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

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

eNeuro

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

Robustness to noise in the auditory system: A distributed and predictable property

Souffi S.[1, 2], Lorenzi C. [3], Huetz C. [1,2], Edeline J.-M.* [1,2]

Running title: Noise-resistant neurons in the auditory system

[1] Paris-Saclay Institute of Neuroscience (Neuro-PSI), Department Integrative and Computational Neuroscience, UMR CNRS 9197.
[2] Université Paris-Sud, Bâtiment 446, 91405 Orsay cedex, France.
[3] Laboratoire des systèmes perceptifs, UMR CNRS 8248, Département d'Etudes Cognitives, Ecole Normale Supérieure, Université Paris Sciences & Lettres, Paris, France.

* *Corresponding Author:*
Jean-Marc Edeline
Paris-Saclay Institute of Neuroscience (Neuro-PSI)
UMR CNRS 9197 Université Paris-Sud, Bâtiment 446,
91405 Orsay cedex, France
email: jean-marc.edeline@u-psud.fr

Number of pages: 33
Number of Figures: 9
Number of Extended figures: 4
Number of Tables: 2
Number of words in the abstract: 222
Number of words in the significance statement: 119
Number of words in the introduction: 724
Number of words in the discussion: 2633

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

Acknowledgments

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

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

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-

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

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

Materials and Methods

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

Subjects

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

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

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

Surgical procedures

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

For auditory cortex recordings (areas A1 and VRB), a craniotomy was performed above the left temporal cortex. The dura above the auditory cortex was removed under binocular control and the cerebrospinal fluid was drained through the cisterna to prevent the occurrence of oedema. For the recordings in MGB, a craniotomy was performed above the most posterior part of the MGB (8mm posterior to Bregma) to reach the left auditory thalamus at a location where the MGB is mainly composed of its ventral, tonotopic, part (Redies et al., 1989, Edeline et al., 1999; Anderson et al., 2007; Wallace et al., 2007). For IC recordings, a craniotomy was performed above the left IC and portions of the cortex were aspirated to expose the surface of the left IC (Malmierca et al., 1995, 1996; Rees et al., 1997). For CN recordings, after opening the skull above the left cerebellum, portions of the cerebellum were aspirated to expose the surface of the left CN (Paraouty et al., 2018).

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

Recording procedures

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

Cortical extracellular recordings were obtained from arrays of 16 tungsten electrodes (TDT, TuckerDavis Technologies; ø: 33 µm, <1 MΩ) composed of two rows of 8 electrodes separated by 1000 µm (350 µm between electrodes of the same row). A silver wire, used as ground, was inserted between the temporal bone and the dura matter on the contralateral side. The location of the primary auditory cortex was estimated based on the pattern of vasculature observed in previous studies (Wallace et al., 2000; Gaucher et al., 2013; Gaucher and Edeline, 2015). The non-primary cortical area VRB was located ventral to A1 and distinguished because of its long latencies to pure tones (Rutkowski et al., 2002; Grimsley et al., 2012). For each experiment, the position of the electrode array was set in such a way that the two rows of eight electrodes sample neurons responding from low to high frequency when progressing in the rostro-caudal direction [see examples in Figure 1 of Gaucher et al., (2012) and in Figure 6A of Occelli et al., (2016)].

In the MGB, IC and CN, the recordings were obtained using 16 channel multi-electrode arrays (NeuroNexus) composed of one shank (10 mm) of 16 electrodes spaced by 110 µm and with conductive site areas of 177µm^2. The electrodes were advanced vertically (for MGB and IC) or with a 40° angle (for CN) until evoked responses to pure tones could be detected on at least 10 electrodes.

All thalamic recordings were from the ventral part of MGB (see above surgical procedures) and all displayed latencies < 9ms. At the collicular level, we distinguished the lemniscal and non-lemniscal divisions of IC based on depth and on the latencies of pure tone responses. We excluded the most superficial recordings (until a depth of 1500µm) and those exhibiting latency >= 20ms in an attempt to select recordings from the central nucleus of IC (CNIC). At the level of the cochlear nucleus, the recordings were collected from both the dorsal (DCN) and ventral (VCN) divisions, but based on the recording depth, we estimate that the DCN recordings were more numerous.

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

Acoustic stimuli

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

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

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

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

Experimental protocol

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

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

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

Data analysis

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

Quantification of responses to pure tones

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

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

Quantification of responses evoked by original vocalizations and noises alone

The responses to vocalizations were quantified using two parameters:

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

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

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

where N is the number of trials and $\sigma x_i x_j$ is the normalized covariance at zero lag between spike trains 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

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

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

Quantification of the Extraction Index

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

EI was computed as follows:

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

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

where $D_{n\text{-}snr}$ is the distance between PSTH $_n$ of noise and PSTH $_{snr}$ of vocalization at a particular SNR, whereas $D_{v\text{-}snr}$ is the distance between PSTH $_v$ of pure vocalization and PSTH $_{snr}$ of vocalization at a particular SNR. EI is bounded between -1 and 1: a positive value indicates that the neural response to noisy vocalization is more vocalization-like, and a negative value implies that the neural response is more noise-like. The EI profile for each recording was determined by computing EI at every SNR level. The normalized inner product was used to compute distance between $_n$ or $_v$ and $_{snr}$, as shown in equation above.

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

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

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

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

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

Bootstrap procedure

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

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

Classification using Linear Discriminant Analysis

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

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

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

Descriptors used for the classifier

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

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

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

Global quantification of category changes with mutual information

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

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

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

In a scenario where the categorization based on the responses in stationary noise do not carry information on the categorization based on the responses in chorus noise, assignment of each recording to a category is equivalent to chance level (here 0.20 because there were 5 different categories) and the MI would be close to zero. In the opposite case, when responses in stationary noise always fall in the same category when recorded with chorus noise, the confusion matrix would be diagonal and the mutual information would tend to log2(5) = 2.3 bits.

Statistics

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

Results

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

Chorus noise impacted more neuronal responses than stationary noise at each stage of the auditory system

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

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

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

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

Robustness to noise is a distributed property in the auditory system

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

Figure 4 presents the five categories derived from the whole data set across the three SNRs and the two noises. Figures 4A and 4F present the EI values of all neurons in the three SNRs in a given background noise (with a color code from blue to red when progressing from low to high EI values). The five categories are indicated by a color bar on the right side and are derived from a continuum of EI values. This color code is used for the 3D representations of the five categories in the stationary (Figure 4B) and chorus (Figure 4G) noise. Figure 4C shows the mean EI values in stationary noise for these five categories across the three SNRs and the percentage of neurons in each category is displayed 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

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

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

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

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

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

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

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

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

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

The neuronal behaviors in stationary and chorus noise are predictable based on response parameters obtained in quiet

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

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

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

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

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

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

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

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

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

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

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

cy of the classification was lower in both noises (43.24% and 49.84% in the stationary and chorus noise respectively, combination 1 in Figure 9) certainly because among our recordings, some did not display reliable enough responses. However, these values of accuracy were still more than twice the chance level. Again, the descriptors based on the TFRP provided an accuracy close to the chance level (31.24% and 23.35% in the stationary and chorus noise respectively, line 8 in Figure 9) and the descriptors based on the response to the signal alone (line 12 in Figure 9) and response to the masker alone (line 14 in Figure 9) generated an accuracy that was close from those obtained with the whole set of descriptors (43.03% and 42.38% in stationary noise; 41.62% and 43.24% in chorus noise).

Altogether, our results pointed out that very few neuronal parameters (as the firing rate and the temporal reliability) to the signal alone and to the noise alone are sufficient to predict, up to a maximum of around 70%, the neuronal behaviors in noise.

Discussion

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

Limitations of the study

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

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

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

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

Obviously, it is also important to assess if our results can be generalized to other types of communication 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

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

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

Noise-type sensitivity and noise representation in the auditory system

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

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

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

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

Predictors of neuronal behavior in noise

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

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

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

General conclusion

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

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, Ceballo 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, Quian Quiroga 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.

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.

Mesgarani N & Chang EF (2012) Selective cortical representation of attended speaker in multi-talker speech perception. *Nature* **485**, 233–236.

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.

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.

Narayan R *et al.* (2007) Cortical interference effects in the cocktail party problem. *Nat Neurosci* **10**, 1601–1607.

Ni R, Bender DA, Shanechi AM, Gamble JR & Barbour DL (2017) Contextual effects of noise on vocalization encoding in primary auditory cortex. *Journal of Neurophysiology* **117**, 713–727.

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.

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.

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.

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

Rabinowitz NC, Willmore BDB, King AJ & Schnupp JWH (2013) Constructing Noise-Invariant Representations of Sound in the Auditory Pathway. *PLoS Biol* **11**, e1001710.

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.

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.

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.

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

Schneider DM & Woolley SMN (2013) Sparse and Background-Invariant Coding of Vocalizations in Auditory Scenes. *Neuron* **79**, 141–152.

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.

Shannon, CE (1948) A mathematical theory of communication. Bell Syst Tech J 27(3):379–423.

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. Science. Oct 13;270(5234):303-4.

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.

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.

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.

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.

Town SM, Wood KC, Bizley JK. (2018) Sound identity is represented robustly in auditory cortex during perceptual constancy. Nat Commun. 9(1):4786.

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.

Wallace MN & Palmer AR (2008) Laminar differences in the response properties of cells in the primary auditory cortex. Exp Brain Res 184, 179–191.

Wallace MN, Anderson LA & Palmer AR (2007) Phase-Locked Responses to Pure Tones in the Auditory Thalamus. *Journal of Neurophysiology* **98**, 1941–1952.

Wallace, MN, Rutkowski RG & Palmer AR (2000) Identification and localisation of auditory areas in guinea pig cortex. *Experimental Brain Research* **132**, 445–456.

Zeng FG, Nie K, Stickney GS, Kong YY, Vongphoe M, Bhargave 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.

Figure legends

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

Figure 2. The decrease in EI values is more pronounced in chorus noise than in stationary noise in each auditory structure. A. Raster plots of responses of the four original vocalizations, noisy vocalizations (in both noises) and noise alone recorded in CN, CNIC, MGv, A1 and VRB. The grey part of rasters corresponds to the evoked activity. For each structure, all the rasters correspond to the same recording. **B.** Rasters showing examples of neuronal responses in stationary noise with values of EI > 0 corresponding to a signal-like response (left, IC recording) and EI < 0 corresponding to a masker-like response (right, A1 recording). Top panels show the responses to the original vocalizations, the middle panels the responses to vocalizations at the 0 dB SNR in stationary noise and the bottom panels the responses to stationary noise alone. **C.** Box plots showing the EI values for the three SNRs obtained in CN (in black), CNIC (in green), MGv (in orange), A1 (in blue) and VRB (in purple) alternatively in stationary (SN) and chorus noise (CN). In each box plot, the red dot represents the mean value. The black lines represent significant differences between the mean values (one-way 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, $P^{d}_{0dB, MGv}$ = 2.11e-30, $P^{e}_{0dB, A1}$ = 5.41e-25, $P^{f}_{0dB,VRB}$ = 6.36e-10, $P^{g}_{-10dB, CN}$ = 5.74e-12, $P^{h}_{-10dB, CNIC}$ = 1.83e-59, $P^{i}_{-10dB, MGv}$ = 1.16e-36, $P^{j}_{-10dB, A1}$ = 2.84e-24, $P^{k}_{+10dB, VRB}$ = 4.62e-11).

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 (±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 (±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 (±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 TFRP (Time-Frequency Response Profile, relative to pure tone responses) (**C**).

D. Proportion of each category in the five auditory structures from CN to VRB for the recordings without significant TFRP.
E-H. Same representations as in A, B, C and D for the chorus noise.
See Table 1 (selection type (a)) for referring to the number of selected recordings in each structure.

Figure 4-2. Attempt to separate the ventral and dorsal parts of the cochlear nucleus based on the depth of the recordings.
A-B. Box plots showing the EI values for the three SNRs obtained in the ventral (VCN, in black) and dorsal (DCN, in grey) parts of the cochlear nucleus in stationary (**A**) and chorus noise (**B**). In each box plot, the red dot represents the mean value. There was no significant difference between the EI values obtained in VCN and DCN for the three SNRs and in both noises (one-way ANOVAs, $P>0.05$; with post-hoc paired t tests, $P>0.05$). **C**. Proportion of each category in the ventral part of the cochlear nucleus (VCN, n=87 recordings) and its dorsal part (DCN, n=302 recordings) obtained in stationary (SN) and chorus (CN) noise.

Figure 4-3. Lemniscal and non-lemniscal parts of the inferior colliculus.
Proportion of each category in the lemniscal part of inferior colliculus (central nucleus of inferior colliculus, CNIC, n=339 recordings) and its non-lemniscal parts (dorsal and external cortices of the inferior colliculus, DCIC-ECIC, n=73 recordings) obtained in stationary and chorus noise. See Table 1 (selection type (b)) for referring to the number of selected recordings.

Figure 5. The noise-type sensitivity is found at each stage of the auditory system.
A1. Group-switching matrix representing the percentage of recordings in a given category in chorus noise depending on the category originally assigned in the stationary noise. The abscissas indicate the category identity in the stationary noise and the ordinates represent the category identity in the chorus noise.
For example, signal-like responses in stationary noise are also 50% signal-like in chorus noise but 10% are reclassified as signal-dominated, 35% balanced, 1.5% insensitive and 3.5% masker-like. Note that, in stationary noise, the number of recordings in each category were 139, 346, 83, 540 and 159 in signal-like, signal-dominated, balanced, insensitive and masker-like category respectively.
A2. Mean percentages of recordings changing category from the stationary noise to the chorus noise, first in each category and second in each structure (VRB, (in purple), A1 (in blue), MGv (in orange), CNIC (in green) and CN (in black)).
B1. Group-switching matrix representing the percentages of recordings changing category from the stationary noise to the chorus noise based only on recordings considered as reliable with the bootstrap procedure in the two types of noise (with a confidence interval $\geqslant 95\%$).
B2. Mean percentages of recordings changing category from the stationary noise to the chorus noise, first in each category and second in each structure (VRB, (in purple), A1 (in blue), MGv (in orange), CNIC (in green) and CN (in black)).

Figure 5-1. Confusion matrices obtained for all and reliable recordings.
A. Confusion matrix relative to Figure 5A1 representing the number of recordings in a given category in chorus noise depending on the category originally assigned in the stationary noise.
B. Same as in (A) with only recordings considered as reliable with the bootstrap procedure in the two types of noise (with a confidence interval $\geqslant 95\%$). This figure is relative to the Figure 5B1.
Note that the mutual information (MI, bits) values were low in (A) and (B) corroborating the fact that a large proportion of recordings assigned to a given category in stationary noise change category in chorus noise.

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

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

Figure 7. Descriptors of the categories in chorus noise used by the classifiers.
A-C. Three TFRP parameters were chosen as descriptors: the best frequency (BF) firing rate, the bandwidth and the response duration. **D-E.** Two signal descriptors were selected corresponding to the firing rate and the CorrCoef values obtained in original conditions. **F-G.** Two main masker descriptors were presented corresponding to the firing rate and the CorrCoef values obtained in chorus noise alone. **H-I.** The three other masker descriptors are: (**H**) the ratio between the masker firing rate taken at the time the signal should have occurred and the initial masker firing rate during the first 200 ms of the masker and the number of action potentials emitted during the first (Initial, **I**) and last (Final, **I**) 50 ms of the masker alone over a 564 ms-period.
J-K. Two descriptors of the signal-to-masker ratio are presented and taken into account the firing rate of responses to the signal and to the masker; the two differ only on which part of the response to the masker is taken into account (see Methods). For each violin plot, the red dot represents the median value and the black lines represent significant differences between the median values (Kruskal-Wallis test, $p<0.05$).

Figure 8. The neuronal behaviors in stationary and chorus noise are predictable based on response parameters obtained in quiet.
A. All tested combinations (1-15) based on four types of descriptors (TFRP, Signal, Masker and Signal/Masker) of the categories in stationary noise and their respective percentages of accuracy of the classifier. The gray part means that the descriptor is included in the classifier and the white part means that the descriptor is excluded from the classifier.
B-C. Example of the confusion matrix obtained with all descriptors (combination 1) in stationary (**B**) and chorus (**C**) noise. Each row corresponds to a true category and each column corresponds to a predicted category. The numbers in the confusion matrix correspond to the percentage of recordings of a given true category which have been predicted to belong to a given predicted category.

Figure 9. Generalization of the classification.
In this figure, the classifiers were trained with the reliable neurons and tested on the rest of the population. All tested combinations (1-15) based on four types of descriptors (TFRP, Signal, Masker and Signal/Masker) of the categories in stationary and chorus noise and their respective percentages of accuracy of the classifier. The gray part means that the descriptor is included in the classifier and the white part means that the descriptor is excluded from the classifier.

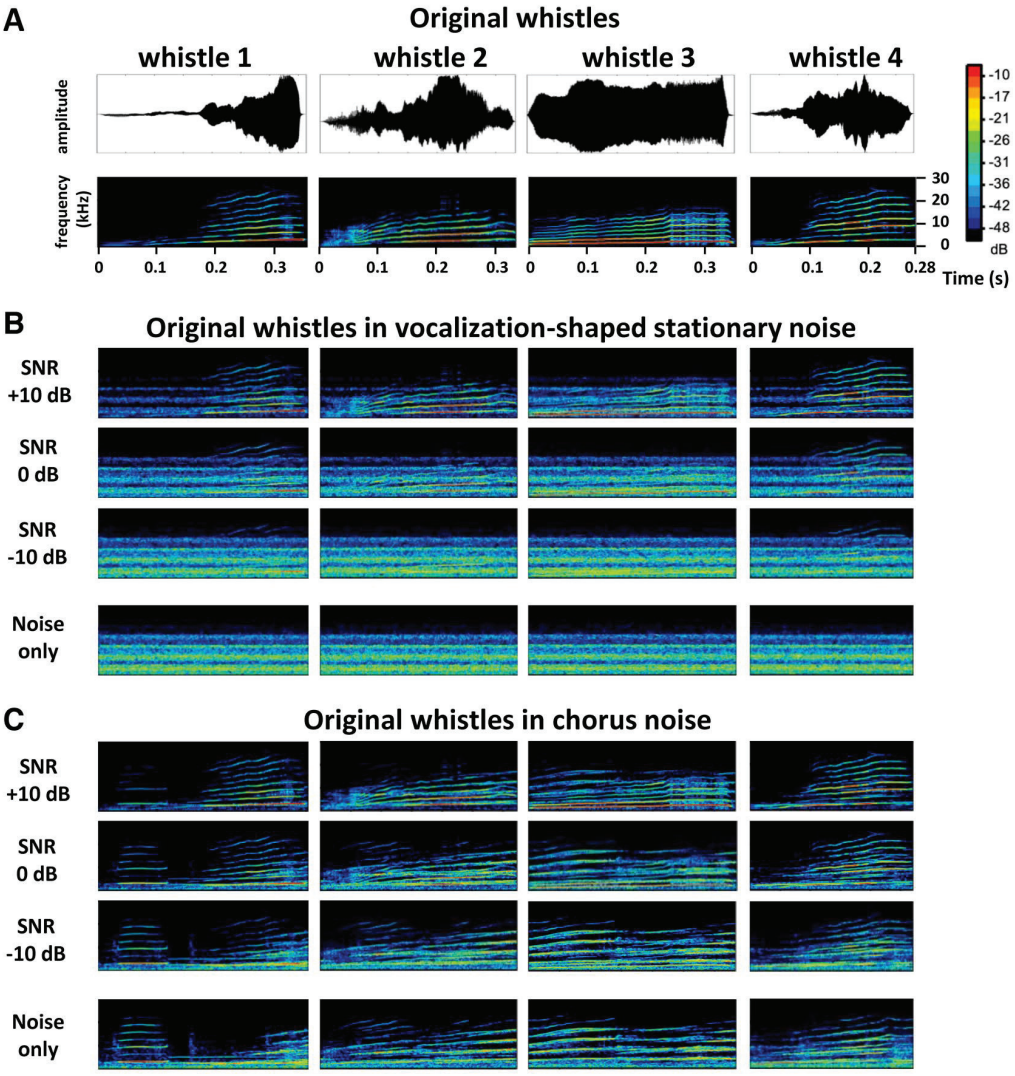
Table titles and legends

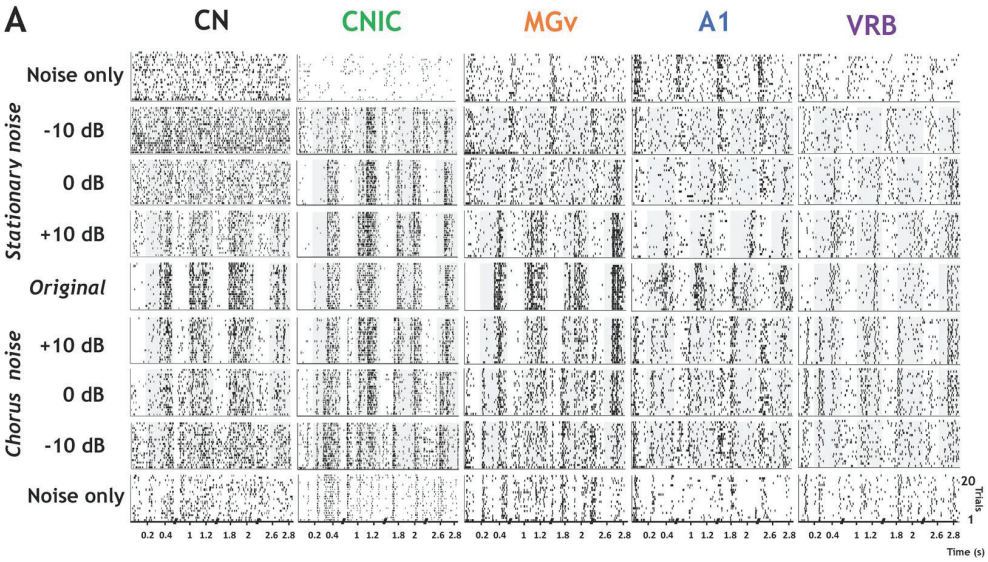
Table 1. Summary of the number of animals and number of selected recordings in each structure. CN: cochlear nucleus, CNIC: central nucleus of inferior colliculus, MGv: ventral part of the medial geniculate body, A1: primary auditory cortex, VRB: ventrorostral belt. A recording corresponds to a channel of a 16-channel electrode.

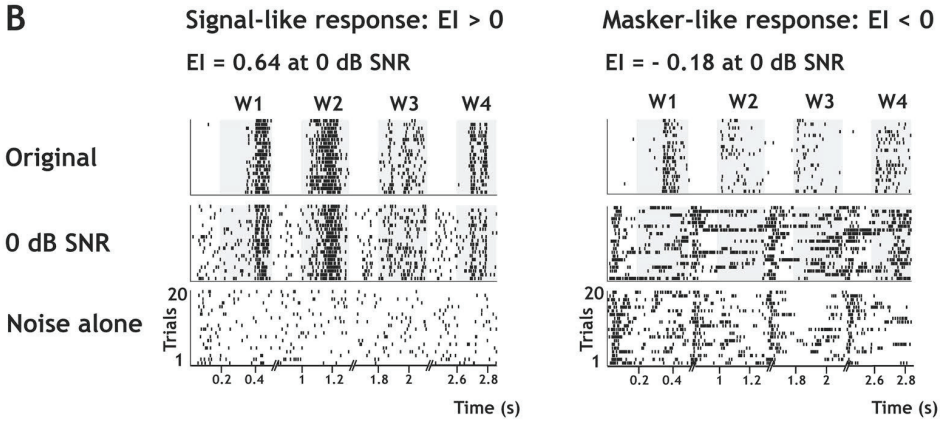
	CN	Lemniscal pathway			Non-lemniscal pathway	Total
		CNIC	MGv	A1	VRB	
Number of animals	10	11	10	11	5	47
Number of recordings tested	672	478	448	544	192	2334
Six Extraction Index (EI) values (for the six SNRs)	617	433	343	488	184	2065
One of the six EI values significantly higher than the $EI_{Surrogate}$	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

	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

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.







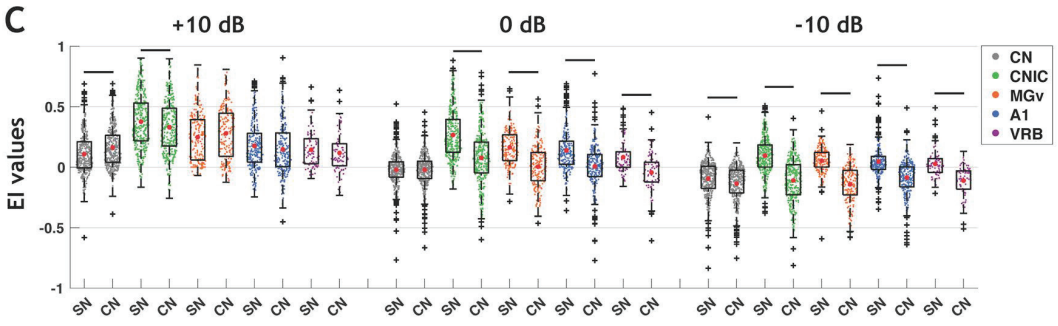


Figure 2.

Determination of the optimal number of clusters

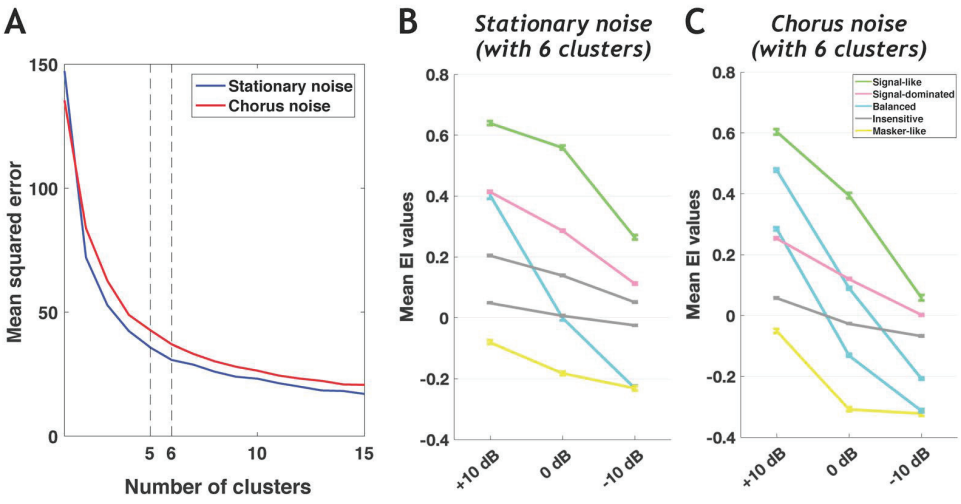
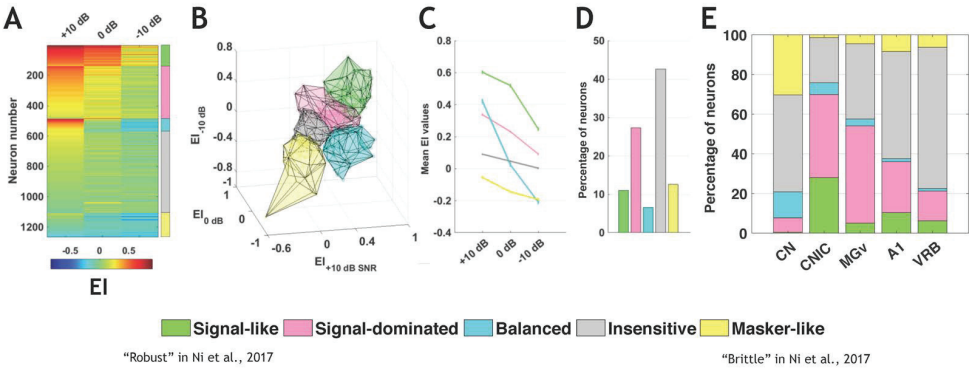


Figure 3.

Stationary noise



Chorus noise

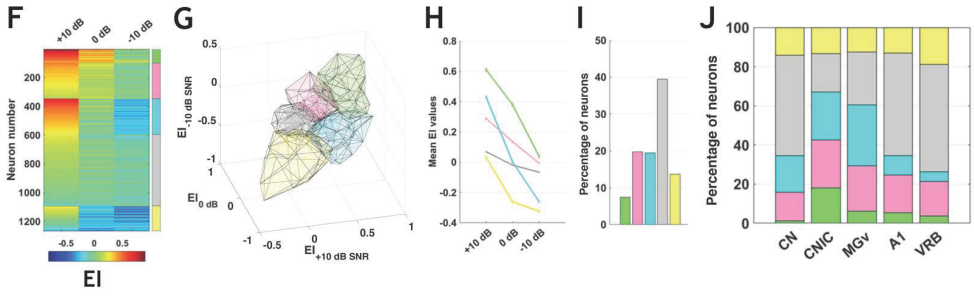


Figure 4.

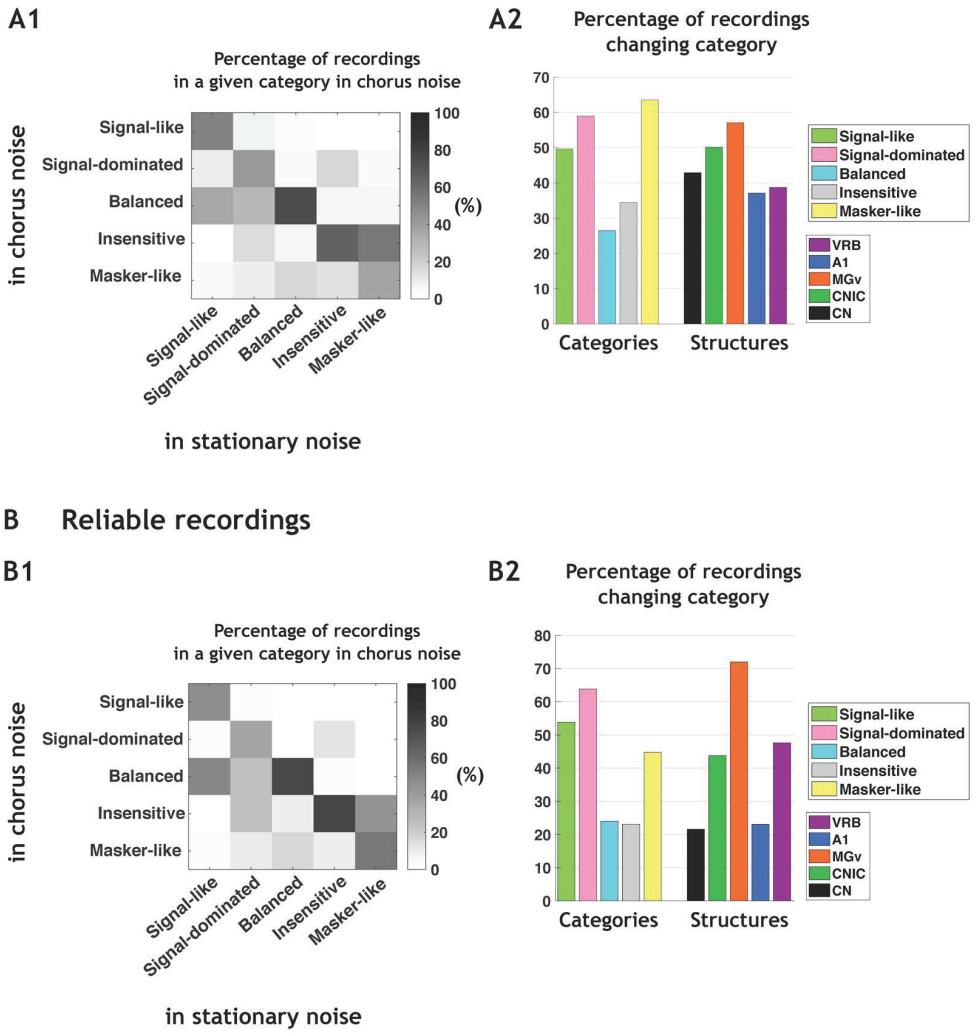


Figure 5.

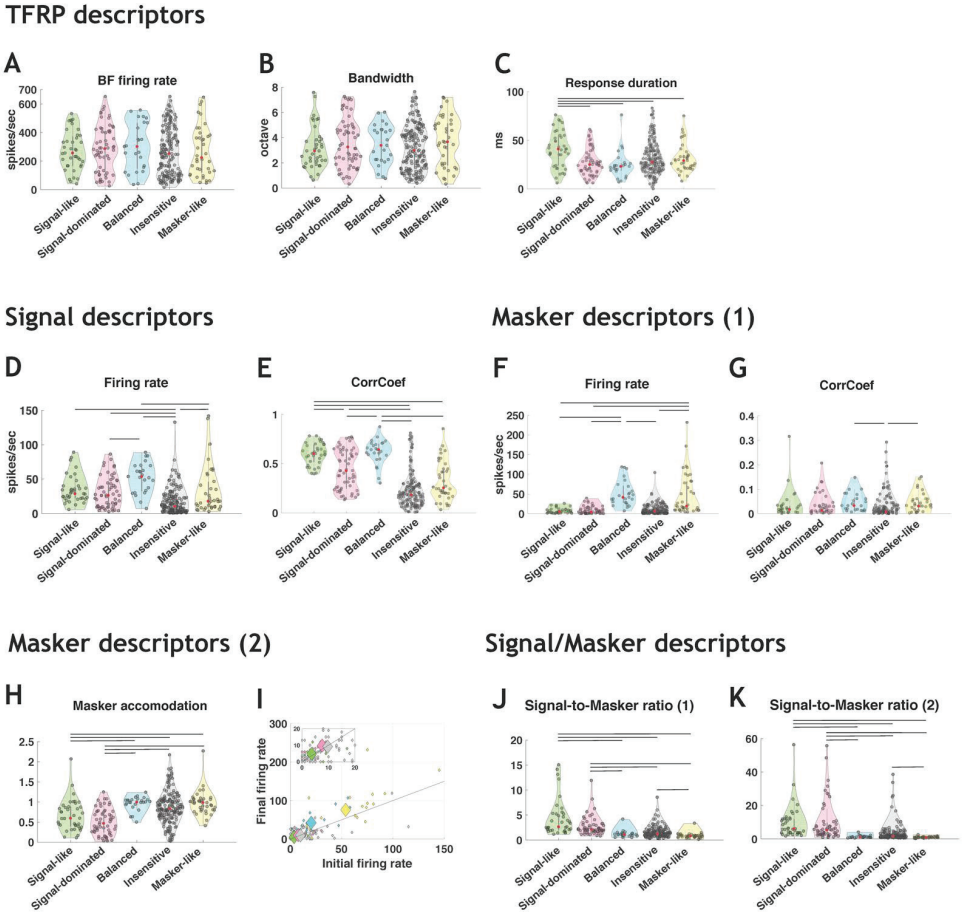


Figure 6.

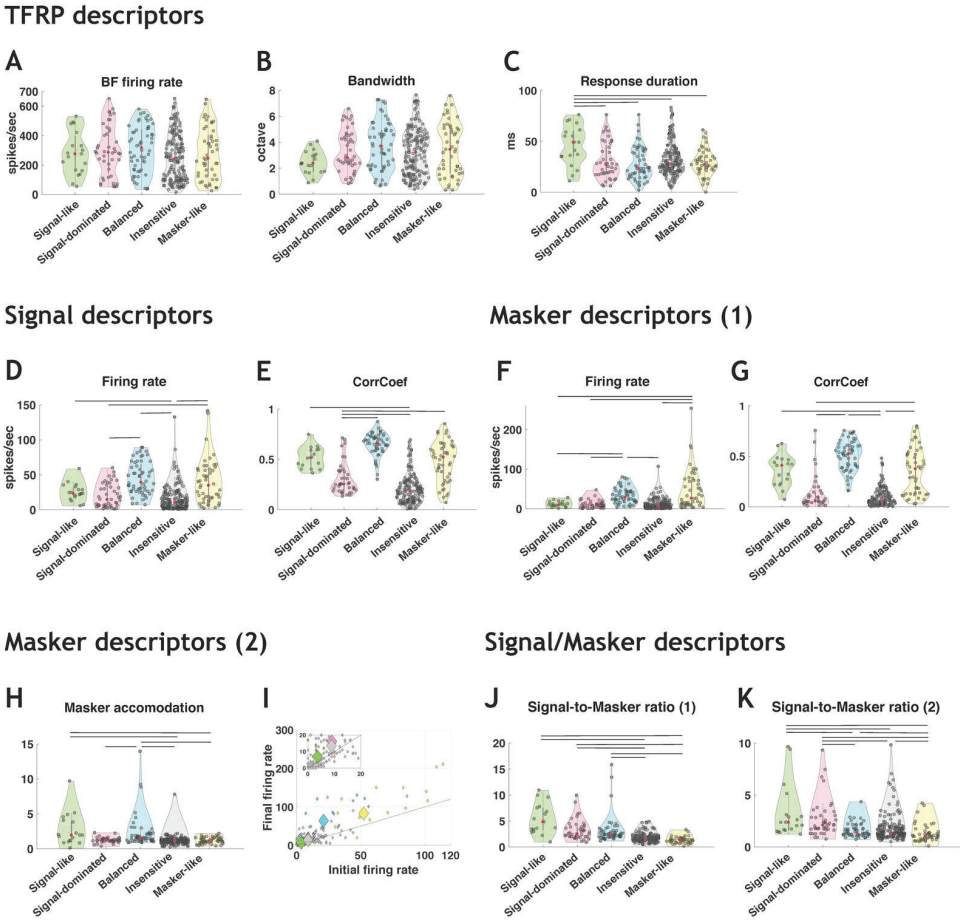


Figure 7.

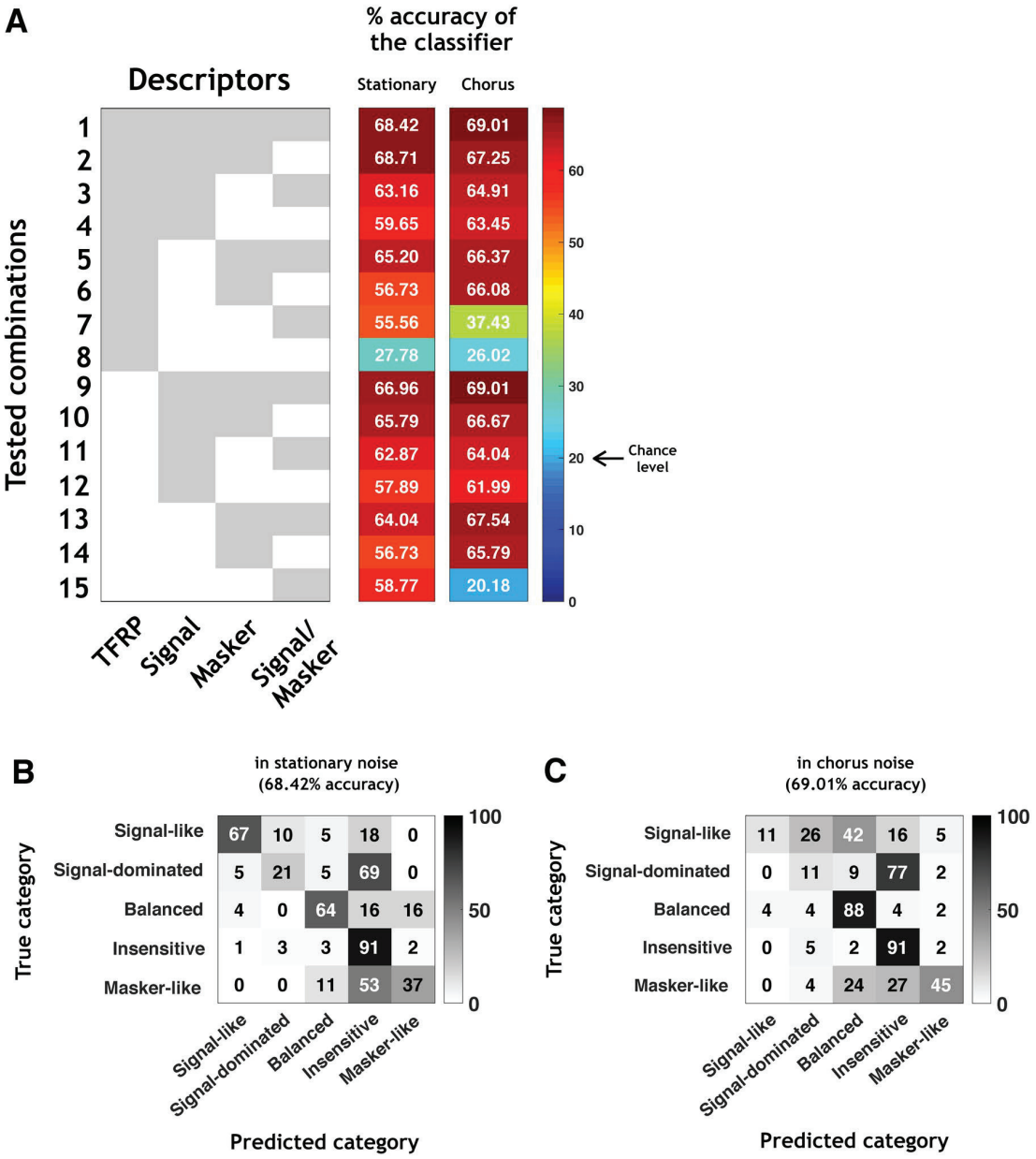


Figure 8.

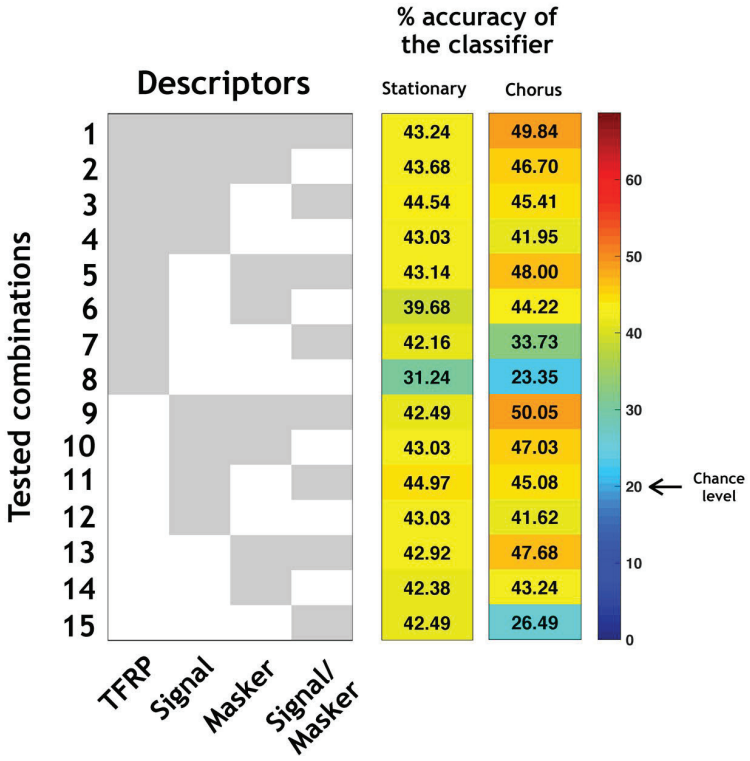


Figure 9.