Noise-Sensitive But More Precise Subcortical Representations Coexist with Robust Cortical Encoding of Natural Vocalizations

Samira Souffi,1,2 Christian Lorenzi,3 Léo Varnet,3 Chloé Huetz,1,2 and Jean-Marc Edeline1,2

1Department Cognition and Behavior, CNRS, UMR 9197, Paris-Saclay Institute of Neurosciences, 91190 Saint-Aubin, France, 2Université Paris-Sud, 91405 Orsay Cedex, France, and 3Laboratoire des systèmes perceptifs, UMR, CNRS 8248, Département d’Etudes Cognitives, Ecole Normale Supérieure, Université PSL, F-75230 Paris Cedex 05, France

Humans and animals maintain accurate sound discrimination in the presence of loud sources of background noise. It is commonly assumed that this ability relies on the robustness of auditory cortex responses. However, only a few attempts have been made to characterize neural discrimination of communication sounds masked by noise at each stage of the auditory system and to quantify the noise effects on the neuronal discrimination in terms of alterations in amplitude modulations. Here, we measured neural discrimination between communication sounds masked by a vocalization-shaped stationary noise from multunit responses recorded in the cochlear nucleus, inferior colliculus, auditory thalamus, and primary and secondary auditory cortex at several signal-to-noise ratios (SNRs) in anesthetized male or female guinea pigs. Masking noise decreased sound discrimination of neuronal populations in each auditory structure, but collicular and thalamic populations showed better performance than cortical populations at each SNR. In contrast, in each auditory structure, discrimination by neuronal populations was slightly decreased when tone-vocoded vocalizations were tested. These results shed new light on the specific contributions of subcortical structures to robust sound encoding, and suggest that the distortion of slow amplitude modulation cues conveyed by communication sounds is one of the factors constraining the neuronal discrimination in subcortical and cortical levels.

Key words: amplitude modulation; auditory system; masking noise; natural sounds; neural discrimination; population recordings

Significance Statement
Dissecting how auditory neurons discriminate communication sounds in noise is a major goal in auditory neuroscience. Robust sound coding in noise is often viewed as a specific property of cortical networks, although this remains to be demonstrated. Here, we tested the discrimination performance of neuronal populations at five levels of the auditory system in response to conspecific vocalizations masked by noise. In each acoustic condition, subcortical neurons better discriminated target vocalizations than cortical ones and in each structure, the reduction in discrimination performance was related to the reduction in slow amplitude modulation cues.

Introduction
Understanding the neural mechanisms used by the auditory system to extract and represent relevant information for discriminating communication sounds in a variety of acoustic environments is a major goal in auditory neurosciences.

Several studies have prompted the view that the perceptual robustness mainly relies on the capacity of cortical neurons to extract invariant acoustic features (Narayan et al., 2007; Schneider and Woolley, 2013; Carruthers et al., 2015; Ni et al., 2017; Town et al., 2018), and it was proposed that this capacity is due to a larger adaptation of cortical cells to the noise statistics compared with subcortical cells (Rabinowitz et al., 2013). Indeed, in the cortical field L—analogous to primary auditory cortex
(A1) in birds—the percentage of correct neuronal discrimination between zebra-finch songs embedded in different types of acoustic maskers decreases proportionally to the target-to-masker ratio and parallels behavioral performance (Narayan et al., 2007). Also, consistent with behavioral data (for review, see Verhey et al., 2003), the comodulation of different frequency bands in background noise improved tone detection in noise of auditory cortical, thalamic, and collicular neurons (Nelken et al., 1999; Las et al., 2005). Moreover, between-vowels discrimination performance of neuronal populations located in A1 resists to a large range of acoustic alterations (including changes in fundamental frequency, spatial location, or level) and is similar to behavioral performance (Town et al., 2018).

The goal of the present study was to challenge this view by identifying the auditory structures responsible for this robust neural discrimination. Background noise has the following three disruptive effects on communication sounds (Noordhoek and Druilman, 1997; Dubbelboer and Houtgast, 2007): it attenuates the power of their amplitude modulation (AM) components (also called “temporal envelope”; Houtgast and Steeneken, 1985; Ewert and Dau, 2000; Biberger and Ewert, 2017); it corrupts their frequency modulation (FM) components (also called “temporal fine structure” (TFS); Shamma and Lorenzi, 2013; Varnet et al., 2017); and it introduces stochastic fluctuations in AM power that generate temporal irregularities (from bin to bin) in the signal temporal envelopes (Ewert and Dau, 2000). Here, electrophysiological recordings were collected from the cochlear nucleus up to a secondary auditory cortical area in anesthetized guinea pigs and the discrimination performance of neuronal populations was assessed for four utterances of the same vocalization category (the whistle; e.g., the guinea pig alarm call) presented against a vocalization-shaping stationary noise at three signal-to-noise ratios (SNRs; +10, 0, −10 dB). An increased discrimination performance may result from the specialization of cortical responses for detecting crucial vocalization features (Wang et al., 1995; Wang and Kadia, 2001; Schneider and Woolley, 2013), whereas a decreased discrimination performance may result from the loss of spectrotemporal details promoting the categorization of sounds into auditory objects (Nelken and Bar-Yosef, 2008; Chechik and Nelken, 2012). Mutual information was used to determine whether the temporal patterns of neuronal responses to the four vocalizations sufficiently differed to assign each response to a particular vocalization. The results obtained in noise were compared with the effects of a deterministic signal-processing scheme, namely, a tone vocoder, which markedly altered the FM cues and progressively attenuated the AM cues (within 38–10 frequency bands). The AM spectra were computed at the output of simulated guinea pig auditory filters for each acoustic alteration. Our results suggest that the attenuation of slow AM cues is one of the factors explaining the decrease in discrimination performance in cortical and subcortical structures. In addition, this study revealed that, for each acoustic distortion tested, the highest level of discrimination was found in subcortical structures, either at the collicular level (in masking-noise conditions) or at the thalamic level (in vocoder conditions).

Materials and Methods

Subjects

These experiments were performed under national license A-91-557 (project 2014-25, authorization 05.202.02) and using procedures 32-2011 and 34-2012, which were validated by Ethics Committee no. 59 [CMEA (Comité d’Ethique en Expérimentation Animale) Paris Center et Sud]. All surgical procedures were performed in accordance with the guidelines established by European Communities Council Directive (2010/63/EU Council Directive Decree).

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

Two to three days before each experiment, the pure-tone audiogram of the animal was determined by testing auditory brainstem responses (ABRs) under isoflurane anesthesia (2.5%), as described in the study by Gourévitch et al. (2009). The ABR was obtained by differential recordings between two subdermal electrodes (SC25, NeuroService) placed at the vertex and behind the mastoid bone. Software (RTLab, Echodia) allowed the averaging of 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 right ear of the animal. 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 down to −10 dB SPL). All animals used in this study had normal pure-tone audiograms (Gourévitch et al., 2009; Gourévitch and Edeline, 2011; Aushana et al., 2018).

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 hindpaw (usually two to four times during the recording session). A single dose of atropine sulfate (0.06 mg/kg, s.c.) was given to reduce bronchial secretions, and a small dose of buprenorphine was administered (0.05 mg/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 into the wound.

For auditory cortex recordings (areas A1 and VRB), a craniotomy was performed above the left temporal cortex. The opening was 8 mm wide starting at the intersection point between parietal and temporal bones and 8–10 mm in height. The dura above the auditory cortex was removed under binocular control and the CSF was drained through the cisterna to prevent the occurrence of edema.

For the recordings in MGB, a craniotomy was performed above the most posterior part of the MGB (8 mm posterior to bregma) to reach the left auditory thalamus at a location where the MGB is mainly composed of its ventral, tonotopic, division (Redies et al., 1989; Edeline et al., 1999; Redies et al., 2007; Anderson et al., 2007; Wallace et al., 2007).

For IC recordings, a craniotomy was performed above the IC, and portions of the cortex were aspirated to expose the surface of the left IC. For CN recordings, after opening the skull above the right cerebellum, portions of the cerebellum were aspirated to expose the surface of the right 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 multunit recordings were collected in five auditory structures, the non-primary cortical area VRB, the primary cortical area A1, the MGB, the IC, and the 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 (ø: 33 μm, <1 MΩ) composed of two rows of eight electrodes separated by 1000 μm (350 μm between electrodes of the same row). A silver wire,
used as a ground, was inserted between the temporal bone and the dura mater on the contralateral side. The location of the primary auditory cortex was estimated based on the pattern of vasculature observed in previous studies (Edeline and Weinberger, 1993; Manunta and Edeline, 1999; Wallace et al., 2000; Edeline et al., 2001). 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 rostrocaudal direction [see examples of tonotopic gradients recorded with such arrays in Gaucher et al., 2012 (their Fig. 1) and in Occelli et al., 2016 (their Fig. 6A)].

All the remaining extracellular recordings (in MGB, IC, and CN) were obtained using 16-channel multielectrode arrays (NeuroNexus) composed of one shank (10 mm) of 16 electrodes spaced by 110 μm and with conductive site areas of 177 μm². 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 were <9 ms. At the collicular level, we distinguished the lemniscal and nonlemniscal divisions of IC based on depth and on the latencies of pure-tone responses. We excluded the most superficial recordings (until a depth of 1500 μm) and those exhibiting latency ≥20 ms 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 and ventral divisions.

The raw signal was amplified 10,000 times [Medusa, Tucker-Davis Technologies (TDT)]. It was then processed by an RX5 multichannel data acquisition system (TDT). The signal collected from each electrode was filtered (610–10,000 Hz) to extract multunit activity (MUA). The trigger level was set for each electrode to select the largest action potentials from the signal. Online and offline examination of the waveforms suggests that the MUA collected here was made of action potentials generated by a few neurons in the vicinity of the electrode. However, as we did not use 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 (0.5–1 MΩ) and that the spike amplitudes had similar values (100–300 μV) 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.

**Acoustic stimuli**

Acoustic stimuli were generated using MATLAB (MathWorks), transferred to a RP2.1-based sound delivery system (TDT) and sent to a Forstek Speaker (model FE87E). The speaker was placed at 2 cm from the right ear of the guinea pig, a distance at which the speaker produced a flat spectrum (±3 dB) between 140 Hz and 36 kHz. The stimulation was not pure monaural, but the head and body of the animal largely attenuated binaural cues. Calibration of the speaker was made using noise and pure tones recorded by a Brüel & Kjær microphone (model 4133) coupled to a preamplifier (model 2169, B&K) and a digital recorder (model PMD671, Marantz). A large set of whistle calls was loaded in the Audition software (Audition 3, Adobe) and four representative examples of whistle were selected. As shown in Figure 1A, despite the fact that the maximal energy of the four selected whistles was in the same frequency range (typically, between 4 and 26 kHz), these calls displayed slight differences in their spectrogram (Fig. 1A–B). In addition, their global temporal envelopes clearly differed (Fig. 1C). The four selected whistles were processed by three- tone vocoders (Gnansia et al., 2009, 2010). In the following figures, the unprocessed whistles will be referred to as the original versions, and the vocoded versions as Voc38 (Voc20, and Voc10, respectively) for the 38-band (20-, and 10-band, respectively) vocoder. In contrast to previous studies that used noise-excited vocoders (Nagarajan et al., 2002; Ransingh et al., 2012; Ter-Mikaelian et al., 2013), a tone vocoder was used here, because noise vocoders introduce random (i.e., non-informative) intrinsic temporal envelope fluctuations distorting the crucial spectrotemporal modulation features of communication sounds (Kates, 2011; Stone et al., 2011; Shamma and Lorenzi, 2013).

**Figure 1D** displays the spectrograms of the 38-band vocoded (first column), the 20-band vocoded (second column) and the 10-band vocoded (third column) versions of the four whistles. The three vocoders differed only in terms of the number of frequency bands (i.e., analysis filters) used to decompose the whistles (38, 20, or 10 bands). The 38-band vocoding process is briefly described below, but the same principles apply to the 20- and the 10-band vocoders. Each digitized signal was processed through a bank of fifth-order Gammatone filters (Patterson, Sayles and Winter, 2010). For the AM spectrum, we analyzed the temporal envelope in each frequency band of 50 gammatone filters than for the vocoding (range, 0.1–10,000 Hz) to extract multiunit activity (MUA). The weighting compensated for imperfect superposition of the impulse responses of the bands at the desired group delay. The weights were optimized numerically to achieve a flat frequency response. **Figure 1E** shows the long-term power spectrum of the 38-, 20-, and 10-band vocoded whistles, and **Figure 1F** shows their global temporal envelopes (which were relatively well preserved by the vocoding process).

The four whistles were also presented in a frozen noise ranging from 10 to 24,000 Hz. To generate this noise, recordings were performed in the colony room where a large group of guinea pigs were housed (30–40, 2–4 animals/cage). Several 4 s audio recordings were added up to generate a “chorus noise,” the power spectrum for which was computed using the Fourier transform. This spectrum was then used to shape the spectrum of white Gaussian noise. The resulting vocalization-shaped stationary noise therefore matched the chorus noise audio spectrum, which explains why frequency bands were overrepresented in the vocalization-shaped stationary noise. **Figure 1G** displays the spectrograms of the four whistles in the vocalization-shaped stationary noise with SNRs of +10, 0, and −10 dB SPL. **Figure 1H** shows the long-term power spectrum of the four whistles at the +10, 0, and −10 dB SNRs, and **Figure 1J** shows their global temporal envelopes (which were severely altered at the 0 and −10 dB SNRs).

AM spectra were computed for the original, vocoded, and noisy versions of each vocalization by decomposing each sound with the same bank of 50 gammatone filters than for the vocoding (range, 0.1–50 kHz). The AM component (envelope) thus corresponds to the magnitude of the analytic signal, whereas the TFS corresponds to its unwrapped instantaneous phase.

For the AM spectrum, we analyzed the temporal envelope in each frequency band through a bank of AM filters using a method adapted from the human study by Varten et al. (2017) for the hearing range of guinea pigs (one-third octave wide first-order Butterworth bandpass...
filters overlapping at $-3\,\text{dB}$, with center frequencies between 0.1 and 410 Hz). The root mean square amplitude of the filtered output was multiplied by a factor of $\sqrt{2}$. For each AM filter, and a modulation index was calculated by dividing the output by the mean amplitude of the AM component for the vocalization sample in the corresponding gammatone filter.

**Experimental protocol**

As inserting an array of 16 electrodes in a brain structure almost systematically induces a deformation of this structure, a 30 min 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 TFRPs were used...
to assess the quality of our recordings and to adjust the depth of electrodes. 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 ~7 mm 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, as follows: we required that the recorded neurons displayed at least three successive similar TFRPs (each lasting 6 min) 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 original whistles in their natural versions, followed by the vocoded versions with 38, 20, and 10 bands at 75 dB SPL. The same set of original whistles was then presented in the vocalization-shaped stationary noise presented 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 min. The protocol was restarted 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

Quantification of responses to pure tones. The TFRPs were obtained by constructing post-stimulus time histograms (PSTHs) 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, TFRPs are matrices of 100 bins on the abscissa (time) multiplied by 129 bins on the ordinate (frequency). All TFRPs were smoothed with a uniform 5 × 5 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 excitatory response were automatically identified using the following procedure: an excitatory peak in the TFRP was defined as a contour of firing rate above spontaneous activity plus six times the SD of the spontaneous activity. Recordings without a significant excitatory peak of responses or with only inhibitory responses were excluded from the data analyses. The bandwidth (BW) was defined as the sum of all peak widths in octaves. The response duration was the time difference between the first and last spikes of the significant peaks. The response strength was the total number of spikes falling in the significant peaks (in Action Potentials/s, AP/s).

Quantification of responses evoked by vocalizations. The responses to vocalizations were quantified using the following two parameters: (1) the firing rate of the evoked response, which corresponds to the total number of action potentials occurring during the presentation of the stimulus minus spontaneous activity (Shannon et al., 1948) calculated as the firing rate of the evoked response, which corresponds to the total number of action potentials occurring during the presentation of the stimulus minus spontaneous activity (Shannon et al., 1948) calculated as the firing rate of the evoked response of MIbias values was very similar between auditory structures: depending on the size range, the MIbias values varied between 0.102 and 0.107 in the CN, from 0.107 to 0.110 in the IC, from 0.105 to 0.114 in the MGB, from 0.107 to 0.111 in A1, and from 0.106 to 0.116 in VRB. There was no significant difference between the mean MIbias values in the different structures (unpaired t test, all p > 0.25).

The information carried by a group of recordings was estimated by the MI_population using the same method described above, as follows: responses of several simultaneous multiunit recordings were concatenated and considered as a single pattern. To assess the influence of the group size of simultaneous multiunit recordings on the information...
carried by that group (MIPopulation), the number of sites used for computing MIPopulation varied from two to the maximal possible size (which is equal to 1 

\[ C_{n,k} \] 

where \( k \) is the group size and \( n \) the number of responsive sites in a recording session), a threshold was fixed to save computation time, as follows: when the number of possible combinations exceeded 100, 100 combinations were randomly chosen, and the mean of all combinations was taken as the MIPopulation for this group size.

For the MIPopulation, the values of bias were also computed as follows: on average and for all sets of nine simultaneous recordings, MIPopulation was 0.104 in the CN, 0.111 in the IC, 0.114 in the MGB, 0.107 in A1, and 0.106 in VRB. There was no significant difference between the mean MIPopulation bias values in the different structures (unpaired t test, all \( p > 0.20 \)).

Statistics. To assess the significance of the multiple comparisons (vocoding process, four levels; masking noise conditions, three levels; auditory structure, five levels), we used an ANOVA for multiple factors to analyze the whole dataset. *Post hoc* pairwise tests were performed between the original condition and the vocoding or noisy conditions. They were corrected for multiple comparisons using Bonferroni corrections and were considered as significant if their \( p \) value was <0.05. All data are presented as mean values ± SEM.

Results

From a database of 2334 recordings collected in the five auditory structures, two criteria were used to include neuronal recordings in our analyses. A recording had to show significant responses to pure tones (see Materials and Methods) and an evoked firing rate significantly above spontaneous firing rate (200 ms before each original vocalization) for at least one of the four original vocalizations. Applying these two criteria led to the inclusion of 499 recordings in CN, 386 recordings in CNIC, 262 recordings in the ventral division of the MGB (MGv), 354 recordings in A1, and 95 recordings in VRB.

Table 1

<table>
<thead>
<tr>
<th>CN</th>
<th>CNIC</th>
<th>MGv</th>
<th>A1</th>
<th>VRB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of animals</td>
<td>10</td>
<td>11</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td>Number of recordings tested</td>
<td>672</td>
<td>478</td>
<td>421</td>
<td>286</td>
</tr>
<tr>
<td>TFRP only</td>
<td>499</td>
<td>386</td>
<td>386</td>
<td>354</td>
</tr>
<tr>
<td>TFRP and significant response to at least one vocalization</td>
<td>560</td>
<td>421</td>
<td>285</td>
<td>455</td>
</tr>
<tr>
<td>TFRP quantifications</td>
<td>0.18 – 18</td>
<td>0.34 – 16</td>
<td>0.33 – 33</td>
<td>0.14 – 16</td>
</tr>
<tr>
<td>Mean bandwidth (octave)</td>
<td>3.91</td>
<td>2.88</td>
<td>4.16</td>
<td>2.07</td>
</tr>
<tr>
<td>Mean response duration (ms)</td>
<td>26.83</td>
<td>35.37</td>
<td>17.31</td>
<td>43.73</td>
</tr>
<tr>
<td>Response strength (AP/s)</td>
<td>77.23</td>
<td>82.25</td>
<td>41.61</td>
<td>37.69</td>
</tr>
</tbody>
</table>

Figure 2. Evolution of the CorrCoef and MI mean values as a function of temporal precision in each structure. **A**, The trial-to-trial temporal reliability, quantified by the CorrCoef, was calculated from responses to original vocalizations with a width of Gaussian window varying from 1 to 50 ms in CN (in black), CNIC (in green), MGv (in orange), A1 (in blue), and VRB (in purple). In our study, a 10 ms width Gaussian window (dashed black line) was selected for the data analysis in each structure. **B**, Mutual information (in bits) was calculated from neuronal responses to original vocalizations with a bin size varying from 1 to 40 ms in CN (in black), CNIC (in green), MGv (in orange), A1 (in blue), and VRB (in purple). In this study, the value of 8 ms was selected for the data analysis because in each structure the MI value was maximal (dashed black line). This holds true also in the different conditions of acoustic alterations, both in noisy and vocoded conditions (data not shown).

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

**Determination of optimal parameters for temporal analyses of spike trains in the five auditory structures**

Before quantifying the neuronal discrimination performance in the five investigated structures, we first looked for the optimal parameters for analyzing the temporal patterns of spike trains in the five structures.
First, the CorrCoef index, which quantifies the trial-to-trial temporal reliability, was computed with a Gaussian window ranging from 1 to 50 ms. As a general rule, the largest the Gaussian window, the largest the CorrCoef mean value, whatever the structure was. We questioned whether selecting a particular value for the Gaussian window influenced the between-structure differences in CorrCoef mean values. Based on the responses to the original vocalizations, Figure 2A shows that the relative ranking between auditory structures remained unchanged whatever the width of the Gaussian window was. Therefore, we kept the value of 10 ms for the Gaussian window (Fig. 2A, dashed line) as it was used in previous cortical studies (Huët et al., 2009; Gaucher et al., 2013; Gaucher and Edeline, 2015; Aushana et al., 2018).

Second, at the cortical level, it was previously shown that the maximal value of MI based on temporal patterns was obtained, on average, with a bin size of 8 ms (Schnupp et al., 2006; Gaucher et al., 2013). However, it has never been demonstrated that the same bin size was optimal at all levels of the auditory system. Figure 2B shows the evolution of MI as a function of temporal precision ranging from 1 to 40 ms based on the responses to the original vocalizations. In our experimental conditions, and with our set of acoustic stimuli, the 8 ms temporal precision was found to be optimal for all auditory structures, in the original (Fig. 2B, dashed line), vocoded, and noisy conditions (data not shown). Therefore, the MI value obtained for a temporal precision of 8 ms was subsequently used in our analyses.

Subcortical structures better discriminate the original vocalizations

Figure 3A displays neuronal responses of two simultaneous multiunit recordings obtained at five levels of the auditory pathway (CN, CNIC, MGv, A1, and VRB). The neuronal responses were strong and sustained in the CN and CNIC, more transient in MGv, phasic in A1, and more diffuse in VRB. For most of the recordings, temporal patterns of response were clearly reproduci-
ble from trial to trial, but they differed from one vocalization to another both at the cortical and subcortical levels. The PSTHs displayed in Figure 3B show that at each level of the auditory system, the four whistles triggered distinct temporal patterns of responses.

Quantifications of evoked responses to original vocalizations over all the recordings are presented in Figure 3C–F for each auditory structure. These analyses clearly pointed out large differences among the mean values of evoked firing rate, CorrCoef, and MI quantified at the cortical level versus the subcortical level. First, the evoked firing rate was significantly higher in the subcortical structures than in the cortex (unpaired t test, lowest p value < 0.001). It was also higher in CN compared with the other subcortical structures (Fig. 3C). Second, the CorrCoef values were significantly higher in CNIC and MGv compared with A1 and VRB (Fig. 3D), indicating that the trial-to-trial reliability of evoked responses was stronger in these structures than in CN, A1, and VRB. Third, the MI\textsubscript{individual} values obtained at the subcortical level were significantly higher than at the cortical level (unpaired t test, highest p value < 0.001 between the cortex and the other structures; Fig. 3E). At the subcortical level, the MI\textsubscript{individual} values were significantly higher in MGv than in CNIC and CN (unpaired t test, p < 0.01) with the CN exhibiting the lowest MI values at the subcortical level. The MI\textsubscript{individual} values were also significantly lower in VRB than in A1 (p = 0.037). Recordings in MGv displayed the highest MI\textsubscript{individual} mean values, suggesting that, on average, thalamic neurons discriminate the four original whistles better than the other auditory structures. As shown in Figure 3G, in each auditory structure high MI\textsubscript{individual} values were strongly correlated with high values of trial-to-trial temporal reliability (indexed by the CorrCoef value; 0.77 < r < 0.88; p < 0.001). Finally, MI was also computed based on the temporal patterns obtained from 2 to 16 simultaneous multiunit recordings to determine whether the discrimination performance of neural networks confirms the results obtained at the individual (i.e., single-recording) level. MI\textsubscript{population} quantifies how well the four whistles can be discriminated based on temporal patterns expressed by neuronal populations distributed on the tonotopic map. The MI\textsubscript{population} values computed from nine simultaneous multiunit recordings show that neural populations in subcortical structures discriminate the four original whistles better than the cortical populations (unpaired t test, highest p value < 0.002 between CN and VRB) without any statistical difference among the three subcortical structures (Fig. 3F).

We next investigated the diversity of the MI\textsubscript{individual} and MI\textsubscript{population} values obtained in the different structures. The distributions of MI\textsubscript{individual} values were plotted as a function of temporal precision for each structure (Fig. 4A–E). The waterfall plots showed that whatever the temporal precision, there were more curves with high MI\textsubscript{individual} values in the subcortical structures than in the cortical areas (Fig. 4A–E, red curves). The examination of the evolution of the MI\textsubscript{population} as a function of the number of simultaneous multiunit recordings in the different structures revealed that the growth functions rapidly reached high values in all subcortical structures, whereas there were only a few such curves in A1 and VRB whatever the number of recordings considered (Fig. 4B–E).

With a temporal resolution of 8 ms, we presented the cumulative percentages of neurons for the MI\textsubscript{individual} (Fig. 5A) and the MI\textsubscript{population} values (Fig. 5B) in each structure. Above a value of 1.5 bits (indicating that at least three stimuli can be discriminated), there were 39% of MGv neurons, and 18% and 14%, respectively, of CNIC and CN neurons; but only 3.5% and 2%, respectively, of A1 and VRB neurons. This proportion was significantly higher in MGv than in CN and CNIC (p = 0.017 and p = 0.04) and was also significantly higher in subcortical structures compared with the cortical structures (all p values < 0.01). The same conclusions were reached for the MI\textsubscript{population} values. More than 90% of the MGv neuronal populations were >1.5 bits, and 83% and 75%, respectively, of the CNIC and CN populations, whereas these populations represented <40% at the cortical level (36% and 34%, respectively, in A1 and VRB).

Thus, both at the level of individual recordings, and at the population of simultaneous multiunit recordings, subcortical neurons are more accurate in discriminating the four original whistles than cortical ones.

Modest effects of tone vocoding

Figure 6A displays rasters of recordings obtained in the five structures in response to the original and tone-vocoded vocalizations using 38 (Voc38), 20 (Voc20), and 10 (Voc10) frequency bands. As illustrated by the rasters and the PSTHs presented in Figure 6B, in all structures neurons still vigorously responded to the vocoded stimuli even for 10-band vocoded stimuli.

Figure 6C–F summarizes the vocoding effects on the four parameters quantifying neuronal responses. Compared with the responses to the original vocalizations, the evoked firing rate obtained in all structures in response to vocoded stimuli only showed modest variations (Fig. 6C): apart from an increase in firing rate in the CN with the 38-band vocoded stimuli, a
Figure 3. Subcortical neurons discriminate better the original vocalizations than cortical neurons. **A**, From bottom to top, neuronal responses were recorded in CN, CNIC, MGv, A1, and VRB simultaneously under 16 electrodes but only two are represented here, with alternated black and red colors. Each dot represents the emission of an action potential, and each line corresponds to each presentation of one of four original whistles. The gray part of rasters corresponds to evoked activity. For each example, the values of the BF (in kHz) and of the BW (in octaves) obtained when testing the responses to pure tones are indicated on the left side. The waveforms of the four original whistles are displayed under the rasters. **B**, PSTHs of each neuronal response presented in **A**. For each neuronal recording, the four PSTHs of the four original whistles have been overlayed. **C–F**, The mean values of the evoked firing rate (in spikes/s; **C**), the trial-to-trial temporal reliability quantified by the CorrCoef value (**D**), the neuronal discrimination assessed by MI\textsubscript{individual} values (bits; **E**), and MI\textsubscript{population} (bits; **F**), with populations of nine simultaneous multiunit recordings obtained with the four original vocalizations in CN (in black), CNIC (in green), MGv (in orange), A1 (in blue), and VRB (in purple). The evoked firing rate corresponds to the total number of action potentials occurring during the presentation of the stimulus minus spontaneous activity (200 ms before each acoustic stimulus). In each structure, error bars represent the SEM of the mean values, and black lines represent significant differences between the mean values (unpaired t test, \(p < 0.05\)). The evoked firing rate decreases from the CN to VRB, but both the trial-to-trial temporal reliability (CorrCoef) and the discrimination performance (MI) reach a maximal value in MGv. Note also that at the population level, all the subcortical structures discriminate better the original vocalizations than cortical areas. **G**, Scatter plots showing in each structure, the strong correlations (0.77 < \(r < 0.88\)) between the CorrCoef and the MI\textsubscript{individual} (bits) values obtained in original conditions.
Figure 4. Diversity of neuronal discrimination performance in quiet for each structure at the individual and population levels. **A1–A5**, Neural discrimination performance in response to original vocalizations in each auditory structure. Waterfall plots show the MI (bits) as a function of temporal resolution (1–256 ms) for the selected recordings in CN (A1), CNIC (A2), MGv (A3), A1 (A4), and VRB (A5). In each structure, the units are ranked by the mean MI value obtained for all bin sizes. Note that there was a larger proportion of neurons with high values of MI (close to the maximal value of 2 bits) in MGv, CNIC, and CN (red curves) compared with a much lower proportion in the cortical areas A1 and VRB. **B1–B5**, Population information quickly reaches high values with simultaneous multiunit recordings at the subcortical level but not the cortical level. For each auditory structure, each thin line represents a particular case of simultaneous recording with a maximum number of electrodes (maximum of 16 simultaneous multiunit recordings), and each thick line represents the mean value of $MI_{\text{Population}}$ in CN (B1, in black), CNIC (B2, in green), MGv (B3, in orange), A1 (B4, in blue), and VRB (B5, in purple). Note that the mean $MI_{\text{Population}}$ value quickly reaches high values close to the maximum value of 2 bits in the subcortical structures (CN, CNIC, and MGv) compared with the two cortical areas (A1 and VRB).

Figure 5. High discrimination performance neurons are more numerous in subcortical structures than in auditory cortex in original conditions. **A, B**, Cumulative percentage of the neuronal discrimination performance obtained in original vocalizations assessed by $MI_{\text{Individual}}$ (in bits; A) and $MI_{\text{Population}}$ (in bits; B) with populations of nine simultaneous multiunit recordings in CN (in black), CNIC (in green), MGv (in orange), A1 (in blue), and VRB (in purple).
Figure 6. Vocoding slightly alters neuronal responses at each stage of the auditory system. A, From left to right, examples of raster plots representing the responses to the four original whistles (Original) and their vocoded versions (Voc38, Voc20, and Voc10). The gray part of rasters corresponds to evoked activity. From bottom to top, Neuronal responses were recorded in CN, CNIC, MGv, A1, and VRB. For each example, the values of the BF (in kHz) and of the BW (in octaves) obtained when testing the responses to pure tones are indicated on the left side. For each example, the mean evoked firing rate (in spikes/s) obtained in each condition is indicated below the rasters.

B, PSTHs of each neuronal response presented in A. For each neuronal recording, the four PSTHs of the original and vocoded conditions have been overlayed. The gray part of the PSTHs corresponds to evoked activity.

C–F, The mean values (±SEM) represent the vocoding effects on the evoked firing rate (in spikes/s; C), the temporal reliability represented by the CorrCoef value (D), the neuronal discrimination assessed by MIIndividual (in bits; E), and MIPopulation (in bits; F) with populations of nine simultaneous multunit recordings in CN (in black), CNIC (in green), MGv (in orange), A1 (in blue), and VRB (in purple; one-way ANOVA, p < 0.05; with post hoc paired t tests, *p < 0.05). The evoked firing rate corresponds to the total number of action potentials occurring during the presentation of the stimulus minus spontaneous activity.
significant decrease in evoked firing rate in response to the 10-band vocoded vocalizations was found only at the subcortical level (for all subcortical structures; with one-way ANOVA: \(F_{CNIC(3,1357)} = 22.6; F_{CN(3,1125)} = 8.85; F_{MGv(3,1047)} = 6.55; p < 0.001\); with post hoc paired t tests, \(p < 0.05\)), whereas there was no decrease in either A1 or VRB. Vocoding also decreased the mean CorrCoef values in every structure except in VRB (Fig. 6D). This decrease was significant with the 10-band vocoded vocalizations in CN, MGv, and A1 (one-way ANOVA: \(F_{CN(3,1415)} = 96.27; F_{MGv(3,1047)} = 7.77; F_{A1(3,1125)} = 3.42; \) highest \(p\) value, \(p < 0.02\); with post hoc paired t tests, \(p < 0.05\)). The decrease in CorrCoef value was already significant with 20-band vocoded vocalizations in the CNIC (one-way ANOVA: \(F_{CNIC(3,1391)} = 26.19, p < 0.001\); with post hoc paired t tests, \(p < 0.05\)).

Similarly, vocoding decreased the MI\(_{\text{Individual}}\) values in each structure except in VRB (Fig. 6E). Here too, the decrease was significant with the 10-band vocoded vocalizations in CN, MGv, and A1 (one-way ANOVA: \(F_{CN(3,1415)} = 12.23, F_{MGv(3,1047)} = 3.75, F_{A1(3,1320)} = 3.59; \) highest \(p\) value, \(p < 0.02\); with post hoc paired t tests, \(p < 0.05\)), and it was already significant with 20-band vocoded vocalizations in the CNIC (one-way ANOVA: \(F_{CNIC(3,1323)} = 13.17, p < 0.001\); with post hoc paired t tests, \(p < 0.05\)). At the population level (MI\(_{\text{Population}}\)), compared with the values obtained in response to the original vocalizations, the MI\(_{\text{Population}}\) values computed with the 10-band vocoded vocalizations were significantly lower in the subcortical structures (one-way ANOVA: \(F_{CN(3,3127)} = 6.46, F_{CNIC(3,3115)} = 6.28, F_{MGv(3,867)} = 4.62; \) highest \(p\) value, \(p < 0.005\); with post hoc paired t tests, \(p < 0.05\)) but not at the cortical level (Fig. 6F). The evolution of MI\(_{\text{Population}}\) as a function of the number of simultaneous multiunit recordings (Fig. 7A–E) showed that in each subcortical structure, the curves rapidly reached high MI\(_{\text{Population}}\) values (close to the maximal value of 2 bits) in each vocoding conditions, whereas in A1 and VRB the curves slowly reached the maximum MI\(_{\text{Population}}\) values.

In conclusion, for the five auditory structures, the neuronal responses to 10-band vocoded vocalizations were slightly weaker, temporally less accurate, and less discriminative than the responses to the original vocalizations. Nonetheless, on average, subcortical structures still maintained the highest discrimination performance between tone-vocoded vocalizations, both at the level of individual recordings and at the population level.

Pronounced effects of masking noise on neuronal discrimination

The rasters presented in Figure 8A illustrate the effects induced by presenting the original vocalizations against a vocalization-shaped stationary noise at three SNRs (+10, 0, and −10 dB). As illustrated by the rasters and the PSTHs presented in Figure 8B, masking noise attenuated neuronal responses at each level of the auditory system. However, the auditory structures were differentially affected by noise. The responses in the CNIC did not change up to a 0 dB SNR, decreasing only at a −10 dB SNR. This was not the case in the other auditory structures where the responses decreased either at a +10 dB SNR (MGv and CN) or at a 0 dB SNR (A1 and VRB).

Figure 8C–F summarizes the effects of masking noise on the different parameters quantifying neuronal responses. Masking noise significantly reduced the evoked firing rate in each auditory structure already at the +10 dB SNR (Fig. 8C; one-way ANOVA: \(F_{CN(3,1415)} = 309.33, F_{CNIC(3,1543)} = 220.64, F_{MGv(3,1047)} = 155.07, F_{A1(3,1415)} = 96.27; p < 0.001\); with post hoc paired t tests, \(p < 0.05\)) except in VRB.

At the subcortical level, masking noise strongly reduced the CorrCoef values in CN and MGv at the highest SNR (+10 dB) tested here (Fig. 8D; one-way ANOVA: \(F_{CN(3,1884)} = 382.22, F_{MGv(3,791)} = 155.82, p < 0.001\); with post hoc paired t tests, \(p < 0.05\)), whereas in the CNIC, this reduction was significant only at the 0 dB SNR (one-way ANOVA: \(F_{CNIC(3,1537)} = 154.12, p < 0.001\); with post hoc paired t tests, \(p < 0.05\)). At the cortical level, the CorrCoef values were significantly reduced in A1 at the +10 dB SNR and in VRB at the 0 dB SNR (one-way ANOVA: \(F_{A1(3,1093)} = 60.83, F_{VRB(3,335)} = 29.56, p < 0.001\); with post hoc paired t tests, \(p < 0.05\)) at the subcortical level, noise reduced the MI\(_{\text{Individual}}\) values, but again, there was a marked difference between the CNIC and the other subcortical structures: the MI\(_{\text{Individual}}\) mean value in CN and MGv was significantly reduced at the +10 dB SNR (Fig. 8E; one-way ANOVA: \(F_{CN(3,819)} = 56.75, F_{MGv(3,621)} = 63.61, p < 0.001\); with post hoc paired t tests, \(p < 0.05\)), whereas the MI\(_{\text{Individual}}\) value in the CNIC was only significantly reduced at the 0 dB SNR (one-way ANOVA: \(F_{CNIC(3,1078)} = 32.08, p < 0.001\); with post hoc paired t tests, \(p < 0.05\)). At the cortical level, noise significantly reduced the average MI\(_{\text{Individual}}\) in A1 only at the −10 dB SNR (one-way ANOVA: \(F_{A1(3,649)} = 9.49, p < 0.001\); with
Noise strongly reduces neuronal responses in all structures, but to a lesser extent in the central nucleus of the inferior colliculus. A. From left to right, raster plots of responses of four original whistles (Original) and their noisy versions embedded in the vocalization-shaped stationary noise at three SNRs: +10, 0, and −10 dB. The gray part of rasters corresponds to evoked activity. From bottom to top, neuronal responses were recorded in CN, CNIC, MGv, A1, and VRB. For each example, the values of the BF (in kHz) and BW (in octaves) obtained when testing the responses to pure tones are indicated on the left side. For each example, the mean evoked firing rate (in spikes/s) obtained in each condition is indicated below the rasters. The green dashed box indicates a typical example of CNIC neuronal responses that are resistant to the noise addition.

B, PSTHs of each neuronal response presented in A. For each neuronal recording, the four PSTHs of the original and noisy conditions have been overlayed. The gray part of the PSTHs corresponds to evoked activity.

C–F, The mean values (± SEM)
In the CNIC, noise induced SNR-dependent reduction in the MIPopulation values, the reduction being modest at a 10 dB SNR. At the cortical level, there was a significant reduction of MIPopulation values only at the −10 dB SNR (one-way ANOVA: \( F_{(3127)} = 16.63, F_{(VRB(3,23)} = 11.41, p < 0.001 \); with post hoc paired t tests, \( p < 0.05 \)). However, at least 20% of the CN recordings at the 10 dB SNR but more important at a −10 dB SNR. In the MGv, noise progressively lowered the curves of the MIPopulation. In the cortex, the MIPopulation growth functions were not strongly impacted except at the −10 dB SNR.

The effects of masking noise on the network discrimination performance were quantified with the MIpopulation (Fig. 8F). At the cortical level, there was a significant reduction of MIpopulation values only at the −10 dB SNR (one-way ANOVA: \( F_{AI(3111)} = 16.63, F_{VRB(3,23)} = 11.41, p < 0.001 \); with post hoc paired t tests, \( p < 0.05 \)), whereas there was a significant decrease in CN already at the +10 dB SNR (one-way ANOVA: \( F_{CN(3127)} = 51.49, p < 0.001 \); with post hoc paired t tests, \( p < 0.05 \)). In MGv and CNIC, neuronal populations still displayed the highest discrimination performance, although the decrease in MIpopulation value was significant at the 0 dB SNR (one-way ANOVA: \( F_{MGv(3,67)} = 41.59, F_{CNIC(3115)} = 22.59, p < 0.001 \); with post hoc paired t tests, \( p < 0.05 \)).

Note that, in VRB, the CorrCoef and MIpopulation were much more decreased in the noise conditions than in the vocoding conditions, suggesting that the lack of significant decreases in vocoding conditions was not a “floor effect” due to the low initial values.

The evolution of the MIpopulation as a function of the number of simultaneous multiunit recordings in the different structures (Fig. 9A–E) revealed that, regardless of the number of neurons considered, noise effects were similar up to the 0 dB SNR: the curves display the noise effects on the MIPopulation growth functions for each structure and at each SNR (indicated by a gradient colors) in CN (A, in black), CNIC (B, in green), MGv (C, in orange), A1 (D, in blue), and VRB (E, in purple). In general, background noise largely altered the growth functions of the MIPopulation in each structure (but to a lesser extent in the CNIC). In the CN, noise induced a stronger reduction of the MIpopulation, which was clearly a function of SNR. In the CNIC, noise induced SNR-dependent reduction in the MIPopulation values, the reduction being modest at a +10 and 0 dB SNR but more important at a −10 dB SNR. In the MGv, noise progressively lowered the curves of the MIPopulation. In the cortex, the MIPopulation growth functions were not strongly impacted except at the −10 dB SNR.

The distributions of the TFRP parameters (best frequency, bandwidth, response duration, response strength) from this specific subpopulation of CN neurons did not differ from the neurons exhibiting MIindividual values <1 bit at the +10 dB SNR in terms of best frequency and bandwidth, but significantly differ in terms of response duration and response strength (\( \chi^2 \) tests, \( p < 0.05 \)). More precisely, the CN recordings exhibiting higher MIindividual values at +10 dB SNR had longer duration responses and stronger evoked firing rates to pure tones.

A more general question is to evaluate whether the TFRP characteristics in the different auditory structures (Fig. 11A, examples) influenced the noise effects quantified by the MIindividual Values (Fig. 11B,C). As indicated in Figure 11, there was no relationship between the best frequency values and the changes in MIindividual values (Fig. 11B) and no relationship between the frequency bandwidth and the changes in MIindividual values (Fig. 11C). Thus, in all structures, the noise-induced alterations in MIindividual values seem to be independent from the characteristics of pure tone responses.

To summarize, masking noise differently impacted the discrimination performance of neurons at the subcortical and cortical levels. Although cortical neurons were more resistant to changes in noise level, the thalamic and collicular neurons maintained higher MI values, with the CNIC neurons displaying the highest discrimination performance both at the individual and population level in the most challenging condition (i.e., at the −10 dB SNR).

Alteration of slow amplitude modulations as one of the factors explaining the changes in neuronal discrimination

Masking noise produced spectrotemporal degradations: it reduced the AM cues in the different audio frequency bands, introduced irrelevant envelope fluctuations, and altered the TFS of the sound. Tone vocoding removed almost all the TFS but also progressively filtered out the fast AM. As a vast literature...
demonstrated that slow AM cues are crucial for speech understanding in normal and degraded conditions (Houtgast and Steeneken, 1985; Drullman et al., 1994; Drullman, 1995; Shannon et al., 1995; Dubbelboer and Houtgast, 2007; Jørgensen and Dau, 2011), we quantified the alterations of AM cues (due to masking noise and to vocoding) and looked for potential relationships with the alterations in neural discrimination (MIPopulation) in the five structures.

The AM spectra obtained in vocoding and noise conditions showed that the AM cues were attenuated compared with the original condition (Fig. 12A). The +10 dB SNR condition produced a flattening of the AM modulation spectrum, which was further accentuated in the 0 and −10 dB SNR conditions. In these two most degraded conditions, noise also introduced non-relevant fluctuations at high rates. In contrast, vocoding preserved the general shape of the AM spectra while progressively filtering out the AM fluctuations.

We investigated the relationships between these degradations of AM cues and neural discrimination (MIPopulation) in the five structures for each experimental condition (Fig. 12B). More precisely, for all conditions, Figure 12B shows the changes in MIPopulation for each auditory structure as a function of the attenuation of AM cues (computed as the mean modulation index between 1 and 20 Hz).

Figure 12B reveals that in all structures other than the CN, MIPopulation is barely affected as long as the reduction of the AM index (D modulation index) remains ≤25%; beyond this limit, the MIPopulation is reduced (i.e., at the 0 and −10 dB SNR). The straightforward conclusion is that the reduction of slow AM cues is one of the factors controlling the decrease in MIPopulation at the cortical and subcortical levels. In the cochlear nucleus, the decrease in the MIPopulation is much larger than in the other structures, suggesting that the alteration of AM cues has more impact on the MIPopulation at the most...
peripheral level. Alternatively, one should keep in mind that the neuronal discrimination in noise can be based on other acoustic cues such as the FM cues (in particular pitch cues), spectral regularity, and harmonicity cues, and the simultaneous rising slope of energy across channels. Thus, in the cochlear nucleus, but also in the other structures, the strong decrease in MIPopulation can potentially result from alterations of one, or several, of these parameters.

Dissecting the contributions of each of these parameters to neuronal discrimination and its decrease in degraded conditions will require manipulations of controlled stimuli in independent conditions. Confirming that the slow AM cues are the main factor for discrimination in degraded conditions could theoretically be achieved by keeping the exact same AM cues and modifying only one of the acoustic parameters listed above. Using a computational model of the peripheral auditory system will help to estimate the respective representations of the envelope and temporal

fine structure after acoustic degradations (Moon et al., 2014; Wirtzfeld et al., 2017). For example, the search for “equivalent” experimental conditions in terms of amounts of neural degradation of AM and FM cues could be performed by using the FAME vocoder (Zeng et al., 2005) to alter systematically AM and FM parameters (i.e., cutoff frequency, modulation strength, modulation phase) of the vocalizations used as stimuli. The results of this type of experiment should also be generalized with other categories of guinea pig calls and other types of communication sounds from other species, and should be included in other types of masking noises.

Discussion
Here, we demonstrated that for each acoustic distortion, subcortical neurons displayed the highest level of discrimination performance of natural vocalizations, either at the collicular level (in
masking noise conditions) or at the thalamic level (in vocoder conditions). More precisely, background noise markedly reduces neural discrimination performance in all auditory structures with larger effects in the cochlear nucleus, whereas the vocoder induced little effect in each auditory structure. Interestingly, the discrimination performance of cortical neurons was less impacted, making these neurons more robust to all acoustic alterations. Moreover, the comparison of neural data collected in response to noisy versus vocoded vocalizations suggests that the transmission of slow AM cues (70 Hz) selected for the data analysis.

Figure 12. Reduction of slow AM cues as one of the factors explaining the neuronal discrimination performance at the subcortical and cortical levels. A, Vocoding and noise effects on the AM spectra. The plot represents the averaged modulation spectra of the four original vocalizations (in black), vocoded vocalizations (Voc38, Voc20, and Voc10: red, green, and blue, respectively, solid lines), and vocalizations in noise at three SNRs (+10, 0, and −10 dB: red, green, and blue respectively, dashed lines). Vertical black dashed line corresponds to the maximum frequency (20 Hz) selected for the data analysis. B, Percentage of ΔMIPopulation as a function of modulation index computed for each structure from mean MIPopulation or mean modulation index values obtained in all adverse conditions and mean values in the original condition. Each dot represents neuronal data (ΔMIPopulation) in CN (in black), CNIC (in green), MGv (in orange), A1 (in blue), and VRB (in purple). Polynomial curves fitting all acoustic conditions have been generated (color lines). In all conditions (vocoding or noise), there is a limit of AM reduction from which the ΔMIPopulation decreases in cortical and subcortical structures.

Subcortical structures represent natural vocalizations more precisely than primary and nonprimary cortical areas

In contrast with previous cortical studies, which have quantified the discrimination between calls that belong to different categories making the discrimination easy for cortical neurons (Narayan et al., 2006, 2007; Ter-Mikaelian et al., 2013; Ni et al., 2017), we used four vocalizations that belong to the same category making the discrimination more difficult for cortical neurons. We showed that on average subcortical populations discriminated the original vocalizations better than cortical populations. Moreover, smaller populations of subcortical neurons compared with cortical ones were sufficient to discriminate between the stimuli used in this study. These results corroborate the finding by Chechik et al. (2006) that the MGB and A1 responses contain 2- to 4-fold less information than the responses of IC neurons. Here, the discrimination performance in MGv was closer than the ones displayed by the other subcortical structures. A potential explanation is that Chechik et al. (2006) recorded from all MGB divisions, including the medial and dorsal divisions, whereas our thalamic recordings were limited to MGv and exhibited tonic responses to vocalizations similar to those observed in the CNIC and the CN (Figs. 3A, 5A). The stimulus sets also differ, as we used four utterances of the same category (the whistle), whereas Chechik et al. (2006) used chirps from three birds and variants of these stimuli, leading potentially to an easier classification between groups of stimuli compared with our protocol. An interesting result was that the optimal bin size for computing MI was similar for all structures (8 ms bin; Fig. 2B). Importantly, with a smaller or a larger bin, the mutual information would have been underestimated, but this would not have changed the differences reported here: whatever the bin size, subcortical neurons will still discriminate better in the original vocalizations than in the cortical areas (Fig. 2B). Potentially, the optimal bin size depends more on the stimuli durations than on the auditory structure. When computing mutual information from IC, MGB, and A1 neuronal responses, Chechik et al. (2006) usually found an optimal bin size of 4 ms, which was different from that in our study, probably because their stimulus durations are shorter than our stimuli (67–111 ms vs 280–363 ms, respectively). Recently, we also found shorter optimal bin sizes when computing MI with shorter (12–65 ms) communication sounds (Royer et al., 2019).

Our original stimuli differed in terms of temporal envelope, and, as a consequence, the most efficient way to discriminate them is probably to follow the time course of AM cues. It is well known that when progressing from the lower to the upper stages of the auditory system, the ability of neurons to follow AM cues considerably changes (Joris et al., 2004; Escabi and Read, 2005). Brainstem neurons phase lock on AM modulations up to hundreds of hertz (Frisina et al., 1990; Rhode and Greenberg, 1994), whereas thalamic neurons do so for a few tens of hertz (Creutzfeldt et al., 1980; Preuss and Müller-Preuss, 1990) and cortical neurons do so for even lower rates (Schreiner and Urban, 1988; Gaese and Ostwald, 1995). As a consequence, subcortical neurons, (but not cortical ones) can follow the largest and fastest AM cues (7–15 Hz) contained in the original vocalizations (Fig. 12A, peak of the black curve in AM spectra). This likely explains why subcortical neurons better discriminate the original stimuli.
both at the individual and population levels. Cortical neurons only follow the weakest and slowest AM cues (1–5 Hz) of the original vocalizations, which potentially explains why cortical neurons weakly discriminate the original stimuli and tend to encode them as a single category (Mesgarani et al., 2014).

Alterations of slow amplitude modulation cues is one of the factors explaining the changes in cortical and subcortical discrimination

Previous studies using vocoded vocalizations reported that cortical responses were not drastically reduced even with two frequency bands (Nagarajan et al., 2002; Ranasinghe et al., 2012; Ter-Mikaelian et al., 2013; Aushana et al., 2018). At the level of A1, studies have pointed out the relationships between the noise impact on the cortical and behavioral discrimination performance. In bird field L (homologous to A1), neuronal responses to song motifs were strongly reduced by three types of masking noises, and the neural discrimination performance was progressively reduced when the SNR decreased, in parallel with the behavioral performance (Narayan et al., 2007). Our VRB results are reminiscent of those obtained in area NCM (homologous to a secondary area) where feedforward inhibition allowed the emergence of invariant neural representations of target songs in noise conditions (Schneider and Woolley, 2013). Similar to the results in the study by Ranasinghe et al. (2012), our IC neuronal responses were found to be resistant to drastic spectral degradations.

Only one study directly compared the impact of vocoding and masking noise on cortical responses to vocalizations (Nagarajan et al., 2002). In this study, auditory cortex responses were robust to spectral degradations even in response to 2-band vocoded vocalizations. Also, broadband white noise reduced neuronal responses at 0 dB SNR. Last, temporal envelope degradations strongly reduced the evoked firing rate and the neural synchronization to the vocalization envelope. Importantly, bandpass filtering the vocalizations between 2–30 Hz did not reduce firing rate and neural synchronization to the vocalization envelope. This is in agreement with the following results in our conditions: when the $\Delta$Modulation index (computed between 1 and 20 Hz) revealed modest AM alterations, there was little effect on the neuronal discrimination, but when the AM alterations reached approximately $\geq 20$–$30\%$, the neuronal discriminations were reduced (Fig. 12B). Thus, our results are consistent with the hypothesis that one of the factors constraining auditory discrimination at the cortical and subcortical levels is the fidelity of transmission and processing of slow AM cues.

When quantifying how different noise levels alter neuronal coding in the auditory system, it was found that the neural representation of natural sounds becomes progressively independent of the level of background noise from the auditory nerve to the IC and A1 (Rabinowitz et al., 2013). It was proposed that this tolerance to background noise results from an adaptation to the noise statistics, which is more pronounced at the cortical than at the subcortical level. In agreement with this study, we found that populations of cortical neurons (A1 and VRB) were more resistant to noise than subcortical ones. However, we did not observe a monotonic evolution of resistance to noise in the auditory system: at the subcortical level, the discrimination performance of CN neuronal populations drastically dropped as early as $+10$ dB SNR, the populations of CNIC neurons maintained the highest discrimination performance even at the $-10$ dB SNR, and those of thalamic neurons largely decreased at 0 dB SNR, whereas cortical neurons showed the lowest discrimination performance at all SNRs but were more robust to noise. In the IC, previous work showed that background noise changes the shape of the temporal modulation transfer function of individual neurons from bandpass to lowpass (Lesica and Grothe, 2008). The CNIC is a massive hub receiving probably the highest diversity of inhibitory and excitatory inputs (Malmierca, 2004; Ayala et al., 2016), and potentially the large diversity of these inputs allows this structure to extract crucial temporal information about the stimulus temporal envelope, even at a relatively low SNR.

Limitations of the study

We previously did not find evidence for higher cortical discrimination in awake animals compared with anesthetized animals (Huetz et al., 2009): with normal and reversed whistle stimuli, the percentage of cortical cells with significant MI values was higher in anesthetized (71%) than in awake animals (44%; Huetz et al., 2009, their Table 1). In addition, the $H_{\text{max}}$ value (equivalent of MI) was higher in anesthetized than in awake animals (0.38 vs 0.24; Huetz et al., 2009, their Table 2). Last, the trial-to-trial temporal reliability of cortical cells to whistle calls was not different in anesthetized and awake guinea pigs (anesthetized, 0.48; vs awake, 0.42; Huetz et al., 2009, their Fig. 8). A recent study (Town et al., 2018) revealed that the cortical discrimination performance between vowels observed in awake animals using acoustic degradations were similar in anesthetized animals (Bizley et al., 2009). Therefore, based on these two studies, the cortical discrimination performance can only be slightly lower or similar in awake compared with anesthetized animals. At the subcortical level, it seems that there is not a large difference between the phase-locking properties of neurons in anesthetized and awake animals (Joris et al., 2004). Temporal properties of IC neurons are only mildly affected by anesthesia (Ter-Mikaelian et al., 2013), indicating that collicular neurons will still be far better than cortical ones to follow the 10–20 Hz temporal cues contained in the four vocalizations. Together, these studies suggest that the hierarchy between cortical and subcortical structures in discriminating communication sounds should be more pronounced or should remain the same in awake animals.

Another limitation of the present study lies in the use of a limited set of stimuli that is restricted to the same four whistles. However, the four whistles used here were clearly representative of our whole database of whistles in terms of frequency range, duration, range of frequency, and amplitude modulations. Changing the four whistles from one recording to another can help in generalizing the results, but the main advantage of using exactly the same four whistles is that from one recording to the next, and from one structure to another, we were sure that the same acoustic cues were available for the neural discrimination. However, the whistles are a subset of the guinea pig repertoire, and therefore the present results may not generalize to other communication sounds, and larger sets of stimuli should be used to confirm that the slow AM cues control the neural discrimination. Even if amplitude modulations seem the main cues for speech understanding (Druillem et al., 1994; Shannon et al., 1995), other factors (the pitch, the frequency modulation, the harmony/cues) can also be involved.

As our results are based on multunit recordings, we do not know whether the same number of neurons was present in the cluster recordings from the different structures, and whether the individual discrimination performances of the cell types found in each structure are equivalent. On the other hand, the MI evaluated here is the reflection of a local computation performed by a small population of individual neurons, which gives us a good estimation of the whole discrimination performance of a given structure.
Functional implications

In humans, speech sounds (such as phonemes) showing similar acoustic properties trigger similar responses and are represented as a single category in the superior temporal gyrus (Mesgarani et al., 2014). As already proposed by Chechick and Nelken (2012), auditory cortex neurons extract abstract auditory entities rather than detailed spectrotemporal features. Obviously, this urges the definition of the acoustic features that form a category of auditory objects. It is relatively easy to delimit broad categories such as environmental sounds, animal vocalizations, music, and speech (Singh and Theunissen, 2003; Gygi et al., 2004, 2007; Woolley et al., 2005; Gygi and Shafiro, 2013) in terms of modulation cues, but within these categories, defining invariant features is a difficult task. Here, the use of vocalizations belonging to the same category of the guinea pig repertoire (i.e., “whistles”) may explain both the relatively poor discrimination abilities of cortical neurons compared with subcortical neurons and the robustness of cortical responses to vocoding and background noise.

From the present study, it appears that the subcortical structures engage significantly more neurons (20–40%) with high discrimination performance than the cortical areas (2–3%; Fig. 5A), confirming that the neural code is rather sparse at the cortical level (Hromádka et al., 2008), which might not be the case at the subcortical level. However, it is also possible that top-down projections coming from auditory cortex and reaching the thalamus, inferior colliculus, and cochlear nucleus (Jacomme et al., 2003; Malmierca and Ryugo, 2011) influence the neural discrimination at the subcortical level, especially in awake, behaving, animals. Thus, we can envision that in behaving animals, learning-induced cortical plasticity also contributes to enhancing the subcortical neural discrimination via the corticofugal projections. Further studies are required to determine to what extent these subcortical representations influence auditory abilities in animals and humans.

References


