0.001$ ).

463 Thus, in all structures, both noises altered the evoked responses promoting masker-like responses, the  
464 chorus noise promoting, on average, a significantly higher proportion of masker-like responses than  
465 the stationary noise.

466

#### 467 **Robustness to noise is a distributed property in the auditory system**

468

469 We initially aimed at determining whether the four categories of cortical neurons (robust, balanced,  
470 insensitive and brittle) described by Ni and colleagues (2017) can also be found at each stage of the  
471 auditory system. For each neuronal recording, we computed six EI values (three for the stationary  
472 noise and three for the chorus noise, relative to the three SNRs). To do the clustering, we used the  
473 three EI values of all responses (i.e., the 1267 recordings) separately, in stationary and chorus noise.  
474 However, analyzing our whole database with the same clustering method and the same criterion (el-  
475 bow method) as in Ni and colleagues (2017) led us to consider either five or six clusters in both noises  
476 (Figure 3), potentially because our recordings came from three subcortical structures in addition to  
477 two cortical areas. When six clusters were defined, two of them displayed very similar behaviors with  
478 only slight differences in EI values at the three SNRs (see Figure 3B-C), which urged us to consider  
479 only five clusters and suggests also that a larger number of clusters would have been non-informative  
480 as similar behaviors should re-appear. Compared to the four categories of Ni and colleagues (2017),  
481 we added one new category, which represents an attenuated version of their robust category. These  
482 neurons cannot be neglected as they represent in fact a large proportion of our database (25% and 18%  
483 in the stationary and chorus noise).

484

485 Figure 4 presents the five categories derived from the whole data set across the three SNRs and the  
486 two noises. Figures 4A and 4F present the EI values of all neurons in the three SNRs in a given back-  
487 ground noise (with a color code from blue to red when progressing from low to high EI values). The  
488 five categories are indicated by a color bar on the right side and are derived from a continuum of EI  
489 values. This color code is used for the 3D representations of the five categories in the stationary (Fig-  
490 ure 4B) and chorus (Figure 4G) noise. Figure 4C shows the mean EI values in stationary noise for  
491 these five categories across the three SNRs and the percentage of neurons in each category is dis-  
492 played in figure 4D. Approximately 10% of the neurons exhibit signal-like responses characterized by

493 mean EI values greater than 0.5 at +10 and 0 dB SNRs. More than 25% display signal-dominated re-  
494 sponses characterized by mean EI values greater than 0.2 at +10 and 0 dB SNRs. About 5% of the  
495 neuronal responses are balanced and more than 40% of the total population has a mean EI value  
496 around 0 at all SNRs which corresponds to the insensitive responses. More than 10% of the auditory  
497 neurons have negative mean EI values at the three SNRs, which correspond to masker-like responses.  
498 Figures 4H and 4I show the mean EI values for these five categories in the chorus noise with, roughly,  
499 similar proportions of the five categories as in the stationary noise. However, in the chorus noise there  
500 were less signal-like (from 10% to 7.5%) and signal-dominated (from 27% to 20%) responses and  
501 more balanced responses (from 6.5% to 19.5%), whereas the proportion of insensitive responses re-  
502 mained similar (42-39.5%). Note also that in the chorus noise, the categories of signal-like and signal-  
503 dominated responses showed, on average, lower EI values at the 0 and -10 dB SNR than in stationary  
504 noise (compared Figure 4C and Figure 4H). Based upon these quantifications performed in the entire  
505 auditory system, we found five similar neuronal behaviors in the stationary and chorus noise.

506

507 What are the proportions of these categories in each structure? For each auditory structure, the per-  
508 centage of neurons from each category is presented in stationary noise (Figure 4E) and in chorus noise  
509 (Figure 4J). In stationary noise, signal-like and signal-dominated responses were mostly present in the  
510 inferior colliculus and thalamus, while the three other categories of neuronal responses classified as  
511 balanced, insensitive, and masker-like were mostly present in the cochlear nucleus and in the two cor-  
512 tical fields. Statistical analyses confirmed that the proportions of the different categories differed in IC  
513 and MGv compared with the three other structures (all Chi-Square;  $p < 0.001$ ). In chorus noise, there  
514 was a large increase in the proportion of balanced responses and a decrease in signal-like responses in  
515 all structures, but these latter neurons remained in higher proportions in IC and MGv than in CN and  
516 in cortex. Moreover, the masker-like responses were in equivalent proportions in all structures (be-  
517 tween 12.5% and 19%). In the CN, there was also an increase in the proportion of signal-dominated  
518 responses (from 7 to 14.5%). Statistical analyses confirmed that, in the chorus noise too, the propor-  
519 tions of the different categories differed in the IC and MGv compared with the three other structures  
520 (all Chi-Square;  $p < 0.001$ ).

521 Note that including the recordings with no significant TFRP ( $n=82$ ) led to exactly the same propor-  
522 tions of recordings in the different categories in both noises (Figure 4-1). These neurons, which did  
523 not respond to pure tones, displayed no signal-like response in subcortical structures and very rarely in  
524 the auditory cortex (see Figure 4-1D-H). In addition, in an attempt to evaluate the differences between  
525 the ventral and dorsal parts of the CN, we compared the EI values for a set of 87 recordings collected  
526 at the deepest electrode placements (assuming that they were potentially located in or close to VCN)  
527 versus the rest of the CN population considered as collected in DCN (Figure 4-2). The mean EI values



528 did not significantly differ between these two populations at the three SNRs in both noises (Figure 4-2  
529 A-B) and the proportions of the different categories were relatively similar in these two populations.  
530 Note also, that in the inferior colliculus, the recordings potentially obtained from the non-lemniscal  
531 divisions displayed (i) more balanced responses than in the lemniscal division in the stationary noise,  
532 and (ii) more insensitive responses than in the lemniscal division in the chorus noise (Figure 4-3).

533

534 To sum up, in both noises, the neurons with a high-fidelity representation of the signal were mostly  
535 present at two subcortical levels, in the inferior colliculus and thalamus. The insensitive responses  
536 showing no preference either for the signal or the noise were found in majority in the cochlear nucleus  
537 and in the auditory cortex. The balanced responses represented a small fraction of neurons in station-  
538 ary noise but were more numerous in the chorus noise, especially in IC and MGv. Interestingly, in  
539 both types of noise, the proportion of these balanced responses decreased as one ascends in the audito-  
540 ry system suggesting a decrease in sensitivity to SNR at the cortical level. Finally, the neurons with a  
541 high-fidelity representation of the noise were mostly localized in the cochlear nucleus in the stationary  
542 noise but were in an equivalent proportion in all structures in the chorus noise (between 12.5% and  
543 19%).

544

#### 545 **The noise-type sensitivity is found at each stage of the auditory system**

546

547 In the auditory cortex of awake marmoset, Ni and colleagues (2017) have pointed out that the neu-  
548 ron behavior in noise can be context-dependent: the behavior of a given neuron in a particular noise  
549 does not predict its behavior in another noise. Is this a property characterizing cortical neurons, or is it  
550 a general property that exists at all levels of the auditory system?

551 In each auditory structure, the neuronal behaviors were partly, but not completely, preserved in the  
552 two noises. On Figure 5A1, the group-switching matrix represents the percentage of neurons in a giv-  
553 en cluster in chorus noise depending on the cluster originally assigned in the stationary noise. The  
554 preservation of the same neuronal behavior in both noises is indicated by the percentages in the diag-  
555 onal line. About 50% of the signal-like and 40% of the signal-dominated neuronal responses in the  
556 stationary noise remained so in the chorus noise. Most of the balanced (73.5%) and insensitive neu-  
557 ron responses (65.5%) in the stationary noise remained also in the same category in the chorus  
558 noise. Only 36.5% of the masker-like neuronal responses remained so in the chorus noise. Figure 5A2  
559 indicates that the signal-like, signal-dominated and masker-like neuronal responses were the three  
560 categories with the highest percentages of category changes ( $\chi^2$   $p < 0.001$ ). In the different structures,  
561 the percentage of category changes was between 37 and 57% without significance difference between  
562 structures.

563 A bootstrap procedure was used to estimate the percentage of category changes, which can occur  
564 simply due to response variability (see Methods section). We suspected that, independently to the  
565 change in noise type, a recording can potentially change category because of its response variability  
566 and/or because it was located at the frontier between two clusters. Briefly, for each recording, and  
567 from the pool of 20 trials obtained for each stimulus, we resampled 20 trials (allowing repetitions),  
568 recomputed the Extraction Index and reallocated each resampled recording in the closest category.  
569 This entire procedure was performed 100 times for each recording. In all the following results, we  
570 only considered the recordings which remained in the same category 95 times or more (out of the 100  
571 bootstraps) in both noises, that is, recordings that displayed a very high reliability of their responses  
572 and were assigned to a particular category with a 95% confidence or more.

573 Figure 5B1 shows, for this population of 342 recordings, that a non-negligible fraction (36-75%) of  
574 the neurons assigned to a given category remained in the same category when shifting from the sta-  
575 tionary to the chorus noise. When analyzing the percentage of neurons changing categories, we found  
576 a similar pattern as the whole population of 1267 neurons, i.e., the largest proportions of neurons  
577 switching category from the stationary to the chorus noise were in the signal-like, signal-dominated  
578 and masker-like categories ( $\chi^2$   $p < 0.05$ , Figure 5B2). Analyzing the percentage of neurons changing  
579 clusters according to the structure revealed that the lowest percentages of category changes were in  
580 the cochlear nucleus and in A1 (21 and 22%) and the highest proportion in MGv (72%,  $\chi^2$   $p < 0.05$ ).

581 Note that when computing the mutual information based on all neurons (Figure 5-1A) or only reliable  
582 neurons (Figure 5-1B), we obtained low MI values (0.53 bits for all neurons and 0.7 bits for reliable  
583 neurons) confirming that a large proportion of neurons change category between noises.

584 Thus, even when using a bootstrap procedure with a severe selection criterion, a non-negligible frac-  
585 tion of recordings change categories from one background noise to another and this noise-type sensi-  
586 tivity is found at each stage of the auditory system.

587

588 **The neuronal behaviors in stationary and chorus noise are predictable based on response pa-**  
589 **rameters obtained in quiet**

590

591 A fundamental question is whether the assignment of a given recording to a particular category can be  
592 predicted from the response characteristics obtained by presenting pure tones and/or the original vo-  
593 calizations in quiet and/or the maskers alone. To address this question, we firstly focused on the neu-  
594 rons reliably categorized with the bootstrap procedure in both noises (with a confidence interval  $>$   
595 95%,  $n=342$ ).



596 To determine if the assignment of a recording to a particular category can be predicted based upon  
597 response characteristics, we trained an artificial classifier (linear discriminant analysis) with all com-  
598 binations of descriptor types.

599 Three descriptors were extracted from the responses to pure tones (TFRP): the best frequency (BF)  
600 firing rate, the bandwidth of tuning and the response duration. From the responses to the signal alone  
601 (original vocalizations) and the maskers alone, two descriptors were extracted: the firing rate and the  
602 temporal reliability (CorrCoef index, see Methods). Three other masker descriptors were used in both  
603 noises in an attempt to quantify the firing rate short-term adaptation to the masker (see Methods). Fi-  
604 nally, we included two descriptors corresponding to the ratios between the responses to the signal and  
605 to the masker (see Methods).

606 In stationary noise, for the descriptors extracted from the TFRP (Figure 6A-C), the distributions of  
607 best frequency (BF) firing rate and bandwidth of tuning did not point out significant differences across  
608 the five categories (Figures 6A-B), but the signal-like category showed significantly longer response  
609 duration than the other categories (Kruskal-Wallis test,  $p < 0.05$ , Figure 6C). For the descriptors ex-  
610 tracted from the signal responses (Figure 6D-E), the firing rate was significantly lower for the insensi-  
611 tive responses compared to all other response categories (Kruskal-Wallis test,  $p < 0.05$ , Figure 6D) and  
612 the signal-like and balanced responses had significantly higher CorrCoef values compared to three  
613 other categories (Kruskal-Wallis test,  $p < 0.05$ , Figure 6E). For the descriptors extracted from the re-  
614 sponses to the masker (Figure 6F-I), the signal-like, signal-dominated and insensitive neuronal re-  
615 sponses showed lower firing rate compared to the balanced and masker-like responses (Kruskal-  
616 Wallis test,  $p < 0.05$ , Figure 6F), whereas the CorrCoef values did not differ across categories except  
617 for the insensitive neuronal responses that showed the lowest CorrCoef values (Kruskal-Wallis test,  
618  $p < 0.05$ , Figure 6G). The masker accommodation (Figure 6H-I) was significantly lower for the signal-  
619 like and signal-dominated categories than the other categories, indicating that these two categories  
620 adapted more to the masker (their firing rate decreased during presentation of the masker alone). For  
621 the two descriptors combining the firing rate to the signal and to the masker (Figure 6J-K), the values  
622 were higher for the signal-like and signal-dominated categories than for the other categories (Figure  
623 6J-K) indicating that these two categories displayed marked preference for the signal over the masker  
624 (Kruskal-Wallis test,  $p < 0.05$ , Figure 6G).

625 In general, similar results were obtained in the chorus noise for TFRP and signal descriptors (Figure  
626 7A-E). However, more differences emerged across categories for the masker descriptors as the Cor-  
627 rCoef index (Figure 7G). The differences observed across categories were relatively comparable to the  
628 differences observed in the CorrCoef values obtained for the signal (Figure 7E).

629 Altogether, these analyses pointed out that there were little or no between-category differences based  
630 upon the TFRP descriptors and that the descriptors combining responses to the signal and to the

631 maskers provide information allowing a correct separation between the signal-like and signal-  
632 dominated categories vs. the others.

633 When the classifier was trained with all the descriptors in stationary noise (combination 1 in Figure  
634 8A, left column), the accuracy of the classifier reached 68.42%, which is more than three times the  
635 chance level (20% as there were five categories). Figure 8B presents the confusion matrix obtained  
636 with this classifier in stationary noise and revealed that the percentage of correct classification de-  
637 pends on the category. Indeed, the signal-like, the balanced and the insensitive neuronal responses  
638 were well predicted (67%, 64% and 91% respectively) whereas only 21% and 37% of the signal-  
639 dominated and masker-like neuronal responses were correctly predicted.

640 Next, for isolating which type of descriptors allowed the higher predictability, we used 14 different  
641 classifiers corresponding to the 14 possible combinations from the four types of descriptors (combina-  
642 tions 2-15 in Figure 8A). Several important results emerged from these analyses. First, the TFRP de-  
643 scriptors alone (line 8 in Figure 8A) led to the lowest accuracy, 27.78%, which is close from the  
644 chance level indicating that the responses to pure tones are insufficient to predict the behavior of a  
645 given neuron in noise. Second, the descriptors extracted either from the responses to the signal alone  
646 (line 12 of Figure 8A), or from the responses to the masker alone (line 14 of Figure 8A), or from the  
647 signal-to-masker alone (line 15 of Figure 8A) led to an accuracy between 56.73% and 58.77%, which  
648 was only 10% less than the best performance when combining all the descriptors. Finally, the de-  
649 scriptors extracted from the responses to the signal, to the masker and to the signal-to-masker ratio  
650 (line 9 of Figure 8A), generated an accuracy of 66.96%, which is close from the global level reached  
651 with all the descriptors. Therefore, the classifier reached a relative good performance in stationary  
652 noise (66.96%) by combining the three descriptors extracted from the responses to the signal alone  
653 and to the masker alone, the TFRP descriptors only slightly increasing the performance of the classifi-  
654 er (around 3%).

655 Globally, the results were similar in the chorus noise: Based on all the descriptors, the accuracy of the  
656 classifier was 69%, it dropped to 26% with the TFRP descriptors alone, and between 62% and 66%  
657 with the signal descriptors alone or masker descriptors alone (Figure 8A, left column). The only dif-  
658 ference was that the accuracy of the classifier with only the signal/masker descriptors was at the  
659 chance level (20.18%) probably because the firing rate to the masker is very close to the firing rate to  
660 the signal, suggesting that with a masker composed by a mixture of different type of signals (as the  
661 chorus noise), the masker firing rate does not significantly increase the performance of the classifier.  
662 Figure 8C illustrates the confusion matrix obtained by combining the four types of descriptors in cho-  
663 rus noise, which led to the maximum accuracy of the classifier (69.01%).

664 To verify that the descriptors are relevant predictors on the whole set of recordings, we trained the  
665 classifiers on the reliable neurons and tested them on the rest of the population (Figure 9). The accura-

666 cy of the classification was lower in both noises (43.24% and 49.84% in the stationary and chorus  
667 noise respectively, combination 1 in Figure 9) certainly because among our recordings, some did not  
668 display reliable enough responses. However, these values of accuracy were still more than twice the  
669 chance level. Again, the descriptors based on the TFRP provided an accuracy close to the chance level  
670 (31.24% and 23.35% in the stationary and chorus noise respectively, line 8 in Figure 9) and the de-  
671 scriptors based on the response to the signal alone (line 12 in Figure 9) and response to the masker  
672 alone (line 14 in Figure 9) generated an accuracy that was close from those obtained with the whole  
673 set of descriptors (43.03% and 42.38% in stationary noise; 41.62% and 43.24% in chorus noise).  
674 Altogether, our results pointed out that very few neuronal parameters (as the firing rate and the tem-  
675 poral reliability) to the signal alone and to the noise alone are sufficient to predict, up to a maximum  
676 of around 70%, the neuronal behaviors in noise.  
677

678

## Discussion

679

680 Here, we demonstrate that the robustness to noise of neuronal responses relies on a distributed net-  
681 work along the auditory system. Signal-like and signal-dominated responses were detected at each  
682 level of the auditory system but were in higher proportions at the collicular and thalamic levels. In  
683 terms of proportions, the highest fidelity representation of the signal or the noise was found at the  
684 subcortical level whereas at the cortical level, the majority of neuronal responses showed no prefer-  
685 ence for the signal or the noise suggesting that cortical neurons are less sensitive to the spectro-  
686 temporal details of the noisy vocalizations. Our results also indicate that neurons sensitive to the type  
687 of noise are present at each stage of the auditory system. Finally, a few neuronal parameters extracted  
688 from both the responses to the signal alone and to the noise alone convey enough information to pre-  
689 dict the neuronal behavior in noise up to a maximum of 70%.

690

### 691 Limitations of the study

692

693 Based on all EI values, the five categories rather form a continuum with no strict boundaries between  
694 them, which inevitably led us to « impose » the categories. Nonetheless, using a severe criterion of the  
695 bootstrap testing (a confidence interval  $> 95\%$ ), we found reliable neurons in each category, in each  
696 structure and in both noises. In addition, for a given recording, prediction about its assignment to a  
697 particular category reached about a maximum of 70% based upon a few descriptors of neuronal re-  
698 sponses. All these results suggest that these five behaviors do exist in the whole auditory system. The  
699 behaviors of cortical neurons found here in anesthetized animals were the same as previously de-  
700 scribed in awake marmosets (Ni et al., 2017). Therefore, the cortical representation of noisy signals by  
701 different neuronal categories characterized either by the preference of the signal, the masker, a sensi-  
702 tivity to SNR or an absence of these three acoustic features, seems independent of the fact that the  
703 animal is awake or anesthetized.

704 One can suspect that the higher proportion of signal-like responses in the CNIC and MGv compared  
705 to auditory cortex, results from the fact that our data were collected in anesthetized animals. Accord-  
706 ing to this view, if the recordings have been collected in awake, behaving animals, the results would  
707 be reversed, with a higher proportion of cortical neurons exhibiting robust responses to acoustic deg-  
708 radations. Confrontation of several results suggests that this simple explanation might not being cor-  
709 rect. First, Lohse and colleagues (2020) have recently demonstrated that collicular neurons of awake  
710 mice displayed the same gain control adaptation to the stimulus statistics than in anesthetized mice.

711 Second, auditory cortex responses collected in behaving ferrets were found to be sufficiently robust to  
712 preserve vowel identity across a large range of acoustic transformations, such as changes in funda-  
713 mental frequency, sound location or level (Town et al., 2018). However, earlier studies from the same  
714 laboratory performed in anesthetized conditions (Bizley et al., 2009, Walker et al., 2011) have reached  
715 similar conclusions for vowels varying in fundamental frequency and virtual acoustic location, indi-  
716 cating that the general principles allowing neuronal discrimination are observable across anesthetized  
717 and behavioral states. Furthermore, at the subcortical level, it seems that there is not a large difference  
718 between the phase-locking properties in anesthetized and awake animals. In fact, in awake animals,  
719 the subcortical neurons, especially collicular ones (Ter-Mikaelian et al., 2007), will still be far better  
720 than cortical ones to follow the 4-20Hz temporal cues contained in the four vocalizations, which are  
721 crucial cues for responding to these signals in noisy conditions (see figure 12 in Souffi et al., 2020).  
722 Thus, the differences observed here between cortical and subcortical structures in detecting and re-  
723 sponding to communication sounds in noisy conditions should remain the same in awake preparations.  
724

725 Using the same methods to classify and determine the optimal number of clusters as Ni and colleagues  
726 (2017), we opted for five categories rather than four, which is one main difference with their study.  
727 Our additional category corresponds to the signal-dominated responses and stands as an intermediate  
728 category between the signal-like and the insensitive responses. These responses might represented a  
729 too small fraction in their cortical data to emerge as a category, but the inclusion of three subcortical  
730 structures and a secondary cortical area might favored their emergence in larger proportions (25% and  
731 18% in the stationary and chorus noise, respectively). We can also wonder if choosing 7, 8, or 9 clus-  
732 ters, would have highlighted other neuronal behaviors. Part of the answer is provided by figures 3B-C,  
733 which show that with 6 clusters, similar behaviors re-appear suggesting that a larger number of cate-  
734 gories would have been non-informative. However, it is possible that more specific behaviors might  
735 have been missed as we collected multi-unit recordings composed of 2-6 shapes of action potentials.  
736 This is potentially the case at the cortical level where a large number of cell types have been described  
737 (Ascoli et al., 2008; DeFelipe et al., 2013) and also in the cochlear nucleus (Cant and Benson, 2003;  
738 Kuenzel, 2019). In fact, in the cochlear nucleus, multi-unit recordings might have masked the distinct  
739 temporal response profiles corresponding to different morphological cell-types. For example, the  
740 pause/build-up cells have been associated with the fusiform cells in the dorsal division of CN (Rhode  
741 et al., 1983; Smith and Rhode, 1985), whereas primary-like, onset and phase-locked patterns have  
742 been associated with the VCN globular bushy cells (Smith and Rhode, 1987). Because of our surgical  
743 approach, our recordings mainly (but not exclusively) come from the DCN.  
744 Obviously, it is also important to assess if our results can be generalized to other types of communica-  
745 tion sounds and to other types of masking noises. In fact, similar results were found in the initial study

746 by Ni and colleagues (2017) with five other different vocalizations and two other masking noises  
747 which suggests that with our experimental conditions, the results should not significantly change, at  
748 least in primary auditory cortex, with other types of communication sounds and other masking noises.

749

#### 750 **Robustness to noise in the auditory system: a localized vs. distributed property?**

751

752 In the A1 of awake marmosets, Ni and colleagues (2017) found about 20-30% of robust responses  
753 (depending on the vocalization), called here signal-like responses. In our cortical data, when pooling  
754 together the signal-like and signal-dominated responses, we obtained about the same proportions as in  
755 the marmoset A1 (33%). In the bird auditory system, Schneider and Woolley (2013) described the  
756 emergence of noise-invariant responses for a subset of cells (the broad spike cells) of a secondary au-  
757 ditory area (area NCM), whereas upstream neurons (IC and A1 neurons in their study) represent vo-  
758 calizations with dense and background-corrupted responses. They suggest that a sparse coding scheme  
759 operating within NCM allows the emergence of this noise-invariant representation. In our study (and  
760 in the mammalian A1 in general), a sparse representation already exists as early as A1 (see the rasters  
761 in Figure 2A, see also Hromádka et al., 2008) allowing signal-like and signal-dominated responses to  
762 be present in about the same proportions in A1 and in the secondary area VRB.  
763 Noise-invariant representations were also reported in A1 of anesthetized ferrets (Rabinowitz et al.,  
764 2013). This study suggested a progressive emergence of noise-invariant responses from the auditory  
765 nerve to IC and to A1, and proposed the adaptation to the noise statistics as a key mechanism to ac-  
766 count for the noise-invariant representation in A1. However, Lohse and colleagues (2020) have re-  
767 cently challenged this result by showing, in anesthetized animals too, (i) that collicular, thalamic and  
768 cortical neurons display the same adaptation to noise statistics and (ii), importantly, that silencing the  
769 auditory cortex did not affect the capacity of IC and MGv neurons to adapt to noise statistics.

770 In fact, Las and colleagues (2005) reported that A1, MGB and IC neurons can detect low-intensity  
771 target tones in a louder fluctuating masking noise and display the so-called “phase-locking suppres-  
772 sion”, that is the interruption of phase-locking to the temporal envelope of background noise. This last  
773 result indicates that both IC and MGB neurons have the same ability as cortical neurons to detect low-  
774 intensity target sounds in louder background noises (even at -15 or -35 dB SNR). Thus, the robustness  
775 of some of our subcortical neurons may stem from this ability to detect the vocalizations even at SNRs  
776 as low as the -10 dB SNR.

777 Based upon the proportion of signal-like and signal-dominated responses, it seems that the robustness  
778 to noise peaks in CNIC, with the MGv neurons being at the intermediate level between IC and A1  
779 (Figures 4E, 4J). In fact, our results point out an abrupt change from a prominent noise-sensitivity in  
780 CN to a prominent noise-robustness in IC, which means that this robustness is generated by neural



781 computation taking place in the central auditory system. Whether this is an intrinsic property emerg-  
782 ing *de novo* in the IC or whether this property emerges as a consequence of the multiple inputs con-  
783 verging upon IC cells (Malmierca and Ryugo, 2011) remains to be determined. Several studies have  
784 clearly demonstrated that IC neurons adapt to the stimulus statistics. First, adaptations of IC neurons  
785 to the average stimulus intensity, stimulus variance and bimodality that has already been described  
786 with a temporal decay of about 160 ms at 75 dB (Dean et al., 2005; 2008). Second, adaptation to the  
787 noise statistics shifted the temporal modulation function (TMF) of IC neurons to slower modulations,  
788 sometimes transforming band-pass TMF to low pass TMF in about 200 ms of noise presentation  
789 (Lesica and Groethe, 2008). In addition, recent studies have shown that a tone-in-noise discrimination  
790 task influences neuronal activity as early as the inferior colliculus (Slee and David, 2015, Shaheen et  
791 al., 2020) suggesting that subcortical structures may participate to complex auditory tasks and should  
792 not be considered as passive relays.

793 A particularly interesting result is that, in both types of noise, the proportion of neurons classified as  
794 balanced (i.e. showing strong SNR-dependence) decreased progressively as one ascends in the audito-  
795 ry system, which is in line with the idea that cortical neurons are less sensitive to SNR than subcorti-  
796 cal ones.

797

#### 798 **Noise-type sensitivity and noise representation in the auditory system**

799

800 Ni and colleagues (2017) found about two-thirds of A1 neurons switching category from one back-  
801 ground noise to another, suggesting that the majority of cortical neurons have a behavior specific to  
802 the type of noise. We preferred to call this phenomenon noise-type sensitivity rather context-  
803 dependence (proposed by Ni and colleagues, 2017) because the latter refers to situations where the  
804 same stimulus is presented in different contexts; whereas inserting signal stimuli in two types of noise  
805 generated different auditory streams.

806 Here, we confirmed that these neurons exist at the level of the primary auditory cortex and extended  
807 this result to the subcortical structures. Therefore, different types of noise streams activate different  
808 subpopulations of neurons at each stage of the auditory system for constructing invariant representa-  
809 tions of communication sounds in noise. In addition, based on a restrictive population of neurons that  
810 have a high response reliability as they remained in the same category with a bootstrap procedure (i.e.,  
811 the reliable neurons), we also found such neurons that switch categories between the two noises in all  
812 auditory structures. However, although we initially found around 40% of such neurons in A1 (Figure  
813 5A2), the bootstrap procedure indicated that more realistic percentages should be much lower, poten-  
814 tially around 20% (Figure 5B2). The response variability, which is probably much larger in awake  
815 than in anesthetized animals (Edeline et al., 2000, 2001; Huetz et al., 2009), can potentially explain

816 the difference between our results and those of Ni and colleagues (2017). Here, these neurons were  
817 detected in auditory cortex but were found in higher proportions in the inferior colliculus and in the  
818 auditory thalamus. This indicates that these subcortical neurons might be more sensitive to the sound  
819 streams in which the signals were embedded. It is interesting to note that signal-like responses in sta-  
820 tionary noise became as much balanced as they remained signal-like in chorus noise. Since in chorus  
821 noise, the signal and masker are very close (both spectrally and temporally), this change of category  
822 from signal-like to balanced was predictable. As already mentioned by Ni and colleagues (2017), if a  
823 larger number of noise types would have been tested, the proportion of neurons within each category  
824 would have been different. For example, a larger fraction of neuronal responses can potentially be  
825 considered as signal-like or masker-like, because masker-like responses in a particular type of noise  
826 can be the signal-like ones in another noise. Our results show that their assumption, if valid, does not  
827 only concern the primary auditory cortex.

828 Robust perception of speech in humans or vocalizations in animals probably also requires a robust  
829 representation of competing sounds (here, masking noise). This can be the functional role of the neu-  
830 rons presenting masker-like responses, which are potentially crucial to determine the characteristics of  
831 the noise type and to provide an accurate representation of it within the auditory stream reaching our  
832 ears at any time. They were detected here, in higher proportion in the CN in stationary noise, but they  
833 became more numerous and in equivalent proportion in all structures in chorus noise. Therefore, the  
834 noise representation can be based upon the neuronal activity in the cochlear nucleus in stationary  
835 noise, whereas this representation can be more distributed in the chorus noise potentially because this  
836 noise has more naturalistic temporal properties leading to activate more neurons than the stationary  
837 noise.

838

### 839 **Predictors of neuronal behavior in noise**

840

841 Using classifiers trained with different types of descriptors, we looked for characteristics from the  
842 responses to the pure tones (i.e., the TFRPs), to the signal alone and to the noise alone for predicting  
843 the assignment of a given recording in a particular category. We pointed out that the TRFP parameters  
844 led to an accuracy of the classifier very close to the chance level indicating that the basic static filter-  
845 ing properties derived from the TFRPs did not predict the neuronal behaviors in noise. When adding  
846 the descriptors of both the signal alone and noise alone (as the firing rate and the temporal reliability),  
847 the classifier reached an accuracy up to around 70% in both noises which strongly suggests that re-  
848 sponses to signal alone and to noise alone contained enough information to predict the behavior in  
849 noise of a given recording. In a previous study testing the responses of IC cells to vocalizations in  
850 noise, it was shown that despite no consistent effect of the mean firing rate, the temporal reliability



851 was decreased by half (Lesica and Grothe, 2008). Under these conditions, IC neurons were still effi-  
852 cient in detecting vocalizations in noise. Using these physiological results in a computational model,  
853 this study also pointed out that under noisy conditions, lowpass filtering the noisy vocalizations is the  
854 most efficient strategy to code the stimuli because it preserved the power at low modulation frequen-  
855 cies and the temporal reliability of responses.

856 Together, these results suggest that the more temporally precise are the synaptic inputs converging  
857 onto a particular neuron, the more robust is the response of that neuron in background noise. This is  
858 true for encoding slow amplitude modulations, which are among the most efficient cues to discrimi-  
859 nate speech and communication sounds (Shannon et al., 1995; Zeng et al. 2005; Souffi et al., 2020).

860

#### 861 **General conclusion**

862

863 Here, we propose that the noise-robustness observed in many studies at the cortical level stems, at  
864 least partially, from subcortical mechanisms (Lesica and Grothe, 2008; Lohse et al., 2020). There-  
865 fore, the auditory cortex potentially inherits adaptation from earlier levels, allowing the cortical net-  
866 works to focus on higher-level processing such as classifying the target stimuli into phonetic or lin-  
867 guistic features (Mesgarani et al., 2014), segregating the different auditory streams (Mesgarani and  
868 Chang, 2012) integrating multimodal information (Deneux et al., 2019), and retaining behaviorally  
869 important stimuli in short term (Huang et al., 2016) or long term memory (Moczulska et al., 2013;  
870 Concina et al., 2019).

871

872

873

874

875

876

877

878

879

880

881

882

883

884

885

886

887

888

889

890

891

892

893

894

895

896

897

898

899

900

901

902

903

904

905

906

907

908

909

910

911

912

913

914

915

916

917

918

919

920

921

922

923

924

925

926

927

928

929

930

931

932

933

## References

- Anderson LA, Wallace MN & Palmer AR (2007) Identification of subdivisions in the medial geniculate body of the guinea pig. *Hearing Research* **228**, 156–167.
- Ascoli GA, Alonso-Nanclares L, et al. (2008) Petilla terminology: nomenclature of features of GABAergic interneurons of the cerebral cortex. *Nat Rev Neurosci.* 9(7):557-568.
- Aushana Y, Souffi S, Edeline JM, Lorenzi C & Huetz C (2018) Robust Neuronal Discrimination in Primary Auditory Cortex Despite Degradations of Spectro-temporal Acoustic Details: Comparison Between Guinea Pigs with Normal Hearing and Mild Age-Related Hearing Loss. *J. Assoc. Res. Otolaryngol.* **19**, 163–180.
- Beetz MJ, García-Rosales F, Kössl M & Hechavarría JC (2018) Robustness of cortical and subcortical processing in the presence of natural masking sounds. *Sci Rep* **8**, 6863.
- Bizley JK, Walker KMM, Silverman BW, King AJ & Schnupp JWH (2009) Interdependent Encoding of Pitch, Timbre, and Spatial Location in Auditory Cortex. *Journal of Neuroscience* **29**, 2064–2075.
- Cant NB, Benson CG (2003) Parallel auditory pathways: projection patterns of the different neuronal populations in the dorsal and ventral cochlear nuclei. *Brain Res Bull.* 60(5-6):457-474.
- Concina, G, Renna, A, Grosso, A & Sacchetti, B (2019) The auditory cortex and the emotional valence of sounds. *Neurosci Biobehav Rev* **98**, 256–264.
- Dean I, Harper NS & McAlpine D (2005) Neural population coding of sound level adapts to stimulus statistics. *Nat. Neurosci.* **8**, 1684–1689.
- Dean I, Robinson BL, Harper NS & McAlpine D (2008) Rapid neural adaptation to sound level statistics. *J. Neurosci.* **28**, 6430–6438.
- DeFelipe J, López-Cruz PL, Benavides-Piccione R, et al. (2013) New insights into the classification and nomenclature of cortical GABAergic interneurons. *Nat Rev Neurosci.* 14(3):202-216.
- Deneux T, Harrell ER, Kempf A, Ceballos S, Filipchuk A, Bathellier B (2019) Context-dependent signaling of coincident auditory and visual events in primary visual cortex. *Elife.* May 23;8:e44006.
- Edeline JM, Dutrieux G, Manunta Y & Hennevin E (2001) Diversity of receptive field changes in auditory cortex during natural sleep. *Eur. J. Neurosci.* **14**, 1865–1880.
- Edeline JM, Manunta Y & Hennevin E (2000) Auditory Thalamus Neurons During Sleep: Changes in Frequency Selectivity, Threshold, and Receptive Field Size. *Journal of Neurophysiology* **84**, 934–952.
- Edeline JM, Manunta Y, Nodal F & Bajo V (1999) Do auditory responses recorded from awake animals reflect the anatomical parcellation of the auditory thalamus? *Hearing Research*, 131, 135-152.
- Franke F, Quiñero R, Hierlemann A, Obermayer K (2015) Bayes optimal template matching for spike sorting - combining fisher discriminant analysis with optimal filtering. *J Comput Neurosci.* 38(3):439-59.
- Gaucher Q & Edeline JM (2015) Stimulus-specific effects of noradrenaline in auditory cortex: implications for the discrimination of communication sounds. *J. Physiol. (Lond.)* **593**, 1003–1020.
- Gaucher Q, Edeline JM & Gourévitch B (2012) How different are the local field potentials and spiking activities? Insights from multi-electrodes arrays. *J. Physiol. Paris* **106**, 93–103.
- Gaucher Q, Huetz C, Gourévitch B & Edeline JM (2013a) Cortical inhibition reduces information redundancy at presentation of communication sounds in the primary auditory cortex. *J. Neurosci.* **33**, 10713–10728.
- Gourévitch B & Edeline JM (2011) Age-related changes in the guinea pig auditory cortex: relationship with brainstem changes and comparison with tone-induced hearing loss. *Eur. J. Neurosci.* **34**, 1953–1965.
- Gourévitch B, Doisy T, Avillac M & Edeline JM (2009) Follow-up of latency and threshold shifts of auditory brainstem responses after single and interrupted acoustic trauma in guinea pig. *Brain Res.* **1304**, 66–79.
- Grimsley JMS, Shanbhag SJ, Palmer AR & Wallace MN (2012) Processing of Communication Calls in Guinea Pig Auditory Cortex. *PLoS ONE* **7**, e51646.
- Hromádka T, DeWeese MR & Zador AM (2008) Sparse Representation of Sounds in the Unanesthetized Auditory Cortex. *PLoS Biol* **6**, e16.
- Huang Y, Matysiak A, Heil P, König R, Brosch M. (2016) Persistent neural activity in auditory cortex is related to auditory working memory in humans and nonhuman primates. *Elife.* Jul 20;5:e15441.
- Huetz C, Philibert B & Edeline JM (2009) A spike-timing code for discriminating conspecific vocalizations in the thalamocortical system of anesthetized and awake guinea pigs. *J. Neurosci.* **29**, 334–350.
- Kuenzel T (2019) Modulatory influences on time-coding neurons in the ventral cochlear nucleus. *Hear Res.*;384:107824.
- Las L, Stern EA & Nelken I (2005) Representation of tone in fluctuating maskers in the ascending auditory system. *J. Neurosci.* **25**, 1503–1513.
- Lesica NA & Grothe B (2008) Efficient Temporal Processing of Naturalistic Sounds. *PLoS ONE* **3**, e1655.
- Lohse M, Bajo VM, King AJ, Willmore BDB. Neural circuits underlying auditory contrast gain control and their perceptual implications. *Nat Commun.* 2020 Jan 16;11(1):324.
- Malmierca MS & Ryugo DK Descending Connections of Auditory Cortex to the Midbrain and Brain Stem. in *The Auditory Cortex* (eds. Winer, J. A. & Schreiner, C. E.) 189–208 (Springer US, 2011).
- Malmierca MS, Le Beau FE, Rees A. (1996) The topographical organization of descending projections from the central nucleus of the inferior colliculus in guinea pig. *Hear Res. Apr*;93(1-2):167-80.

- 934 Malmierca MS, Rees A, Le Beau FE, Bjaalie JG (1995) Laminar organization of frequency-defined local axons within and between the inferior colliculi of the guinea pig. *J Comp Neurol.* Jun 19;357(1):124-44.
- 935
- 936 Mesgarani N & Chang EF (2012) Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*
- 937 **485**, 233–236.
- 938 Mesgarani N, David SV, Fritz JB & Shamma SA (2014) Mechanisms of noise robust representation of speech in primary auditory cortex. *Proceedings of the National Academy of Sciences* **111**, 6792–6797.
- 939
- 940 Moczulska KE *et al.* (2013) Dynamics of dendritic spines in the mouse auditory cortex during memory formation and memory recall. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 18315–18320.
- 941
- 942 Narayan R *et al.* (2007) Cortical interference effects in the cocktail party problem. *Nat Neurosci* **10**, 1601–1607.
- 943 Ni R, Bender DA, Shaneci AM, Gamble JR & Barbour DL (2017) Contextual effects of noise on vocalization encoding in primary auditory cortex. *Journal of Neurophysiology* **117**, 713–727.
- 944
- 945 Occelli F, Suied C, Pressnitzer D, Edeline JM & Gourévitch BA (2016) Neural Substrate for Rapid Timbre Recognition? Neural and Behavioral Discrimination of Very Brief Acoustic Vowels. *Cereb. Cortex* **26**, 2483–2496.
- 946
- 947 Paraouty N, Stasiak A, Lorenzi C, Varnet L & Winter IM (2018) Dual Coding of Frequency Modulation in the Ventral Cochlear Nucleus. *J. Neurosci.* **38**, 4123–4137.
- 948
- 949 Pouzat C, Delescluse M, Viot P, Diebolt J. (2004) Improved spike-sorting by modeling firing statistics and burst-dependent spike amplitude attenuation: a Markov chain Monte Carlo approach. *J Neurophysiol.* 91(6):2910-28.
- 950
- 951 Quiroga RQ, Nadasdy Z, Ben-Shaul Y. (2004) Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Comput.* 16(8):1661-87.
- 952
- 953 Rabinowitz NC, Willmore BDB, King AJ & Schnupp JWH (2013) Constructing Noise-Invariant Representations of Sound in the Auditory Pathway. *PLoS Biol* **11**, e1001710.
- 954
- 955 Redies H, Brandner S & Creutzfeldt OD (1989) Anatomy of the auditory thalamocortical system of the guinea pig. *The Journal of Comparative Neurology* **282**, 489–511.
- 956
- 957 Rees A, Sarbaz A, Malmierca MS & Le Beau FEN (1997) Regularity of Firing of Neurons in the Inferior Colliculus. *Journal of Neurophysiology* **77**, 2945–2965.
- 958
- 959 Rhode WS, Smith PH, Oertel D. (1983) Physiological response properties of cells labeled intracellularly with horseradish peroxidase in cat dorsal cochlear nucleus. *J Comp Neurol.* Feb 1;213(4):426-47.
- 960
- 961 Rutkowski RG, Shackleton TM, Schnupp JWH, Wallace MN & Palmer AR (2002) Spectrotemporal Receptive Field Properties of Single Units in the Primary, Dorsocaudal and Ventrostroral Auditory Cortex of the Guinea Pig. *Audiology and Neurotology* **7**, 214–227.
- 962
- 963 Schneider DM & Woolley SMN (2013) Sparse and Background-Invariant Coding of Vocalizations in Auditory Scenes. *Neuron* **79**, 141–152.
- 964
- 965 Shaheen LA, Slee SJ, David SV (2021) Task Engagement Improves Neural Discriminability in the Auditory Midbrain of the Marmoset Monkey. *J Neurosci.* Jan 13;41(2):284-297.
- 966
- 967 Shannon, CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27(3):379–423.
- 968
- 969 Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. *Science.* Oct 13;270(5234):303-4.
- 970
- 971 Slee SJ, David SV (2015) Rapid Task-Related Plasticity of Spectrotemporal Receptive Fields in the Auditory Midbrain. *J Neurosci.* Sep 23;35(38):13090-102.
- 972
- 973 Smith PH, Rhode WS. (1985) Electron microscopic features of physiologically characterized, HRP-labeled fusiform cells in the cat dorsal cochlear nucleus. *J Comp Neurol.* Jul 1;237(1):127-43.
- 974
- 975 Souffi S, Lorenzi C, Varnet L, Huetz C, Edeline JM. (2020) Noise-sensitive but more precise subcortical representations co-exist with robust cortical encoding of natural vocalizations. *J. Neurosci.* **27**, 5228-5246.
- 976
- 977 Ter-Mikaelian, M., Sanes, D. H. & Semple, M. N. (2007) Transformation of Temporal Properties between Auditory Midbrain and Cortex in the Awake Mongolian Gerbil. *J. Neurosci.* **27**, 6091–6102.
- 978
- 979 Town SM, Wood KC, Bizley JK. (2018) Sound identity is represented robustly in auditory cortex during perceptual constancy. *Nat Commun.* 9(1):4786.
- 980
- 981 Walker KMM, Bizley JK, King AJ & Schnupp JWH (2011) Multiplexed and Robust Representations of Sound Features in Auditory Cortex. *Journal of Neuroscience* **31**, 14565–14576.
- 982
- 983 Wallace MN & Palmer AR (2008) Laminar differences in the response properties of cells in the primary auditory cortex. *Exp Brain Res* **184**, 179–191.
- 984
- 985 Wallace MN, Anderson LA & Palmer AR (2007) Phase-Locked Responses to Pure Tones in the Auditory Thalamus. *Journal of Neurophysiology* **98**, 1941–1952.
- 986
- 987 Wallace, MN, Rutkowski RG & Palmer AR (2000) Identification and localisation of auditory areas in guinea pig cortex. *Experimental Brain Research* **132**, 445–456.
- 988
- 989 Zeng FG, Nie K, Stickney GS, Kong YY, Vongphoe M, Bhargava A, Wei C, Cao K. (2005) Speech recognition with amplitude and frequency modulations. *Proc Natl Acad Sci U S A.* Feb 15;102(7):2293-8.
- 990
- 991
- 992
- 993
- 994

995

## Figure legends

996

997 **Figure 1. Original and noisy vocalizations.** **A.** Waveforms (*top*) and spectrograms (*bottom*) of the  
 998 four original whistles used in this study. **B-C.** Spectrograms of the four whistles in stationary (**B**) and  
 999 chorus (**C**) noise at three SNRs (+10, 0 and -10 dB, *from top to bottom*) and the noise only. The  
 1000 frequency range for all spectrograms is 0–30 kHz and all spectrograms share the same color scale  
 1001 (covering a range of 50 dB).

1002

1003 **Figure 2. The decrease in EI values is more pronounced in chorus noise than in stationary noise**  
 1004 **in each auditory structure.** **A.** Raster plots of responses of the four original vocalizations, noisy  
 1005 vocalizations (in both noises) and noise alone recorded in CN, CNIC, MGv, A1 and VRB. The grey  
 1006 part of rasters corresponds to the evoked activity. For each structure, all the rasters correspond to the  
 1007 same recording. **B.** Rasters showing examples of neuronal responses in stationary noise with values of  
 1008  $EI > 0$  corresponding to a signal-like response (left, IC recording) and  $EI < 0$  corresponding to a  
 1009 masker-like response (right, A1 recording). Top panels show the responses to the original  
 1010 vocalizations, the middle panels the responses to vocalizations at the 0 dB SNR in stationary noise and  
 1011 the bottom panels the responses to stationary noise alone. **C.** Box plots showing the EI values for the  
 1012 three SNRs obtained in CN (in black), CNIC (in green), MGv (in orange), A1 (in blue) and VRB (in  
 1013 purple) alternatively in stationary (SN) and chorus noise (CN). In each box plot, the red dot represents  
 1014 the mean value. The black lines represent significant differences between the mean values (one-way  
 1015 ANOVAs,  $P < 0.001$ ; with post-hoc paired t tests,  $P^a_{+10dB, CN} = 4.03e-18$ ,  $P^b_{+10dB, CNIC} = 1.45e-07$ ,  $P^c_{0dB, CNIC} = 2.47e-45$ ,  
 1016  $P^d_{0dB, MGv} = 2.11e-30$ ,  $P^e_{0dB, A1} = 5.41e-25$ ,  $P^f_{0dB, VRB} = 6.36e-10$ ,  $P^g_{-10dB, CN} = 5.74e-12$ ,  
 1017  $P^h_{-10dB, CNIC} = 1.83e-59$ ,  $P^i_{-10dB, MGv} = 1.16e-36$ ,  $P^j_{-10dB, A1} = 2.84e-24$ ,  $P^k_{+10dB, VRB} = 4.62e-11$ ).

1018

1019

1020

1021

1022

1023

1024

1025

1026

1027

1028

1029

1030

1031

1032

1033

1034

1035

1036

1037

1038

1039

1040

1041

1042

1043

1044

1045

1046

1047

1048

1049

1050

1051

1052

1053

1054

1055

1056

1057

1058

1059

1060

1061

1062

1063

1064

1065

1066

1067

1068

**Figure 3. The choice of five clusters is optimal to reveal the different behaviors in both noises.**

**A.** Mean square error of EI profile clustering as a function of the number of clusters using the K-means algorithm for the stationary and chorus noise. **B-C.** Population average EI profile ( $\pm$ SEM) of each cluster when considering six clusters to separate the data in the stationary noise (**B**) and in the chorus noise (**C**). Note that in both noises, two clusters have similar mean EI profile, i.e., the same EI evolution across the three SNRs, (the two grey clusters in B and the two blue clusters in C) leading us to consider only five clusters in the following results (Figure 4).

**Figure 4. Robustness to noise is a distributed property in the auditory system.**

**A.** Each row corresponds to the EI profile of a given neuronal recording obtained in the five auditory structures in stationary noise with a color code from blue to red when progressing from low to high EI values. On the right, five stacked colors delineate the identity for the five categories of responses. The signal-like category is in green, the signal-dominated category in pink, the balanced category in turquoise, the insensitive category in gray and the masker-like category in yellow. The names of the categories used in the study by Ni and colleagues (2017) are provided for comparison. **B-E.** 3D representation of the five categories in stationary noise (**B**), mean EI values ( $\pm$ SEM) of the five categories (**C**), relative proportions of each category in stationary noise (**D**) and proportion of each category in the five auditory structures from CN to VRB (**E**).

**F-J.** Same representations as in A, B, C, D and E for the responses collected in the chorus noise.

See Table 1 selection type (b) for referring to the number of selected recordings in each structure.

**Figure 4-1. Similar results as in figure 4 are obtained taking into account the neurons without significant TRFP.**

**A-C.** Mean EI values ( $\pm$ SEM) of the five categories across the three SNRs (+10, 0 and -10 dB) (**A**), relative proportions of each category in stationary noise (**B**) and proportion of each category in the five auditory structures from CN to VRB for the recordings with or without significant TRFP (Time-Frequency Response Profile, relative to pure tone responses) (**C**).

1046 **D.** Proportion of each category in the five auditory structures from CN to VRB for the recordings  
1047 without significant TFRP.

1048 **E-H.** Same representations as in A, B, C and D for the chorus noise.

1049 See Table 1 (selection type (a)) for referring to the number of selected recordings in each structure.

1050

1051 **Figure 4-2. Attempt to separate the ventral and dorsal parts of the cochlear nucleus based on**  
1052 **the depth of the recordings.**

1053 **A-B.** Box plots showing the EI values for the three SNRs obtained in the ventral (VCN, in black) and  
1054 dorsal (DCN, in grey) parts of the cochlear nucleus in stationary (A) and chorus noise (B). In each box  
1055 plot, the red dot represents the mean value. There was no significant difference between the EI values  
1056 obtained in VCN and DCN for the three SNRs and in both noises (one-way ANOVAs,  $P > 0.05$ ; with  
1057 post-hoc paired t tests,  $P > 0.05$ ). **C.** Proportion of each category in the ventral part of the cochlear  
1058 nucleus (VCN,  $n=87$  recordings) and its dorsal part (DCN,  $n=302$  recordings) obtained in stationary  
1059 (SN) and chorus (CN) noise.

1060

1061 **Figure 4-3. Lemniscal and non-lemniscal parts of the inferior colliculus.**

1062 Proportion of each category in the lemniscal part of inferior colliculus (central nucleus of inferior  
1063 colliculus, CNIC,  $n=339$  recordings) and its non-lemniscal parts (dorsal and external cortices of the  
1064 inferior colliculus, DCIC-ECIC,  $n=73$  recordings) obtained in stationary and chorus noise. See Table  
1065 1 (selection type (b)) for referring to the number of selected recordings.

1066

1067 **Figure 5. The noise-type sensitivity is found at each stage of the auditory system.**

1068 **A1.** Group-switching matrix representing the percentage of recordings in a given category in chorus  
1069 noise depending on the category originally assigned in the stationary noise. The abscissas indicate the  
1070 category identity in the stationary noise and the ordinates represent the category identity in the chorus  
1071 noise.

1072 For example, signal-like responses in stationary noise are also 50% signal-like in chorus noise but  
1073 10% are reclassified as signal-dominated, 35% balanced, 1.5% insensitive and 3.5% masker-like. Note  
1074 that, in stationary noise, the number of recordings in each category were 139, 346, 83, 540 and 159 in  
1075 signal-like, signal-dominated, balanced, insensitive and masker-like category respectively.

1076 **A2.** Mean percentages of recordings changing category from the stationary noise to the chorus noise,  
1077 first in each category and second in each structure (VRB, (in purple), A1 (in blue), MGv (in orange),  
1078 CNIC (in green) and CN (in black)).

1079 **B1.** Group-switching matrix representing the percentages of recordings changing category from the  
1080 stationary noise to the chorus noise based only on recordings considered as reliable with the bootstrap  
1081 procedure in the two types of noise (with a confidence interval  $\geq 95\%$ ).

1082 **B2.** Mean percentages of recordings changing category from the stationary noise to the chorus noise,  
1083 first in each category and second in each structure (VRB, (in purple), A1 (in blue), MGv (in orange),  
1084 CNIC (in green) and CN (in black)).

1085

1086 **Figure 5-1. Confusion matrices obtained for all and reliable recordings.**

1087 **A.** Confusion matrix relative to Figure 5A1 representing the number of recordings in a given category  
1088 in chorus noise depending on the category originally assigned in the stationary noise.

1089 **B.** Same as in (A) with only recordings considered as reliable with the bootstrap procedure in the two  
1090 types of noise (with a confidence interval  $\geq 95\%$ ). This figure is relative to the Figure 5B1.

1091 Note that the mutual information (MI, bits) values were low in (A) and (B) corroborating the fact that  
1092 a large proportion of recordings assigned to a given category in stationary noise change category in  
1093 chorus noise.

1094

1095

1096 **Figure 6. Descriptors of the categories in stationary noise used by the classifiers.**



1097 **A-C.** Three TFRP parameters were chosen as descriptors: the best frequency (BF) firing rate, the  
 1098 bandwidth and the response duration. **D-E.** Two signal descriptors were selected corresponding to the  
 1099 firing rate and the CorrCoef values obtained in original conditions. **F-G.** Two main masker descriptors  
 1100 were presented corresponding to the firing rate and the CorrCoef values obtained in stationary noise  
 1101 alone. **H-I.** The three other masker descriptors are: (**H**) the ratio between the masker firing rate taken  
 1102 at the time the signal should have occurred and the initial masker firing rate during the first 200 ms of  
 1103 the masker and the number of action potentials emitted during the first (Initial, **I**) and last (Final, **I**) 50  
 1104 ms of the masker alone over a 564 ms-period. **J-K.** Two descriptors of the signal-to-masker ratio are  
 1105 presented and taken into account the firing rate of responses to the signal and to the masker; the two  
 1106 differ only on which part of the response to the masker is taken into account (see Methods). For each  
 1107 violin plot, the red dot represents the median value and the black lines represent significant  
 1108 differences between the median values (Kruskal-Wallis test,  $p < 0.05$ ).

1109  
 1110 **Figure 7. Descriptors of the categories in chorus noise used by the classifiers.**

1111 **A-C.** Three TFRP parameters were chosen as descriptors: the best frequency (BF) firing rate, the  
 1112 bandwidth and the response duration. **D-E.** Two signal descriptors were selected corresponding to the  
 1113 firing rate and the CorrCoef values obtained in original conditions. **F-G.** Two main masker descriptors  
 1114 were presented corresponding to the firing rate and the CorrCoef values obtained in chorus noise  
 1115 alone. **H-I.** The three other masker descriptors are: (**H**) the ratio between the masker firing rate taken  
 1116 at the time the signal should have occurred and the initial masker firing rate during the first 200 ms of  
 1117 the masker and the number of action potentials emitted during the first (Initial, **I**) and last (Final, **I**) 50  
 1118 ms of the masker alone over a 564 ms-period.

1119 **J-K.** Two descriptors of the signal-to-masker ratio are presented and taken into account the firing rate  
 1120 of responses to the signal and to the masker; the two differ only on which part of the response to the  
 1121 masker is taken into account (see Methods). For each violin plot, the red dot represents the median  
 1122 value and the black lines represent significant differences between the median values (Kruskal-Wallis  
 1123 test,  $p < 0.05$ ).

1124  
 1125 **Figure 8. The neuronal behaviors in stationary and chorus noise are predictable based on**  
 1126 **response parameters obtained in quiet.**

1127 **A.** All tested combinations (1-15) based on four types of descriptors (TFRP, Signal, Masker and  
 1128 Signal/Masker) of the categories in stationary noise and their respective percentages of accuracy of  
 1129 the classifier. The gray part means that the descriptor is included in the classifier and the white part  
 1130 means that the descriptor is excluded from the classifier.

1131 **B-C.** Example of the confusion matrix obtained with all descriptors (combination 1) in stationary (**B**)  
 1132 and chorus (**C**) noise. Each row corresponds to a true category and each column corresponds to a  
 1133 predicted category. The numbers in the confusion matrix correspond to the percentage of recordings  
 1134 of a given true category which have been predicted to belong to a given predicted category.

1135  
 1136 **Figure 9. Generalization of the classification.**

1137 In this figure, the classifiers were trained with the reliable neurons and tested on the rest of the  
 1138 population. All tested combinations (1-15) based on four types of descriptors (TFRP, Signal, Masker  
 1139 and Signal/Masker) of the categories in stationary and chorus noise and their respective percentages of  
 1140 accuracy of the classifier. The gray part means that the descriptor is included in the classifier and the  
 1141 white part means that the descriptor is excluded from the classifier.

1142  
 1143

1144

1145

1146

**Table titles and legends**

1147

**Table 1.** Summary of the number of animals and number of selected recordings in each structure.

1148

CN: cochlear nucleus, CNIC: central nucleus of inferior colliculus, MGv: ventral part of the medial

1149

geniculate body, A1: primary auditory cortex, VRB: ventrorostral belt.

1150

A recording corresponds to a channel of a 16-channel electrode.

1151

	CN	Lemniscal pathway			Non-lemniscal pathway	Total
		CNIC	MGv	A1	VRB	
Number of animals	10	11	10	11	5	47
Number of recordings tested	672	478	448	544	192	2334
Six Extraction Index (EI) values (for the six SNRs)	617	433	343	488	184	2065
One of the six EI values significantly higher than the EI <sub>Surrogate</sub>	428	374	230	349	137	1518
Selection type						
(a) Significant response to at least one vocalization and/or significant TFRP (Time-Frequency Response Profile)	401	350	210	279	109	1349
(b) Significant response to at least one vocalization and significant TFRP	389	339	198	261	80	1267

1152

1153

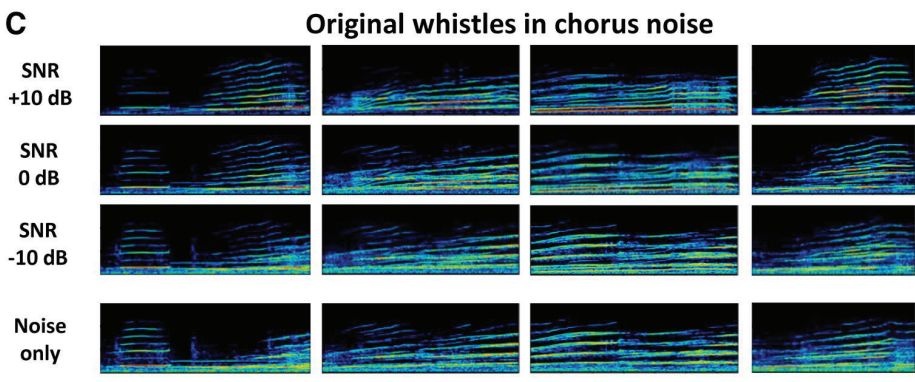
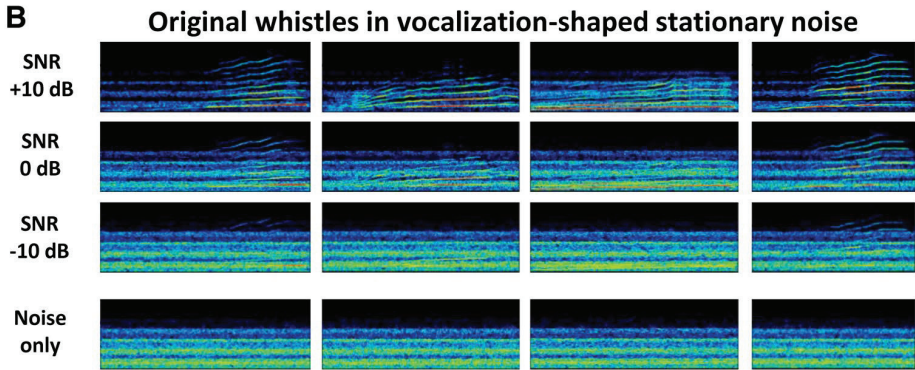
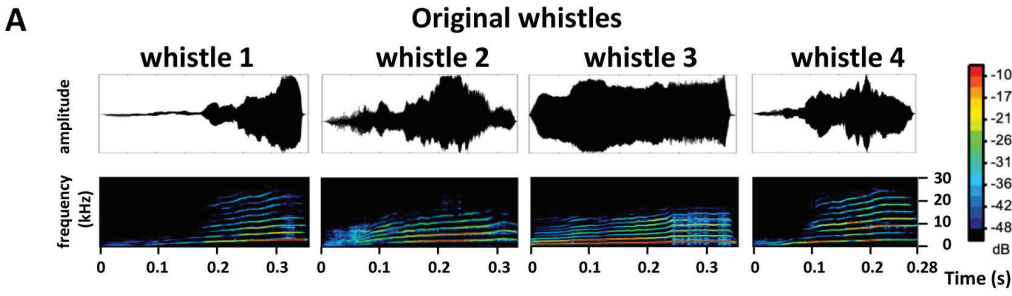
	CN	Lemniscal pathway			Non-lemniscal pathway	Total
		CNIC	MGv	A1	VRB	
Number of recordings reliably categorized in the two noises	139	80	50	52	21	342
Number of recordings reliably categorized and noise-type sensitive	30	35	36	12	10	123
Number of recordings reliably categorized and no noise-type sensitive	109	45	14	40	11	219

1154

1155

**Table 2.** Number of recordings reliably categorized both in stationary and in chorus noise using the bootstrap procedure and number of recordings sensitive to the type of noise within this population.

1156





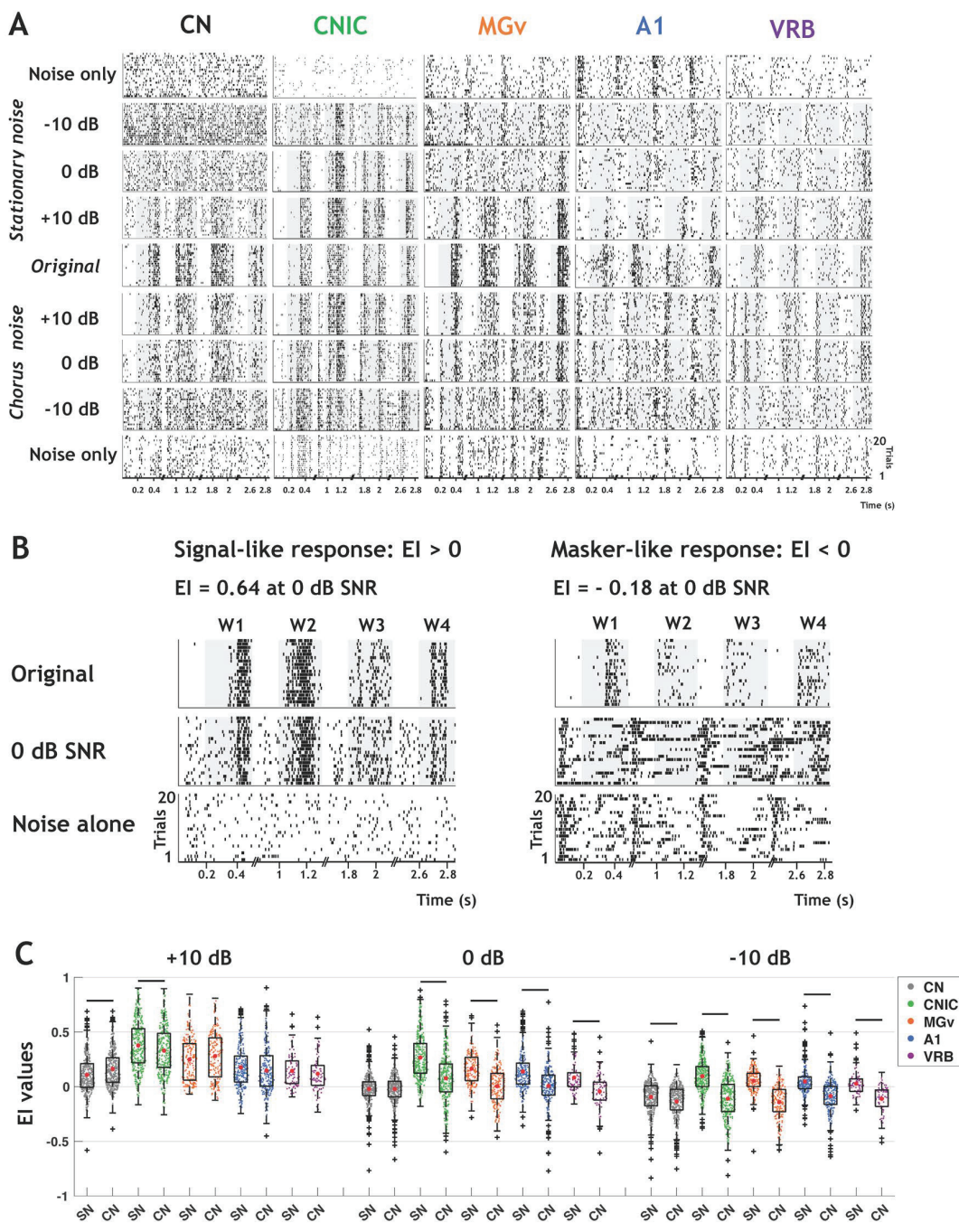


Figure 2.

Determination of the optimal number of clusters

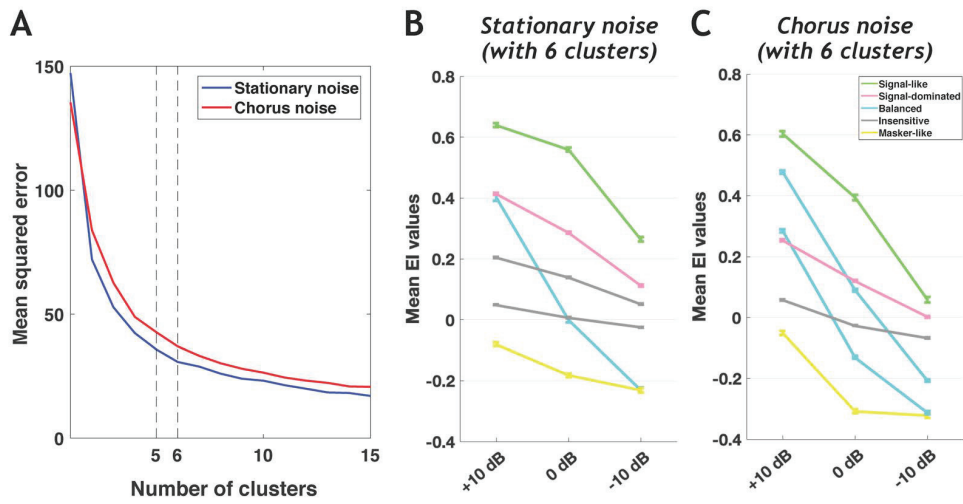
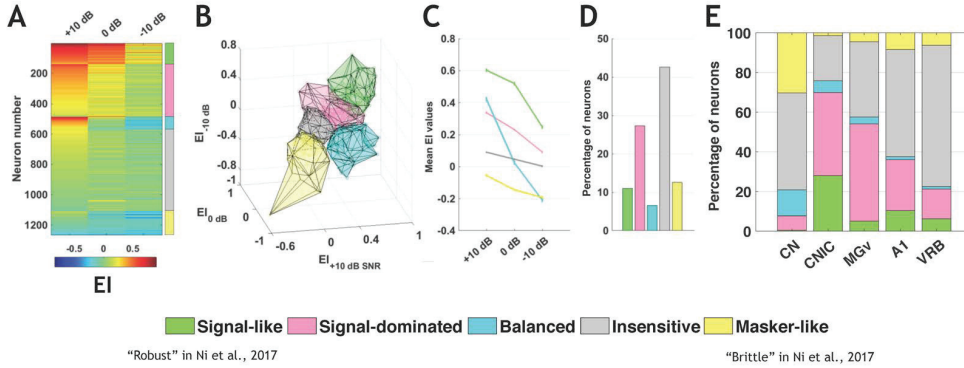


Figure 3.

Stationary noise



Chorus noise

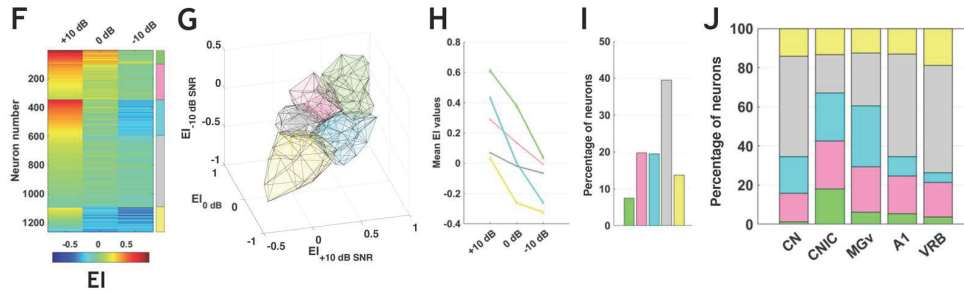
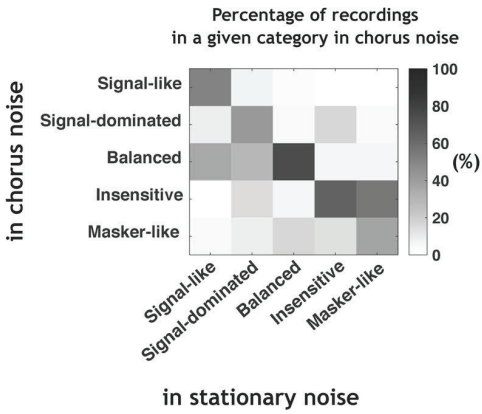


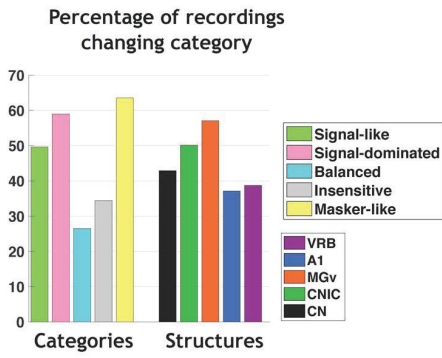
Figure 4.

**A All recordings**

**A1**

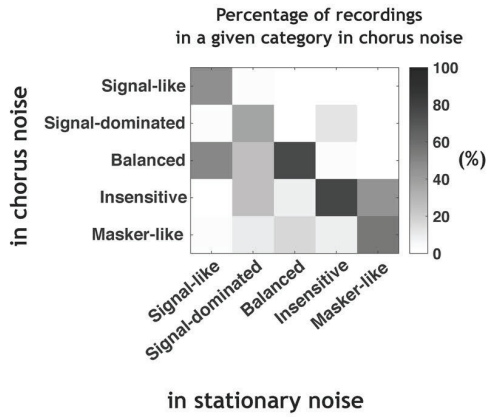


**A2**

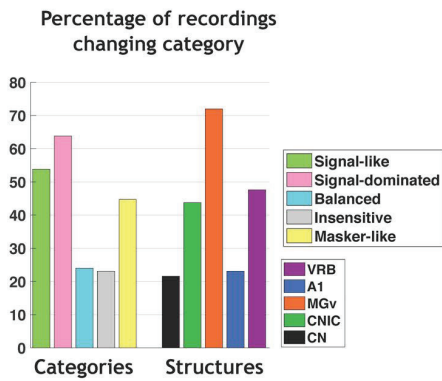


**B Reliable recordings**

**B1**

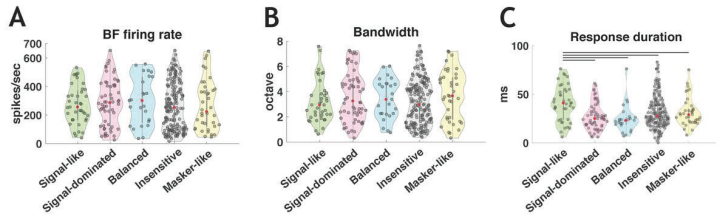


**B2**

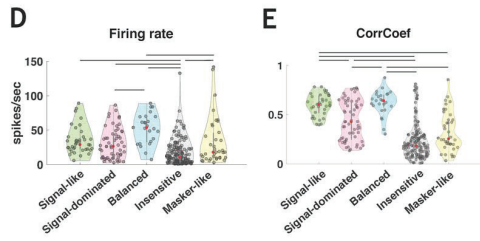


**Figure 5.**

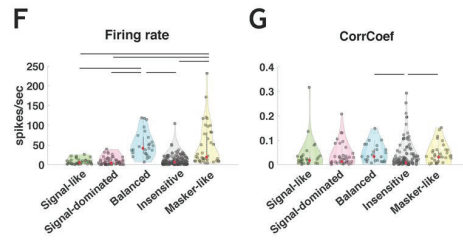
TFRP descriptors



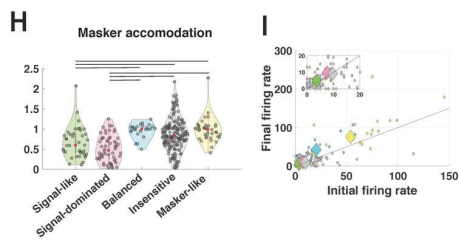
Signal descriptors



Masker descriptors (1)



Masker descriptors (2)



Signal/Masker descriptors

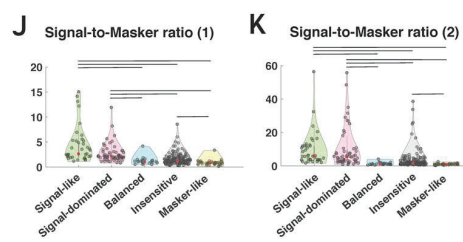
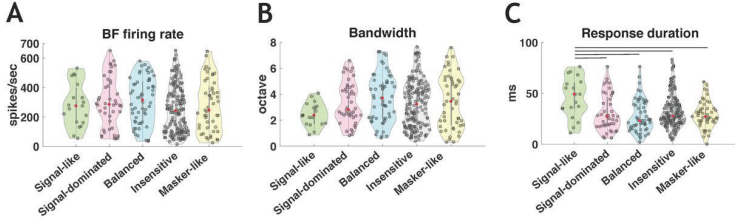
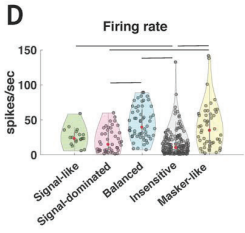


Figure 6.

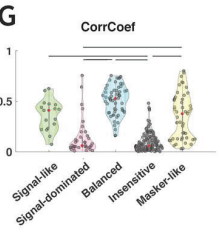
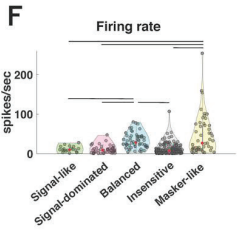
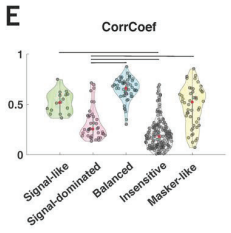
TFRP descriptors



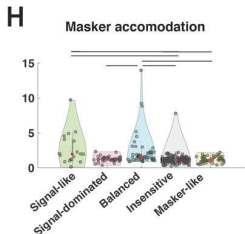
Signal descriptors



Masker descriptors (1)



Masker descriptors (2)



Signal/Masker descriptors

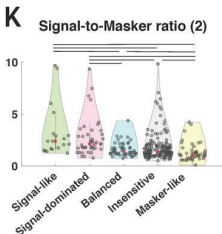
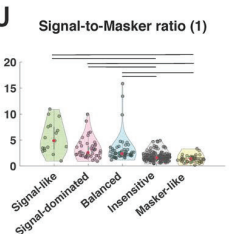
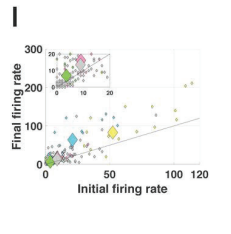


Figure 7.



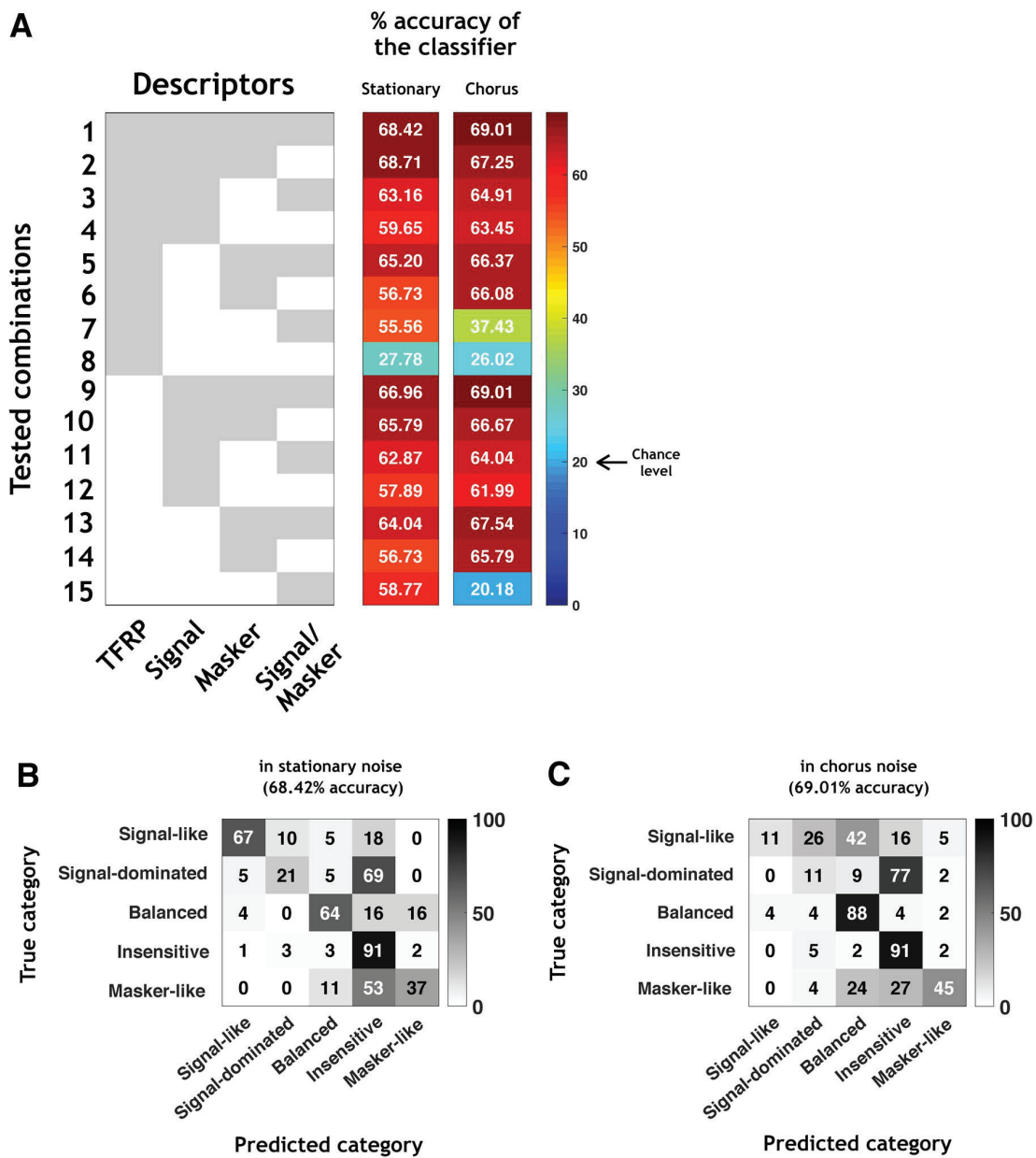


Figure 8.

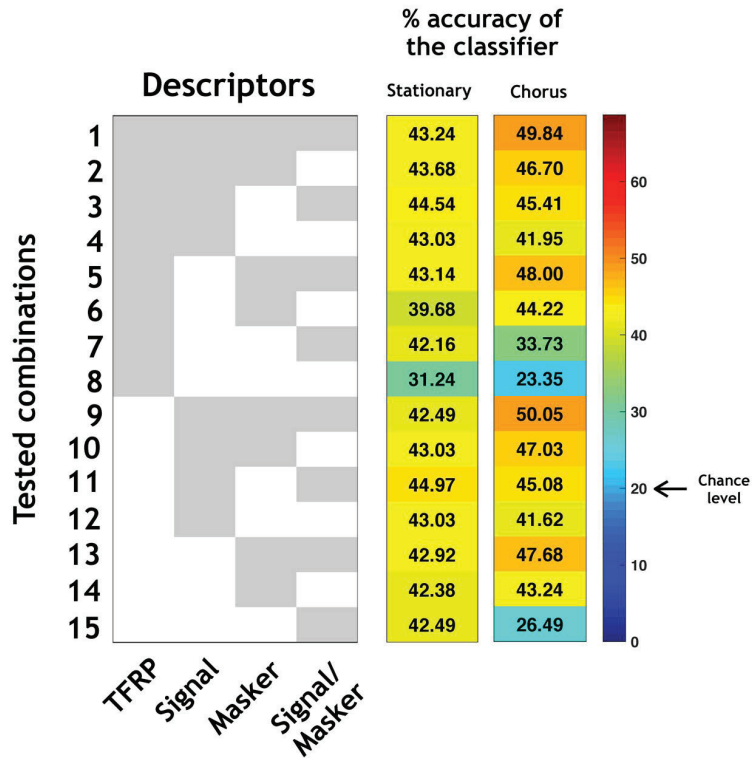


Figure 9.