Noise-sensitive but more precise subcortical representations co-exist with robust cortical encoding of natural vocalizations

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

JNeurosci
THE JOURNAL OF NEUROSCIENCE

Research Report: Regular Manuscript

Noise-sensitive but more precise subcortical representations co-exist with robust cortical encoding of natural vocalizations

https://doi.org/10.1523/JNEUROSCI.2731-19.2020

Cite as: J. Neurosci 2020; 10.1523/JNEUROSCI.2731-19.2020

Received: 18 November 2019
Revised: 8 May 2020
Accepted: 15 May 2020

Noise-sensitive but more precise subcortical representations co-exist with robust cortical encoding of natural vocalizations

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

[1] Paris-Saclay Institute of Neurosciences (Neuro-PSI), Department Cognition and Behavior CNRS UMR 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.

Abbreviated title: Cortical and subcortical discrimination in noise

Number of pages: 38
Number of Figures: 12
Number of Table: 1
Number of words in the abstract: 204
Number of words in the introduction: 598
Number of words in the discussion: 2149

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 and LV were 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 the Entendre group.

We thank Roger Mundry for his detailed and relevant comments on statistical analyses, Nihaad Paraouty for teaching us the cochlear-nucleus surgery and Quentin Gaucher for careful reading of this manuscript. We also wish to thank Mélanie Dumont, Aurélie Bonilla and Céline Dubois for taking care of the guinea-pig colony.

* *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

Abstract

Humans and animals maintain accurate sound discrimination in the presence of loud sources of background noise. It is commonly assumed that this ability relies on the robustness of auditory cortex responses. However, 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 multiunit responses recorded in the cochlear nucleus, inferior colliculus, auditory thalamus, primary and secondary auditory cortex at several signal-to-noise ratios (SNR) 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.

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 - the analogous of primary auditory cortex (A1) in bird - the percentage of correct neuronal discrimination between zebra-finch songs embedded in different types of acoustic maskers decreases proportionally to the target-to-masker ratio and parallels behavioral performance (Narayan et al., 2007). Also, consistent to behavioral data (for review see Verhey et al., 2003), the co-modulation 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 three disruptive effects on communication sounds (Noordhoek and Drullman, 1997; Dubbelboer and Houtgast, 2007): it attenuates the power of their amplitude modulation components (AM, also called “temporal-envelope”; Houtgast and Steeneken, 1985; Ewert and Dau, 2000; Biberger and Ewert, 2017), corrupts their frequency modulation components (FM, also called “temporal fine structure”; Shamma and Lorenzi, 2013; Varnet et al., 2017) and introduces stochastic fluctuations in AM power which 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-shaped 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 spectro-temporal details promoting the categorization of sounds into auditory objects (Nelken and Bar-Yosef, 2008; Chechik and Nelken, 2012). Mutual information was used to determine if 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 to 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 to 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 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 (AI) 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 to three days before each experiment, the animal's pure-tone audiogram was determined by testing auditory brainstem responses (ABR) under isoflurane anaesthesia (2.5 %) as described in Gourévitch et colleagues (2009). The ABR was obtained by differential recordings between two subdermal electrodes (SC25-NeuroService) placed at the vertex and behind the mastoid bone. 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. 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 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 (area 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 height. 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 (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, 2000; 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 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 (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 (Edeline and Weinberger, 1993; Manunta

and Edeline, 1999; Edeline et al., 2001; Wallace et al., 2000). The non-primary cortical area VRB was located ventral to A1 and distinguished because of its long latencies to pure tones (Grimsley et al., 2012; Rutkowski et al., 2002). 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 of tonotopic gradients recorded with such arrays in figure 1 of Gaucher and colleagues (2012) and in figure 6A of Occelli and colleagues (2016)].

All the remaining extracellular recordings (in MGB, IC and CN) 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 and ventral divisions.

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., 2002; 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.

Acoustic stimuli

Acoustic stimuli were generated using MATLAB (The Mathworks, Natick, MA), 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, a distance at which the speaker produced a flat spectrum (± 3 dB) between 140 Hz and 36 kHz. The stimulation was not purely monaural, but the animal's head and body largely attenuated binaural cues. Calibration of the speaker was made using noise and pure tones recorded by a Bruel & Kjaer microphone 4133 coupled to a preamplifier B&K 2169 and a digital recorder Marantz PMD671.

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, despite the fact 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 and spectrum (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, Voc10 respectively) for the 38-band (20-band, 10-band, respectively) vocoded whistles. In contrast to previous studies that used noise-excited vocoders (Nagarajan et al., 2002; Ranasinghe 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 spectro-temporal modulation features of communication sounds (Shamma and Lorenzi, 2013; Kates, 2011; Stone et al., 2011).

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) 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-band or the 10-band vocoders. Each digitized signal was passed through a bank of 38 fourth-order Gammatone filters (Patterson, 1987) with center frequencies uniformly spaced along a guinea-pig adapted ERB (Equivalent Rectangular Bandwidth) scale ranging from 20 to 35505 Hz (Sayles and Winter, 2010). In each frequency band, the temporal envelope was extracted using full-wave rectification and lowpass filtering at 64 Hz with a zero-phase, sixth-order Butterworth filter. The resulting envelopes were used to amplitude modulate sine-wave carriers with frequencies at the center frequency of the Gammatone filters, and with random starting phase. Impulse responses were peak-aligned for the envelope (using a group delay of 16 ms) and the temporal fine structure across frequency channels (Hohmann, 2002). The modulated signals were finally weighted and summed over the 38 frequency bands. The weighting compensated for imperfect superposition of the bands' impulse responses 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-band, 20-band 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-seconds of audio recordings were added up to generate a "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. Figure 1G displays the spectrograms of the four whistles in the vocalization-shaped stationary noise with a SNR of +10 dB SPL, 0 dB SPL, -10 dB SPL. Figure 1H shows the long-term power spectrum of the four whistles at the +10 dB, 0 dB and -10 dB SNR, and figure 1I shows their global temporal envelopes (which were severely altered at the 0 dB and -10 dB SNR).

Amplitude-modulation (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 (spanning the 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 Varnet and colleagues (2017) for the guinea pigs' hearing range (1/3-octave wide first-order Butterworth bandpass filters overlapping at -3 dB, with center frequencies between 0.1 Hz and 410 Hz). The root-mean-square amplitude of the filtered output was multiplied by a factor of $\sqrt{2}$. For each AM filter, 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-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 the electrodes' depth. For auditory cortex recordings (AI 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 4 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 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

Quantification of responses to pure tones

The TFRP were obtained by constructing post-stimulus time histograms (PSTH) 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 are 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 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 standard deviation of the spontaneous activity. Recordings without 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.

Quantification of responses evoked by vocalizations

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 minus spontaneous activity; (ii) the trial-to-trial temporal reliability coefficient (named CorrCoef as in our previous studies: Gaucher et al., 2013a; Huetz et al., 2014; Gaucher and Edeline, 2015; Aushana et al., 2018) which quantifies the trial-to-trial reliability of the responses 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-msec 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).

Quantification of mutual information from the responses to vocalizations

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

The method relies on a pattern-recognition algorithm that is designed to "guess which stimulus evoked a particular response pattern" (Schnupp et al., 2006) by going through the following steps: From all the responses of a cortical site to the different stimuli, a single response (test pattern) is extracted and represented as a PSTH with a given bin size (different sizes were considered as indicated in the Results section). Then, a mean response pattern is computed from the remaining responses (training set) for each stimulus class. The test pattern is then assigned to the stimulus class of the closest mean response pattern. This operation is repeated for all the responses, generating a confusion matrix where each response is assigned to a given stimulus class. From this confusion 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, or in other words, the values taken by the random variables "presented stimulus class" and "assigned stimulus class".

In our case, we used responses to the 4 whistles and selected the first 280 ms of these responses to work on spike trains of exactly the same duration (the shortest whistle being 280

ms long). In a scenario where the responses do not carry information, the assignments of each response to a mean response pattern is equivalent to chance level (here 0.25 because we used 4 different stimuli and each stimulus was presented the same number of times) and the MI would be close to zero. In the opposite case, when responses are very different between stimulus classes and very similar within a stimulus class, the confusion matrix would be diagonal and the mutual information would tend to $\log_2(4) = 2$ bits.

This algorithm was applied with different bin sizes ranging from 1 to 280 ms (see figure 2B for the evolution of MI with temporal precisions ranging from 1 to 40 ms).

The MI estimates are subject to non-negligible positive sampling biases. Therefore, as in Schnupp and colleagues (2006), we estimated the expected size of this bias by calculating MI values for "shuffled" data, in which the response patterns were randomly reassigned to stimulus classes. The shuffling was repeated 100 times, resulting in 100 MI estimates of the bias (MI_{bias}). These MI_{bias} estimates are then used as estimators for the computation of the statistical significance of the MI estimate for the real (unshuffled) datasets: the real estimate is considered as significant if its value is statistically different from the distribution of MI_{bias} shuffled estimates. Significant MI estimates were computed for MI calculated from neuronal responses under one electrode. The range of MI_{bias} values was very similar between auditory structures: depending on the conditions (original, vocoded, noisy vocalizations), it was from 0.102 to 0.107 in the CN, from 0.107 to 0.110 in the IC, from 0.105 to 0.114 in the MGB, 0.107 to 0.111 in A1 and from 0.106 to 0.116 in VRB. There was no significant difference between the mean MI_{bias} values in the different structures (unpaired t-test, all $p>0.25$).

The information carried by a group of recordings was estimated by the population MI ($MI_{Population}$), using the same method described above: 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 ($MI_{Population}$), the number of sites used for computing $MI_{Population}$ varied from 2 to the maximal possible size (which is equal to 16 minus the non-responsive sites). As the number of possible combinations could be extremely large (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: when the number of possible combinations exceeded one hundred, 100 combinations were randomly chosen, and the mean of all combinations was taken as the $MI_{Population}$ for this group size.

For the $MI_{Population}$, the values of bias were also computed: on average and for all sets of 9 simultaneous recordings, it was 0.104 in the CN, 0.111 in the IC, 0.114 in the MGB, 0.107 in

AI and 0.106 in VRB. There was no significant difference between the mean $MI_{Population}$ 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 analysis of variance (ANOVA) for multiple factors to analyze the whole data set. Post-hoc pair-wise 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 below 0.05. All data are presented as mean values ± standard error (s.e.m.).

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 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 MGv, 354 recordings in A1 and 95 recordings in VRB. Table 1 summarizes the range of best frequencies, mean bandwidth, response duration and response strength obtained when testing pure tone responses in each auditory structure. In the following sections, the neuronal responses to the original vocalizations presented in quiet are compared across brain structures and the discrimination performance are described at the individual and population levels. The neuronal discrimination tested with tone-vocoded vocalizations and vocalizations presented against different levels of masking noise are described and compared next.

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 if selecting a particular value for the Gaussian window influenced the between-structure differences in CorrCoef mean values. Based upon 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 (dashed line in Fig. 2A) as it was used in previous cortical studies (Huetz et al., 2009; Gaucher et al., 2013a; Gaucher and Edeline, 2015; Aushana et al., 2018).

Second, at the cortical level, it was previously showed that the maximal value of mutual information (MI) based on temporal patterns was obtained, on average, with a bin size of 8ms (Schnupp et al., 2006; Gaucher et al., 2013a). However, it has never been demonstrated that the same bin size was optimal at all levels of the auditory system. Figure 2B shows the

463 evolution of MI as a function of temporal precision ranging from 1 to 40 ms based on the
464 responses to the original vocalizations. In our experimental conditions, and with our set of
465 acoustic stimuli, the 8-ms temporal precision was found to be optimal for all auditory
466 structures, in the original (dashed line in Fig. 2B), vocoded and noisy conditions (data not
467 shown). Therefore, the MI value obtained for a temporal precision of 8 ms was subsequently
468 used in our analyses.
469
470 **Subcortical structures better discriminate the original vocalizations**
471
472 Figure 3A displays neuronal responses of two simultaneous multiunit recordings obtained at
473 five levels of the auditory pathway (CN, CNIC, MGv, AI and VRB). The neuronal responses
474 were strong and sustained in the CN and CNIC, more transient in MGv, phasic in AI and
475 more diffuse in VRB. For most of the recordings, temporal patterns of response were clearly
476 reproducible from trial-to-trial, but they differed from one vocalization to another both at the
477 cortical and subcortical level. The PSTHs displayed in figure 3B show that at each level of the
478 auditory system, the four whistles triggered distinct temporal patterns of responses.
479 Quantifications of evoked responses to original vocalizations over all the recordings are
480 presented on figures 3C-F for each auditory structure. These analyses clearly pointed out large
481 differences between the mean values of evoked firing rate, CorrCoef and MI quantified at the
482 cortical vs. at the subcortical level. First, the evoked firing rate was significantly higher in the
483 subcortical structures than in the cortex (unpaired t-test, lowest p value $p<0.001$). It was also
484 higher in CN compared to the other subcortical structures (Fig. 3C). Second, the CorrCoef
485 values were significantly higher in CNIC and MGv compared to AI and VRB (Fig. 3D),
486 indicating that the trial-to-trial reliability of evoked responses was stronger in these structures
487 than in CN, A1 and VRB. Third, the $MI_{Individual}$ values obtained at the subcortical level were
488 significantly higher than at the cortical level (unpaired t-test, highest $p<0.001$ between the
489 cortex and the other structures; Fig. 3E). At the subcortical level, the $MI_{Individual}$ values were
490 significantly higher in MGv than in CNIC and in CN (unpaired t-test, $p<0.01$) with the CN
491 exhibiting the lowest MI values at the subcortical level. The $MI_{Individual}$ values were also
492 significantly lower in VRB than in AI ($p = 0.037$). Recordings in MGv displayed the highest
493 $MI_{Individual}$ mean values, suggesting that, on average, thalamic neurons discriminate better the
494 four original whistles than the other auditory structures. As shown in figure 3G, in each
495 auditory structure, high $MI_{Individual}$ values were strongly correlated with high values of trial-to-
496 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 two to sixteen simultaneous multiunit recordings to determine whether the discrimination performance of neural networks confirm the results obtained at the individual (i.e., single recording) level. $MI_{Population}$ quantifies how well the four whistles can be discriminated based on temporal patterns expressed by neuronal populations distributed on the tonotopic map. The $MI_{Population}$ computed from 9 simultaneous multiunit recordings shows that neural populations in subcortical structures discriminate the four original whistles better than the cortical populations (unpaired t-test, highest p value $p<0.002$ between CN and VRB) without any statistical difference between the three subcortical structures (Fig. 3F).

We next investigated the diversity of the $MI_{Individual}$ and $MI_{Population}$ values obtained in the different structures. The distributions of $MI_{Individual}$ values were plotted as a function of temporal precision for each structure (Fig. 4 A1-A5). It showed that whatever the temporal precision, there were more curves with high $MI_{Individual}$ values in the subcortical structures than in the cortical areas (see red curves on Fig. 4 A1-A5). The examination of the evolution of the $MI_{Population}$ 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 of such curves in AI and VRB whatever the number of recordings considered (Fig. 4 B1-B5).

With a temporal resolution of 8 ms, we presented the cumulative percentages of neurons for the $MI_{Individual}$ (Fig. 5A) and the $MI_{Population}$ values (Fig. 5B) in each structure. Above a value of 1.5 bits (indicating that at least 3 stimuli can be discriminated), there were 39% of MGv neurons, 18% and 14% of the neurons in CNIC and CN respectively; but only 3.5% and 2% of the neurons in A1 and VRB respectively. 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 ones (all p values <0.01). The same conclusions were reached for the $MI_{Population}$ values: More than 90% of the MGv neuronal populations were above 1.5 bits, 83 % and 75% of the populations in CNIC and CN respectively, whereas these populations represented less than 40% at the cortical level (36 % and 34% in A1 and VRB respectively).

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 to 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 significant decrease in evoked firing rate in response to the 10-band vocoded vocalizations was only found at the subcortical level (for all subcortical structures, one-way ANOVA: $F_{CN(3,1995)}=22.6$; $F_{CNIC(3,1543)}=8.85$; $F_{MGv(3,1047)}=6.55$, $p<0.001$; with post-hoc paired t tests, $p<0.05$), whereas there was no decrease in either AI or VRB. Vocoding also decreased the CorrCoef mean values in every structure except in VRB (Fig. 6D). This decrease was significant with the 10-band vocoded vocalizations in CN, MGv and in AI (one-way ANOVA: $F_{CN(3,1930)}=26.48$; $F_{MGv(3,889)}=7.7$; $F_{A1(3,1125)}=3.42$, 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: $F_{(3,1391)}=26.19$, $p<0.001$; with post-hoc paired t tests, $p<0.05$).

Similarly, vocoding decreased the $MI_{Individual}$ 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 AI (one-way ANOVA: $F_{CN(3,1445)}=12.23$, $F_{MGv(3,810)}=3.75$, $F_{A1(3,720)}=3.59$, 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: $F_{CNIC(3,1231)}=13.17$, $p<0.001$; with post-hoc paired t tests, $p<0.05$). At the population level ($MI_{Population}$), compared to the values obtained in response to the original vocalizations, the $MI_{Population}$ values computed with the 10-band vocoded vocalizations were significantly lower in the subcortical structures (one-way ANOVA: $F_{CN(3,127)}=6.46$, $F_{CNIC(3,115)}=6.28$, $F_{MGv(3,67)}=4.62$, 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 $MI_{Population}$ as a function of the number of simultaneous multiunit recordings (Fig. 7 A-E) showed that in each subcortical structure, the curves rapidly reached high $MI_{Population}$ values

(close to the maximal value of 2 bits) in each vocoding conditions, whereas in AI and VRB the curves slowly reached the maximum $MI_{Population}$ 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 neurons 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 (AI 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: $F_{CN(3,1995)}=309.33$, $F_{CNIC(3,1543)}=220.64$, $F_{MGv(3,1047)}=155.07$, $F_{A1(3,1415)}=96.27$, $p<0.001$; 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: $F_{CNIC(3,1357)}=154.12$, $p<0.001$; with post-hoc paired t tests, $p<0.05$). At the cortical level, the CorrCoef values were significantly reduced in AI 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 $MI_{Individual}$ values but again, there was a marked difference between the CNIC and the other subcortical structures: the $MI_{Individual}$ 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

599 $MI_{Individual}$ value in the CNIC was only significantly reduced at the 0 dB SNR (one-way
600 ANOVA: $F_{(3,1078)}=32.08$, $p<0.001$; with post-hoc paired t tests, $p<0.05$). At the cortical level,
601 noise significantly reduced the average $MI_{Individual}$ in AI only at the -10 dB SNR (one-way
602 ANOVA: $F_{(3,649)}=9.49$, $p<0.001$; with post-hoc paired t tests, $p<0.05$) whereas the average
603 $MI_{Individual}$ in VRB remained unchanged (Fig 8E).
604 The effects of masking noise on the network discrimination performance were quantified with
605 the $MI_{Population}$ (Fig. 8F). At the cortical level, there was a significant reduction of $MI_{Population}$
606 values only at the -10 dB SNR (one-way ANOVA: $F_{A1(3,111)}=16.63$, $F_{VRB(3,23)}=11.41$, $p<0.001$;
607 with post-hoc paired t tests, $p<0.05$) whereas there was a significant decrease in CN already at
608 the +10 dB SNR (one-way ANOVA: $F_{CN(3,127)}=51.49$, $p<0.001$; with post-hoc paired t tests,
609 $p<0.05$). In MGv and CNIC, neuronal populations still displayed the highest discrimination
610 performance although the decrease in $MI_{Population}$ value was significant at the 0 dB SNR (one-
611 way ANOVA: $F_{MGv(3,67)}=41.59$, $F_{CNIC(3,115)}=22.59$, $p<0.001$; with post-hoc paired t tests,
612 $p<0.05$).
613 Note that, in VRB, the CorrCoef and $MI_{Population}$ were much more decreased in the noise
614 conditions than in the vocoding conditions, suggesting that the lack of significant decreases in
615 vocoding conditions was not a “floor effect” due to the low initial values.
616 The evolution of the $MI_{Population}$ as a function of the number of simultaneous multiunit
617 recordings in the different structures (Fig. 9A-E) revealed that regardless of the number of
618 neurons considered, noise effects were similar up to the 0 dB SNR: the population curves in
619 CNIC and MGv grew up relatively rapidly and reached higher values than the curves obtained
620 in CN and in the two cortical areas. At the -10 dB SNR, the $MI_{Population}$ from the CNIC
621 remained higher (regardless of the number of neurons considered) than in the other structures;
622 whereas there was no increase of the $MI_{Population}$ with the number of neurons in VRB.
623 One puzzling result came from the fact that on average, the values of $MI_{Individual}$ and
624 $MI_{Population}$ decreased more for CN recordings than for the two subsequent subcortical relays.
625 However, at least 20% of the CN recordings at the +10 dB SNR maintained $MI_{Individual}$ values
626 above 1 bit (Fig. 10A, red curves) and $MI_{Population}$ values above 1.5 bits (Fig. 10C, red curves),
627 suggesting that a sub-population of CN neurons were still able to send information about the
628 vocalization identity at higher brainstem centers. This also suggests that the discrimination
629 performed by a group of a fixed number of neurons deteriorates with noise faster in the CN
630 and consequently, more CN neurons are necessary to obtain an equivalent amount of
631 information observed in CNIC.

The distributions of the TFRP parameters (best frequency, bandwidth, response duration, response strength) from this specific sub-population of CN neurons did not differ from the neurons exhibiting $MI_{Individual}$ values below 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-square tests; $p<0.05$, Fig. 10B). More precisely, the CN recordings exhibiting higher $MI_{Individual}$ values at +10 dB SNR had longer duration responses and stronger evoked firing rate to pure tones.

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

To summarize, masking noise differently impacted the neurons' discrimination performance 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 spectro-temporal degradations: it reduced the AM cues in the different audio-frequency bands, introduced irrelevant envelope fluctuations and altered the temporal fine structure (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, 1995; Shannon et al., 1995; Dubbelboer and Houtgast, 2007; Jorgensen 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 ($MI_{Population}$) in the five structures.

The AM spectra obtained in vocoding and noise conditions showed that the AM cues were attenuated compared to 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 dB 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 ($MI_{Population}$) in the five structures for each experimental condition (Fig. 12B). More precisely, for all conditions, figure 12B shows the changes in $MI_{Population}$ 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, $MI_{Population}$ is barely affected as long as the reduction of the AM index (Δmodulation index) remains lower than 25%; beyond this limit, the $MI_{Population}$ is reduced (i.e., at the 0 dB and -10 dB SNR). The straightforward conclusion is that the reduction of slow AM cues is one of the factors controlling the decrease in $MI_{Population}$ at the cortical and subcortical levels. In the cochlear nucleus, the decrease on the $MI_{Population}$ is much larger than in the other structures, suggesting that the alteration of AM cues has more impact on the $MI_{Population}$ at the most peripheral level. Alternatively, one should keep in mind that the neuronal discrimination in noise can be based upon 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 $MI_{Population}$ 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 parameter 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 experiments should also be generalized with other categories of guinea pig calls, other types of communication sounds from other species and should 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, comparison of neural data collected in response to noisy versus vocoded vocalizations suggests that the transmission of slow (< 20 Hz) amplitude modulation information is one of the factors contributing to the neural discrimination decrease in noise at the cortical and subcortical levels.

Subcortical structures represent natural vocalizations more precisely than primary and non-primary 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 to cortical ones were sufficient to discriminate between the stimuli used in this study. These results corroborate the finding by Chechik and colleagues (2006) that the MGB and AI 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 and colleagues (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 (Fig. 3A and 5A). The stimulus sets also differ, as we used four utterances of the same category (the Whistle), whereas Chechik and colleagues (2006) used three birds' chirps and variants of these stimuli leading potentially to an easier classification between groups of stimuli

compared to 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 the original vocalizations than the cortical areas (Fig. 2B). Potentially, the optimal bin size depends more upon the stimuli durations than upon the auditory structure. When computing mutual information from IC, MGB and A1 neuronal responses, Chechik and colleagues (2006) usually found an optimal bin size of 4 ms, different from ours, probably because their stimulus durations are shorter than our stimuli (67-111 ms vs. to 280-363 ms here). Recently, we also found shorter optimal bin size 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 neurons' ability 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., 1990; Preuss and Müller-Preuss, 1990) and cortical neurons for even lower rates (Gaese and Ostwald, 1995; Schreiner and Urbas, 1998). 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 (see the peak of the black curve in AM spectra, Fig. 12A). 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., 2014b).

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 AI, studies have

767 pointed out the relationships between the noise impact on the cortical and behavioral
768 discrimination performance. In bird field L (homologous to AI), neuronal responses to song
769 motifs were strongly reduced by three types of masking noises, and the neural discrimination
770 performance was progressively reduced when the SNR decreased, in parallel with the
771 behavioral performance (Narayan et al., 2007). Our VRB results are reminiscent of those
772 obtained in area NCM (homologous to a secondary area) where feed-forward inhibition
773 allowed the emergence of invariant neural representations of target songs in noise conditions
774 (Schneider and Woolley, 2013). Similarly to the results by Ranasinghe and colleagues (2012),
775 our IC neuronal responses were found to be resistant to drastic spectral degradations.
776 Only one study directly compared the impact of vocoding and masking noise on cortical
777 responses to vocalizations (Nagarajan et al., 2002). In this study, auditory cortex responses
778 were robust to spectral degradations even in response to 2-band vocoded vocalizations. Also,
779 broadband white noise reduced neuronal responses at 0 dB SNR. Last, temporal-envelope
780 degradations strongly reduced the evoked firing rate and the neural synchronization to the
781 vocalization envelope. Importantly, band-pass filtering the vocalizations between 2-30 Hz did
782 not reduce firing rate and neural synchronization to the vocalization envelope. This is in
783 agreement with the results in our conditions: when the Δmodulation index (computed between
784 1 and 20 Hz) revealed modest AM alterations, there was little effect on the neuronal
785 discrimination, but when the AM alterations reached about 20-30% or more, the neuronal
786 discriminations were reduced (Fig. 12B). Thus, our results are consistent with the hypothesis
787 that one of the factors constraining auditory discrimination at the cortical and subcortical level
788 is the fidelity of transmission and processing of slow AM cues.

789 When quantifying how different noise levels alter neuronal coding in the auditory system, it
790 was found that the neural representation of natural sounds becomes progressively independent
791 of the level of background noise from the auditory nerve to the IC and AI (Rabinowitz et al.,
792 2013). It was proposed that this tolerance to background noise results from an adaptation to
793 the noise statistics, which is more pronounced at the cortical than at the subcortical level. In
794 agreement with this study, we found that populations of cortical neurons (AI and VRB) were
795 more resistant to noise than subcortical ones. However, we did not observe a monotonic
796 evolution of resistance to noise in the auditory system: at the subcortical level, the
797 discrimination performance of CN neuronal populations drastically dropped as early as +10
798 dB SNR, the populations of CNIC neurons maintained the highest discrimination performance
799 even at the -10 dB SNR, those of thalamic ones 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 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%, Table 1 in Huetz et al., 2009). In addition, the Hmax value (equivalent of MI) was higher in anesthetized than in awake animals (0.38 vs. 0.24, Table 2 in Huetz et al., 2009). 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; Fig. 8 in Huetz et al., 2009). 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 to 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., 2007), 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 remained 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 four same 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 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 (Drullman et al., 1994; Shannon et al., 1995), other factors (the pitch, the frequency modulation, the harmonicity cues) can also be involved.

As our results are based on multiunit recordings, we do not know whether the same number of neurons were present in the cluster recordings from the different structures, and whether the individual discrimination performance 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., 2014b). As already proposed by Chechick and Nelken (2012), auditory cortex neurons extract abstract auditory entities rather than detailed spectro-temporal features. Obviously, this urges to define 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 (Gygi and Shafiro, 2013; Gygi et al., 2004, 2007; Woolley et al., 2005; Singh and Theunissen, 2003) 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 to subcortical ones 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% see 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)

868 influence the neural discrimination at the subcortical level, especially in awake, behaving,
869 animals. Thus, we can envision that in behaving animals, learning-induced cortical plasticity
870 also contributes to enhancing the subcortical neural discrimination via the corticofugal
871 projections. Further studies are required to determine to what extent these subcortical
872 representations influence auditory abilities in animals and humans.
873

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.

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.

Ayala YA, Pérez-González D, Malmierca MS (2016) Stimulus-specific adaptation in the inferior colliculus: The role of excitatory, inhibitory and modulatory inputs. Biological Psychology 116, 10–22.

Biberger T, Ewert SD (2017) The role of short-time intensity and envelope power for speech intelligibility and psychoacoustic masking. J Acoust Soc Am. 142(2):1098.

Bizley JK, Walker KM, Silverman BW, King A J, Schnupp JWH (2009) Interdependent Encoding of Pitch, Timbre, and Spatial Location in Auditory Cortex. Journal of Neuroscience 29, 2064–2075.

Carruthers IM et al. (2015) Emergence of invariant representation of vocalizations in the auditory cortex. Journal of Neurophysiology 114, 2726–2740.

Chechik G, Nelken I (2012) Auditory abstraction from spectro- temporal features to coding auditory entities. Proc Natl Acad Sci U S A 109(46):18968–18973.

Chechik, G. et al. (2006) Reduction of Information Redundancy in the Ascending Auditory Pathway. Neuron 51, 359–368.

Creutzfeldt O, Hellweg FC, Schreiner C (1980) Thalamocortical transformation of responses to complex auditory stimuli. Experimental Brain Research 39.

Drullman, R. Speech intelligibility in noise: Relative contribution of speech elements above and below the noise level. *The Journal of the Acoustical Society of America* **98**, 1796–1798 (1995).

Drullman, R., Festen, J. M. & Plomp, R. Effect of temporal envelope smearing on speech reception. *The Journal of the Acoustical Society of America* **95**, 1053–1064 (1994).

Dubbelboer F, Houtgast TA (2007) A detailed study on the effects of noise on speech intelligibility. The Journal of the Acoustical Society of America 122, 2865.

Edeline JM, Weinberger NM (1993) Receptive field plasticity in the auditory cortex during frequency discrimination training: selective retuning independent of task difficulty. Behav. Neurosci. 107, 82–103.

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.

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, 934-953.

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.

Escabí MA, Read HL (2005) Neural Mechanisms for Spectral Analysis in the Auditory Midbrain, Thalamus, and Cortex. in International Review of Neurobiology 70, 207–252.

Ewert SD, Dau T (2000) Characterizing frequency selectivity for envelope fluctuations. J Acoust Soc Am. 108(3 Pt 1):1181-96.

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.

Frisina RD, Smith RL, Chamberlain SC (1990) Encoding of amplitude modulation in the gerbil cochlear nucleus: I. A hierarchy of enhancement. Hear. Res. 44, 99–122.

Gaese BH, Ostwald J (1995) Temporal coding of amplitude and frequency modulation in the rat auditory cortex. Eur. J. Neurosci. 7, 438–450.

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.

Gauche 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.

Gnansia D, Péan V, Meyer B, Lorenzi C (2009) Effects of spectral smearing and temporal fine structure degradation on speech masking release. J. Acoust. Soc. Am. 125, 4023–4033.

Gnansia D, Pressnitzer D, Péan V, Meyer B, Lorenzi C (2010) Intelligibility of interrupted and interleaved speech for normal-hearing listeners and cochlear implantees. Hearing Research 265, 46–53.

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.

Gygi B, Shafiro V (2013) Auditory and Cognitive Effects of Aging on Perception of Environmental Sounds in Natural Auditory Scenes. J Speech Lang Hear Res 56, 1373–1388.

Gygi B, Kidd GR, Watson CS (2007) Similarity and categorization of environmental sounds. Perception & Psychophysics 69, 839–855.

Gygi B, Kidd GR, Watson CS (2004) Spectral-temporal factors in the identification of environmental sounds. The Journal of the Acoustical Society of America 115, 1252–1265.

Hohmann V (2002) Frequency analysis and synthesis using a Gammatone filterbank. Acust Acta Acust 88:433–442.

Houtgast T, Steeneken H (1985) A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. The Journal of the Acoustical Society of America. 77 (3): 1069–1077.

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.

Huetz C, Guédin M, Edeline JM (2014) Neural correlates of moderate hearing loss: Time course of response changes in the primary auditory cortex of awake guinea-pigs Frontiers in Systems Neuroscience, 8, 65.

Jacomme AV et al (2003) The projection from auditory cortex to cochlear nucleus in guinea pigs: an in vivo anatomical and in vitro electrophysiological study. Experimental Brain Research 153, 467–476.

968 Jørgensen S, Dau T (2011) Predicting speech intelligibility based on the signal-to-noise envelope
969 power ratio after modulation-frequency selective processing. J Acoust Soc Am. 2011
970 Sep;130(3):1475-87.
971 Joris PX, Schreiner CE, Rees A (2004) Neural processing of amplitude-modulated sounds. Physiol.
972 Rev. 84, 541–577.
973 Kates JM, Arehart KH (2014) The Hearing-Aid Speech Perception Index (HASPI). Speech
974 Communication 65, 75–93.
975 Kates JM (2011) Spectro-temporal envelope changes caused by temporal fine structure
976 modification. The Journal of the Acoustical Society of America 129, 3981–3990.
977 Las L, Stern EA, Nelken I (2005) Representation of Tone in Fluctuating Maskers in the Ascending
978 Auditory System. Journal of Neuroscience 25, 1503–1513.
979 Lesica NA, Grothe B (2008) Efficient Temporal Processing of Naturalistic Sounds. PLoS ONE 3,
980 e1655.
981 Malmierca MS, Ryugo DK (2011) Descending Connections of Auditory Cortex to the Midbrain
982 and Brain Stem. in The Auditory Cortex (eds. Winer, J. A. & Schreiner, C. E.) 189–208
983 (Springer US, 2011).
984 Malmierca MS (2004) The Inferior Colliculus: A Center for Convergence of Ascending and
985 Descending Auditory Information. Neuroembryology and Aging 3, 215–229.
986 Manunta Y, Edeline JM (1999) Effects of noradrenaline on frequency tuning of auditory cortex
987 neurons during wakefulness and slow-wave sleep. Eur. J. Neurosci. 11, 2134–2150.
988 Mesgarani N, Cheung C, Johnson K, Chang EF (2014b) Phonetic feature encoding in human
989 superior temporal gyrus. Science 343 (6174):1006-10.
990 Moon Il J, Won JH, Park MH, Ives DT, Nie K, Heinz MG, Lorenzi C and Rubinstein JT (2014)
991 Optimal combination of neural temporal envelope and fine structure cues to explain speech
992 identification in background noise. Journal of Neuroscience, 34, 12145-12154.
993 Nagarajan SS et al. (2002) Representation of Spectral and Temporal Envelope of Twitter
994 Vocalizations in Common Marmoset Primary Auditory Cortex. Journal of Neurophysiology
995 87, 1723–1737.
996 Narayan R et al. (2007) Cortical interference effects in the cocktail party problem. Nature
997 Neuroscience 10, 1601–1607.
998 Narayan R, Graña G, Sen K (2006) Distinct Time Scales in Cortical Discrimination of Natural
999 Sounds in Songbirds. Journal of Neurophysiology 96, 252–258.
1000 Nelken I, Rotman Y, Yosef OB (1999) Responses of auditory-cortex neurons to structural features
1001 of natural sounds. Nature 397 154–157.
1002 Nelken I, Bar-Yosef O (2008) Neurons and objects: the case of auditory cortex. Front Neurosci.
1003 2(1):107-13.
1004 Ni R, Bender DA, Shanechi AM, Gamble JR, Barbour DL (2017) Contextual effects of noise on
1005 vocalization encoding in primary auditory cortex. Journal of Neurophysiology 117, 713–727.
1006 Noordhoek IM, Drullman R (1997) Effect of reducing temporal intensity modulations on sentence
1007 intelligibility. J. Acoust. Soc. Am. 101, 498–502.
1008 Occelli F, Suied C, Pressnitzer D, Edeline JM, Gourévitch B (2016) A Neural Substrate for Rapid
1009 Timbre Recognition? Neural and Behavioral Discrimination of Very Brief Acoustic Vowels.
1010 Cereb. Cortex 26, 2483–2496.
1011 Paraouty N, Stasiak A, Lorenzi C, Varnet L, Winter IM (2018) Dual Coding of Frequency
1012 Modulation in the Ventral Cochlear Nucleus. J. Neurosci. 38, 4123–4137.
1013 Patterson RD (1987) A pulse ribbon model of monaural phase perception. J. Acoust. Soc. Am. 82,
1014 1560–1586.

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.

Preuss A, Müller-Preuss P (1990). Processing of amplitude modulated sounds in the medial geniculate body of squirrel monkeys. Exp Brain Res 79, 207–211.

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 Biology 11, e1001710.

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

Ranasinghe KG, Vrana WA, Matney CJ, Kilgard MP (2013) Increasing diversity of neural responses to speech sounds across the central auditory pathway. Neuroscience 252, 80–97.

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.

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

Royer J, Occelli F, Huetz C, Edeline JM, Cancela JM (2019) Are auditory cortex neurons better in discriminating communication sounds in mother vs. in virgin mice? An electrophysiological study in C57BL/6 mice. 42th Midwinter Meeting of the Association for Research in Otolaryngology, Baltimore, USA.

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.

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

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J. & Ekelid, M. Speech Recognition with Primarily Temporal Cues. *Science* **270**, 303–304 (1995).

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

Schnupp JWH, Hall TM, Kokelaar RF, Ahmed B (2006) Plasticity of temporal pattern codes for vocalization stimuli in primary auditory cortex. J. Neurosci. 26, 4785–4795.

Schreiner CE, Urbas JV (1988) Representation of amplitude modulation in the auditory cortex of the cat. II. Comparison between cortical fields. Hear. Res. 32, 49–63.

Shamma S, Lorenzi C (2013) On the balance of envelope and temporal fine structure in the encoding of speech in the early auditory system. J Acoust Soc Am 133(5):2818–2833.

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

Singh N. C, Theunissen FE (2003) Modulation spectra of natural sounds and ethological theories of auditory processing. The Journal of the Acoustical Society of America 114, 3394–3411.

Stone MA, Füllgrabe C, Mackinnon RC, Moore BCJ (2011) The importance for speech intelligibility of random fluctuations in ‘steady’ background noise. J. Acoust. Soc. Am. 130, 2874–2881.

Ter-Mikaelian M, Semple MN, Sanes DH (2013) Effects of spectral and temporal disruption on cortical encoding of gerbil vocalizations. Journal of Neurophysiology 110, 1190–1204.

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

Varnet L, Ortiz-Barajas MC, Erra RG, Gervain J, Lorenzi C (2017) A cross-linguistic study of speech modulation spectra. The Journal of the Acoustical Society of America 142, 1976–1989.

Verhey JL, Pressnitzer D, Winter IM. (2003) The psychophysics and physiology of comodulation masking release. Exp Brain Res. Dec;153(4):405-17.

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

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

Wang X, Kadia SC (2001) Differential Representation of Species-Specific Primate Vocalizations in the Auditory Cortices of Marmoset and Cat. Journal of Neurophysiology 86, 2616–2620.

Wang X, Merzenich MM, Beitel R, Schreiner CE (1995) Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. Journal of Neurophysiology 74, 2685–2706.

Wirtzfeld MR, Ibrahim RA, Bruce IC. (2017) Predictions of Speech Chimaera Intelligibility Using Auditory Nerve Mean-Rate and Spike-Timing Neural Cues. J Assoc Res Otolaryngol. 18(5):687-710.

Woolley SM, Fremouw TE, Hsu A, Theunissen FE (2005) Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. Nat Neurosci. 8, 1371-9.

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. USA*, *102*, 2293-2298.

	CN	Lemniscal pathway			Non-lemniscal pathway
		CNIC	MGv	A1	VRB
Number of animals	10	11	10	11	5
Number of recordings tested	672	478	448	544	192
TFRP only	560	421	285	455	126
TFRP and significant response to at least one vocalization	499	386	262	354	95
TFRP quantifications					
BF Range (kHz): min-max	0.18 - 18	0.34 - 36	0.33 - 33	0.14 - 36	0.67 - 36
Mean bandwidth (octave)	3.91	2.88	4.16	2.07	1.79
Mean response duration (ms)	26.83	35.37	17.31	43.73	44.83
Response strength (AP/sec)	77.23	82.25	41.61	37.69	19.97

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

Figure Legends

Figure 1. Spectrograms, spectra and temporal envelopes of the acoustic stimuli. A-C. Spectrograms (A), spectra (B) and temporal envelopes (C) of the four original whistles used in this study. **D-F.** *From left to right*, spectrograms (D), spectra (E) and temporal envelopes (F) of the four vocoded whistles using 38, 20 and 10 frequency bands. **G-I.** *From left to right*, spectrograms (G), spectra (H) and temporal envelopes (I) of the four original whistles embedded in a vocalization-shaped stationary noise at three SNRs (+10, 0 and -10 dB).

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 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 hold true also in the different conditions of acoustic alterations, both in noise and vocoded conditions (data not shown).

Figure 3. **Subcortical neurons discriminate better the original vocalizations than cortical neurons. A.** *From bottom to top, n*euronal 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 grey part of rasters corresponds to evoked activity. For each example, the values of the best frequency (BF in kHz) and of the bandwidth (BW in octave) obtained when testing the responses to pure tones are indicated in the left side. The waveforms of the four original whistles are displayed under the rasters. **B.** Peristimulus time histograms (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 panels show the mean values of (**C**) the evoked firing rate (spikes/sec), (**D**) the trial-to-trial temporal reliability quantified by the CorrCoef value, (**E**) the neuronal discrimination assessed by the mutual information (MI) computed at the level of the individual recording ($MI_{Individual}$, bits) and (**F**) the neuronal discrimination at the population level ($MI_{Population}$, bits) with populations of 9 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 $MI_{Individual}$ (bits) values obtained in original conditions.

Figure 4. Diversity of neuronal discrimination performance in quiet for each structure at the individual and population level.

A. Neural discrimination performance in response to original vocalizations in each auditory structure. Waterfall plots show the mutual information (MI, *bits*) as a function of temporal resolution (1 to 256 ms) for the selected recordings in CN *(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 from the maximal value of 2 bits) in MGv, CNIC and CN (*red curves*) compared to a much lower proportion in the cortical areas AI and VRB.

B. Population information quickly reaches high values with simultaneous multiunit recordings at the subcortical but not cortical level. For each auditory structure, each thin line represents a particular case of simultaneous recording with a maximum number of electrodes (maximum 16 simultaneous multiunit recordings) and each thick line represents the mean value of $MI_{Population}$ 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 $MI_{Population}$ value quickly reaches high values close from the maximum value of 2 bits in the subcortical structures (CN, CNIC and MGv) compared to the two cortical areas (A1 and VRB).

Figure 5. High discrimination performance neurons are more numerous in subcortical structures than in auditory cortex in original conditions. A. Cumulative percentage of the neuronal discrimination performance obtained in original vocalizations assessed by the mutual information (MI) computed at the level of the individual recordings ($MI_{Individual}$, bits) and (**B**) at the population level ($MI_{Population}$, bits) with populations of 9 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 grey 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 best frequency (BF in kHz) and of the bandwidth (BW in octave) obtained when testing the responses to pure tones are indicated in the left side. For each example, the mean evoked firing rate (spikes/sec) obtained in each condition is indicated below the rasters. **B.** Peristimulus time histograms (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 grey part of the PSTHs corresponds to evoked activity. **C-F.** The mean values (±SEM) represent the vocoding effects on (**C**) the evoked firing rate (spikes/sec), (**D**) the temporal reliability represented by the CorrCoef value, (**E**) the neuronal discrimination assessed by the mutual information (MI) computed at the level of the individual recordings ($MI_{Individual}$, bits) and (**F**) the neuronal discrimination at the population level ($MI_{Population}$, bits) with populations of 9 simultaneous multiunit 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 (200 ms before each acoustic stimulus). At the population level, the discrimination performance significantly decreased only for 10 frequency bands in subcortical structures and did not decrease in cortical areas.

Figure 7. Vocoding effects on the $MI_{Population}$ growth functions in each auditory structure. The curves display the average growth functions of the $MI_{Population}$ for each structure in each vocoding condition (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 each structure, the vocoding slightly reduced the $MI_{Population}$ values. At the cortical level, the reduction induced by vocoding was similar at 38 and 20 bands, then a stronger reduction was observed at 10 bands. At the thalamic level, there was almost no change in the growth function of the $MI_{Population}$ with 38 and 20 bands vocalizations, but there was a large decrease in $MI_{Population}$ with the 10-band vocoded stimuli. In the CNIC, the vocoding only induced a reduction of the $MI_{Population}$ for 20 and 10 bands; a similar scenario was observed at the CN level.

Figure 8. 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 grey 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 best frequency (BF in kHz) and of the bandwidth (BW in octave) obtained when testing the responses to pure tones are indicated in the left side. For each example, the mean evoked firing rate (spikes/sec) 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.** Peristimulus time histograms (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 grey part of the PSTHs corresponds to evoked activity. **C-F.** The mean values (±SEM) represent the noise effects on (**C**) the evoked firing rate (spikes/sec), (**D**) the temporal reliability represented by the CorrCoef value, (**E**) the neuronal discrimination assessed by the mutual information (MI) computed at the level of the individual recordings ($MI_{Individual}$, bits) and (**F**) the neuronal discrimination at the population level ($MI_{Population}$, bits) with populations of 9 simultaneous multiunit 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 (200 ms before each acoustic stimulus). At the population level, the discrimination performance significantly decreased in all structures when the SNR decreased, with on average the CNIC populations still able to discriminate 2 out of 4 stimuli ($MI_{Population}$value >1bit).

Figure 9. Noise effects on the $MI_{Population}$ growth functions in each auditory structure. The curves display the noise effects on the $MI_{Population}$ 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 $MI_{Population}$ in each structure (but to a lesser extent in the CNIC). In the CN, noise induced a stronger reduction of the $MI_{Population}$, which was clearly a function of SNR. In the CNIC, noise induced SNR-dependent reduction in the $MI_{Population}$ values, the reduction being modest at a +10 and 0 dB SNR but more important at a -10 dB SNR. In the MGv, noise progressively lowered the curves of the $MI_{Population}$. In the cortex, the $MI_{Population}$ growth functions were not strongly impacted except at the -10 dB SNR.

Figure 10. A subpopulation of CN neurons maintains good neuronal discrimination performance at a +10 dB SNR.
A. Waterfall plot shows the mutual information ($MI_{Individual}$, bits) as a function of temporal resolution (1 to 256 ms) for the CN recordings at +10 dB SNR. The recordings are ranked by the mean MI value obtained for all bin sizes. Note that at this particular SNR, 20% of the CN recordings still maintained $MI_{Individual}$ values above 1 bit, indicating that some CN neurons still send information about the vocalization identity at higher brainstem centers such as the CNIC. **B.** Distributions of the Time-Frequency Response Profile (TFRP) parameters (best frequency, bandwidth, response duration and response strength) for the two neuronal populations in CN depending of the MI value (*in grey*, MI>=1bit and *in black*, MI<1bit). Note that, there were significant differences in terms of response duration and the response strength. **C.** The curves display the individual and average growth functions of the $MI_{Population}$ for the simultaneous CN recordings at the +10 dB SNR. Note that despite the fact that the mean $MI_{Population}$ value was much lower than in the original condition (see figure 4B1), about 20% of the simultaneously recorded populations reached a value of 1.5 bits with 9 neurons or less (*red curve lines*).

Figure 11. No relationship between the mutual information and the parameters of TFRPs (the best frequency, BF and the bandwidth, BW) at each stage of the auditory system.
A. Typical examples of Time-Frequency Response Profile (TFRP) recorded in VRB, AI, MGv, CNIC and CN. These TFRPs are examples of responses to pure tones and the first column also corresponds to the same neurons as those presented in figures 3, 5 and 7. *From left to right*, the maximal firing rate (in spikes/sec) was 100 and 220 in VRB, 195 and 200 in AI, 460 and 420 in MGv, 315 and 250 in CNIC and 340 and 330 in CN. From these TFRPs, we extracted parameters such as the best frequency (in kHz), the bandwidth (in octave), the response duration (in ms) and the response strength (in spikes/sec). **B**. Noise effect on neuronal discrimination ($MI_{Individual}$, bits) according to the best frequency (BF). Scattergrams of the $MI_{Individual}$ values obtained at the +10 dB SNR as a function of the $MI_{Individual}$ values obtained with the original vocalizations based on neuronal responses recorded in CN, CNIC, MGv, A1 and VRB. We separated the recordings in three groups according to the best frequency: BF< 5kHz (in red), 5<= BF <= 15 kHz (in blue) and BF > 15 kHz (in green). $MI_{Individual}$ mean values are represented with a black cross. **C**. Noise effect on neuronal discrimination ($MI_{Individual}$, bits) according to the bandwidth (BW). Scattergrams of the $MI_{Individual}$ values obtained at the +10 dB SNR as a function of the $MI_{Individual}$ values obtained with the original vocalizations based on neuronal responses recorded in CN, CNIC, MGv, A1 and VRB. We separated the recordings in three groups according to the bandwidth: BW <= 2 octaves (in red), 2 <= BW <= 4 octaves (in blue) and BW >= 4 octaves (in green). $MI_{Individual}$ mean values are represented with a black cross. Note that, in all structures, the decrease in $MI_{Individual}$ value from the original conditions to the +10 dB SNR occurred whatever the BF and the BW values.

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 amplitude-modulation (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 $\Delta MI_{Population}$ as a function of Δmodulation index computed for each structure from mean $MI_{Population}$ or mean modulation-index values obtained in all adverse conditions and mean values in the original condition. Each dot represents neuronal data ($\Delta MI_{Population}$) 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 $\Delta MI_{Population}$ decreases in cortical and subcortical structures.

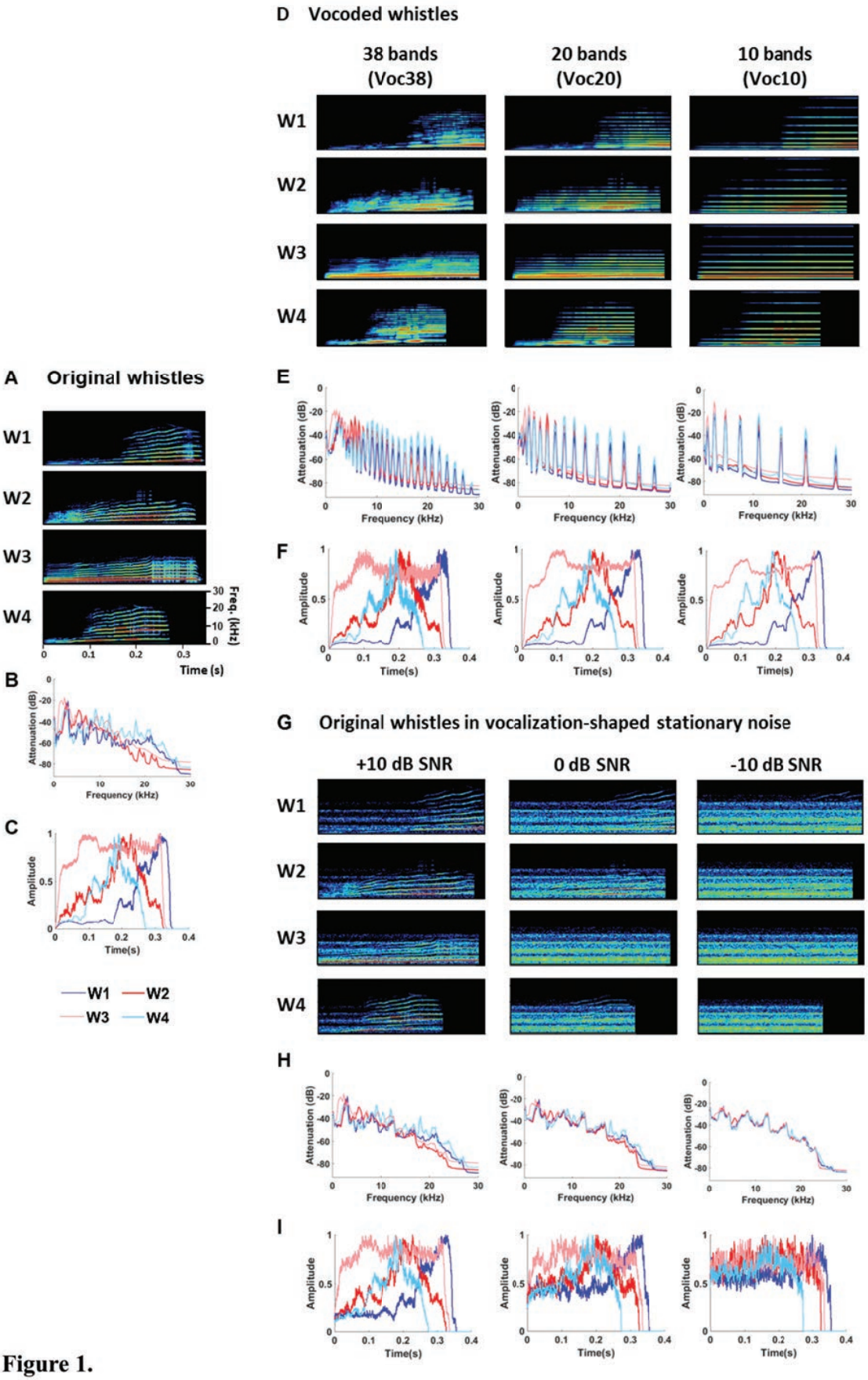

Figure 1.

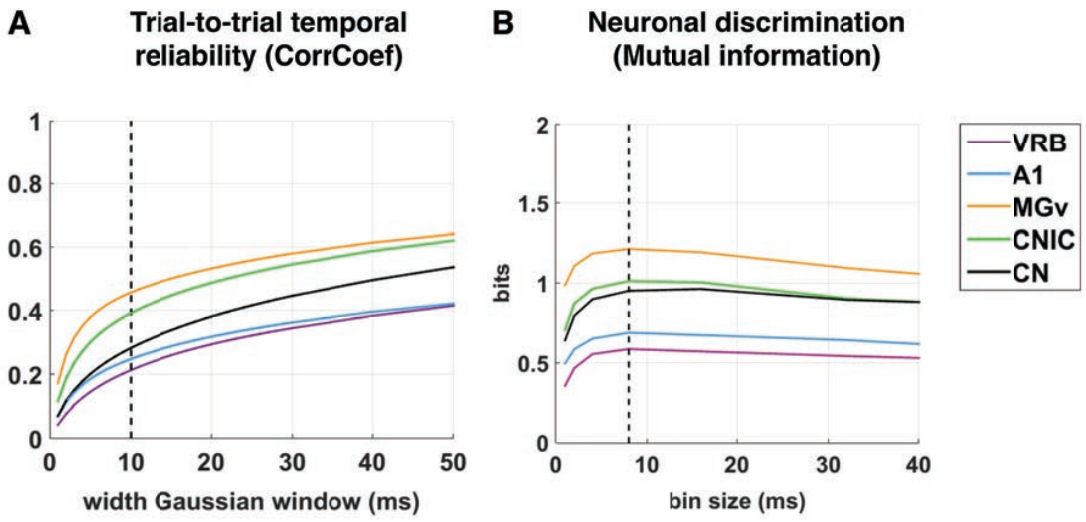

Figure 2.

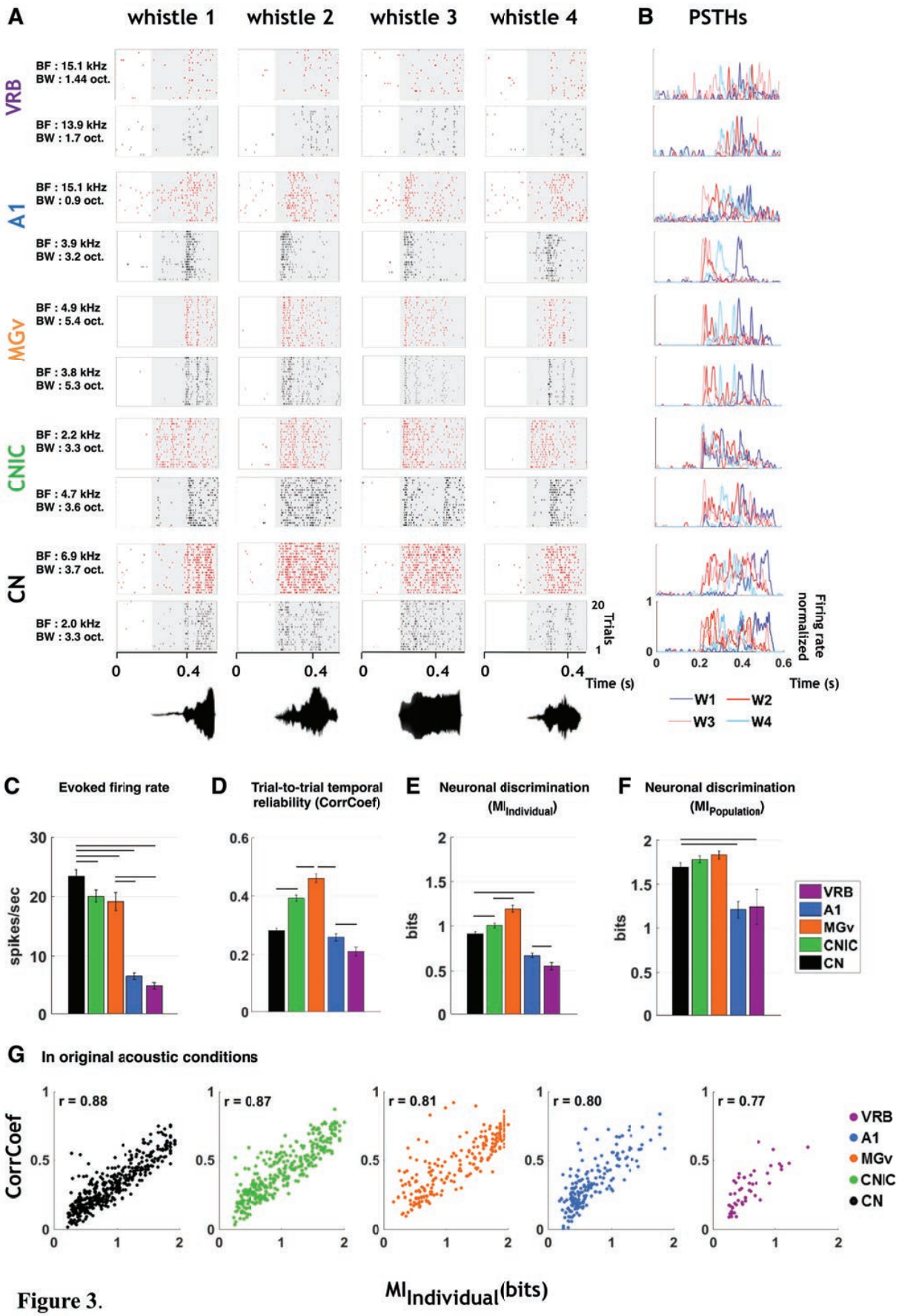

Figure 3.

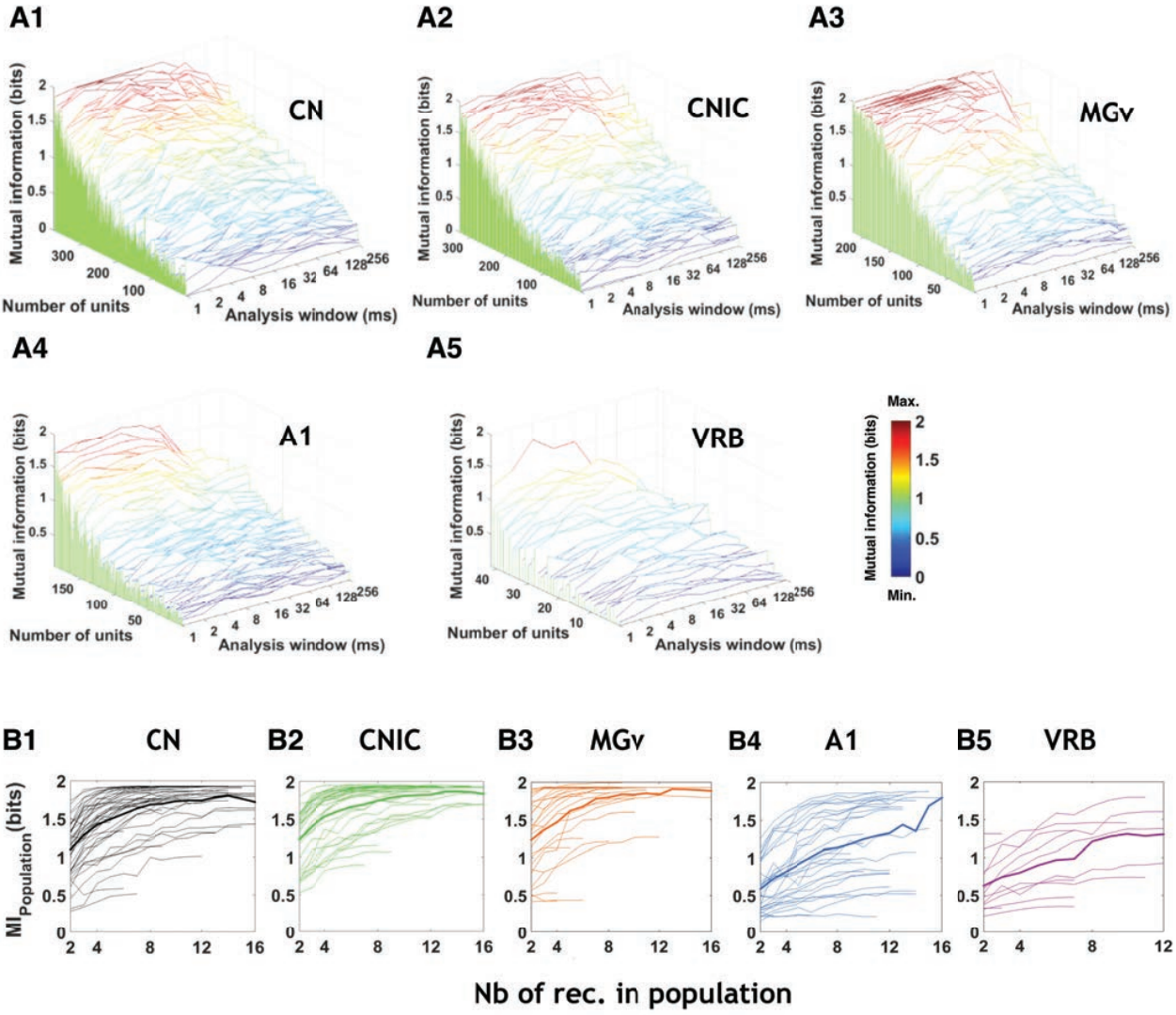

Figure 4.

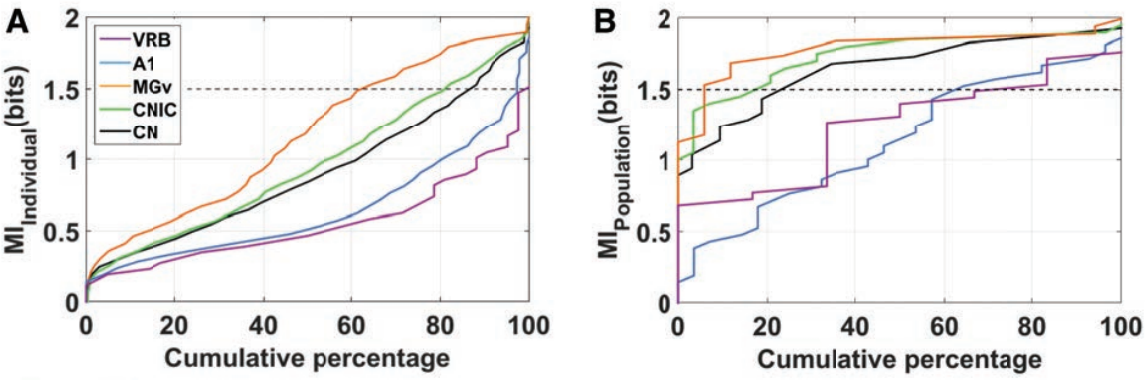

Figure 5.

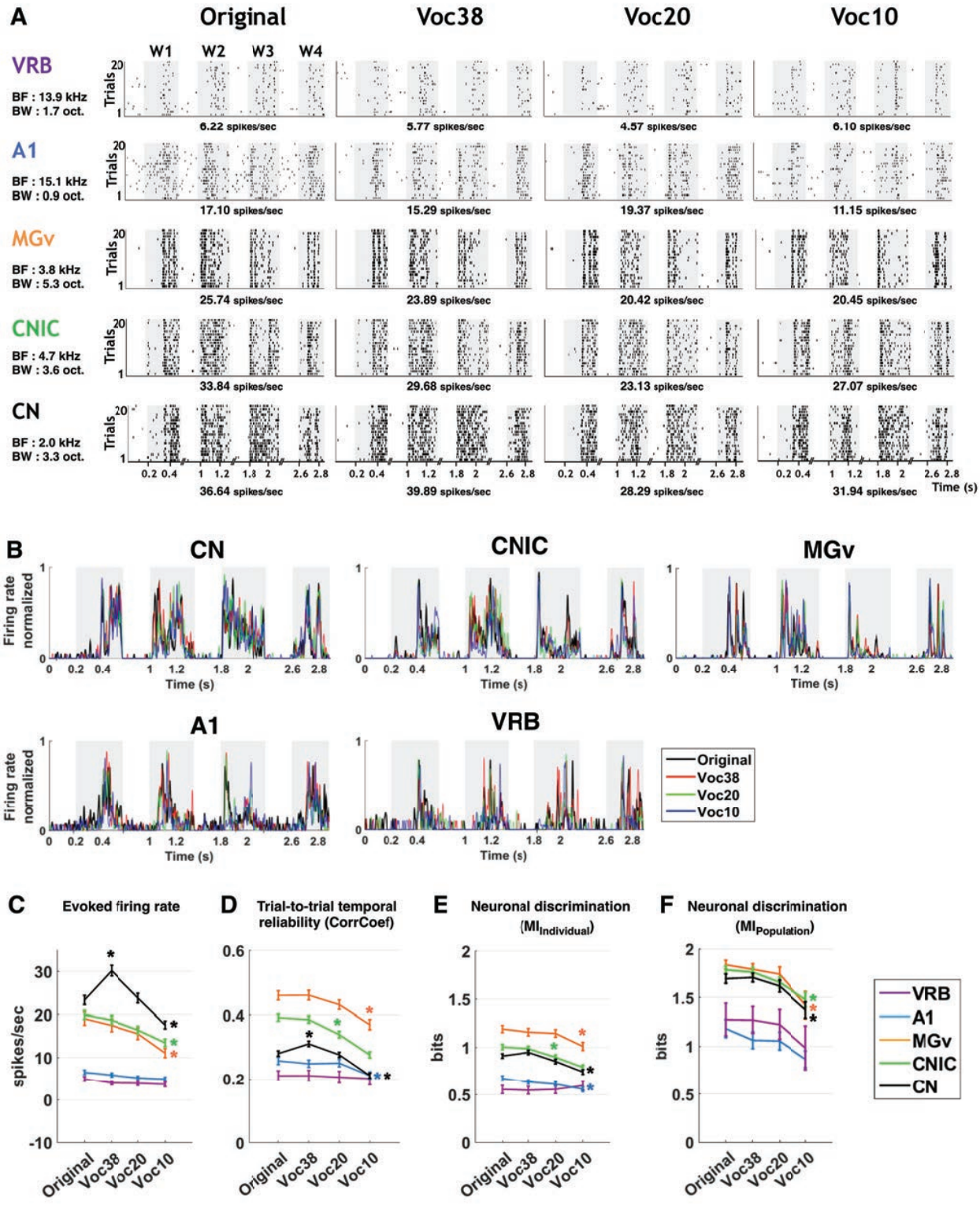

Figure 6.

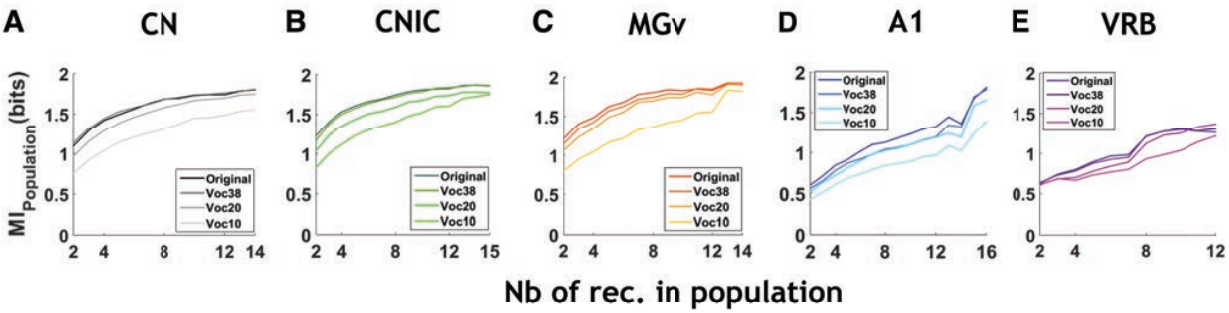

Figure 7.

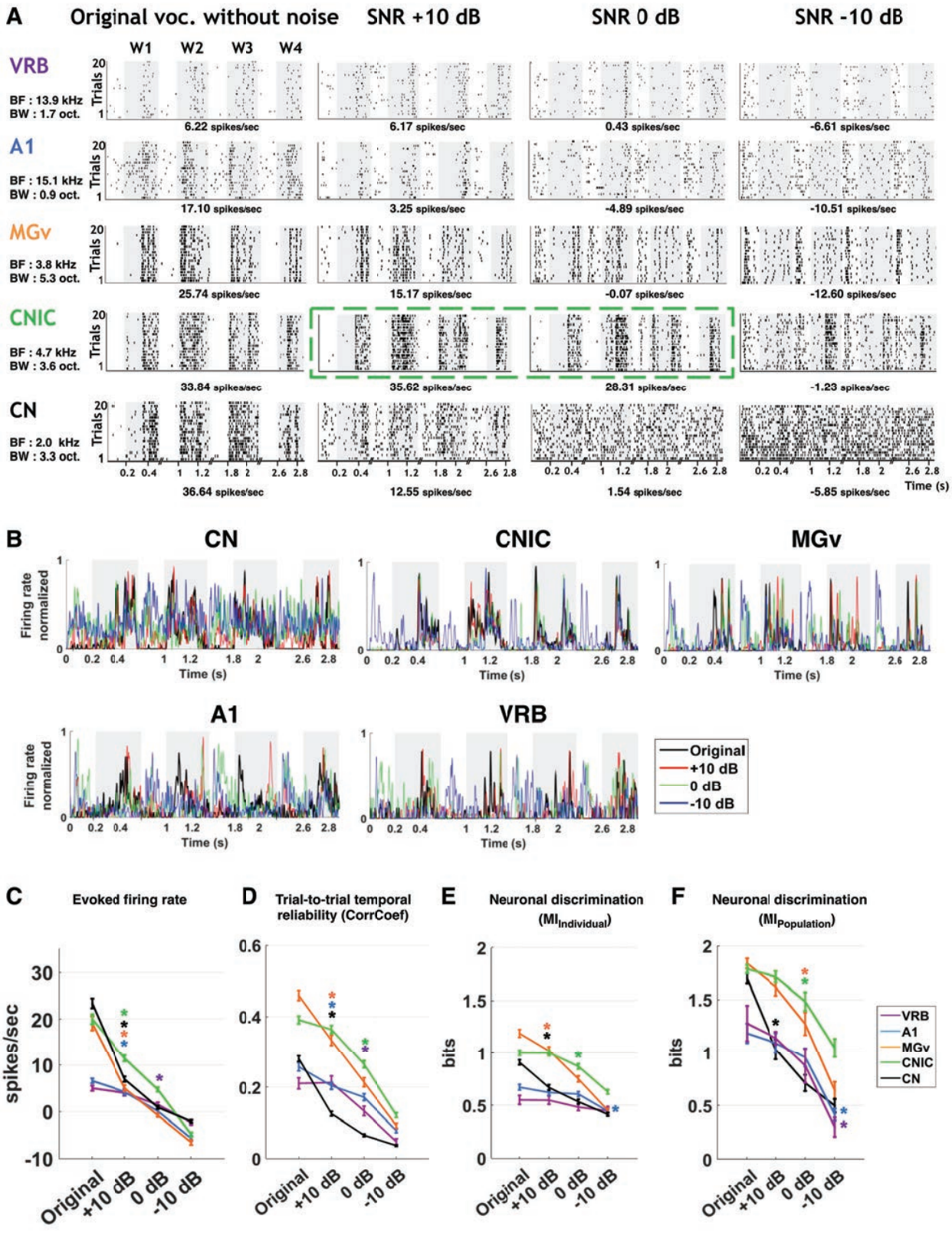

Figure 8.

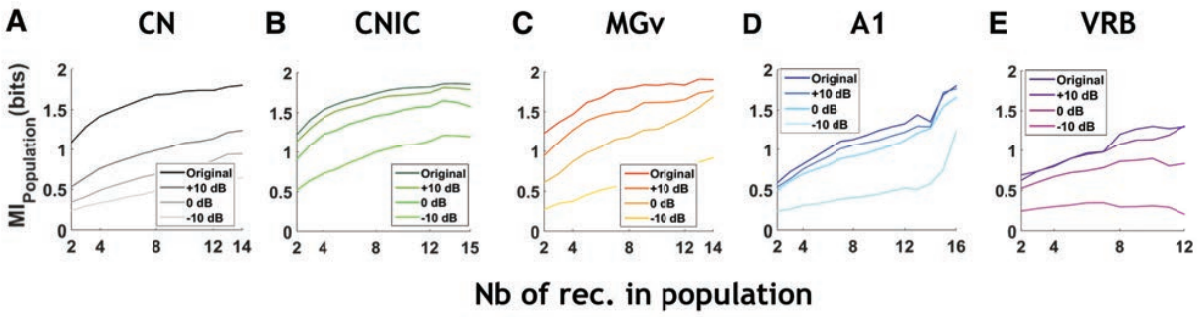

Figure 9.

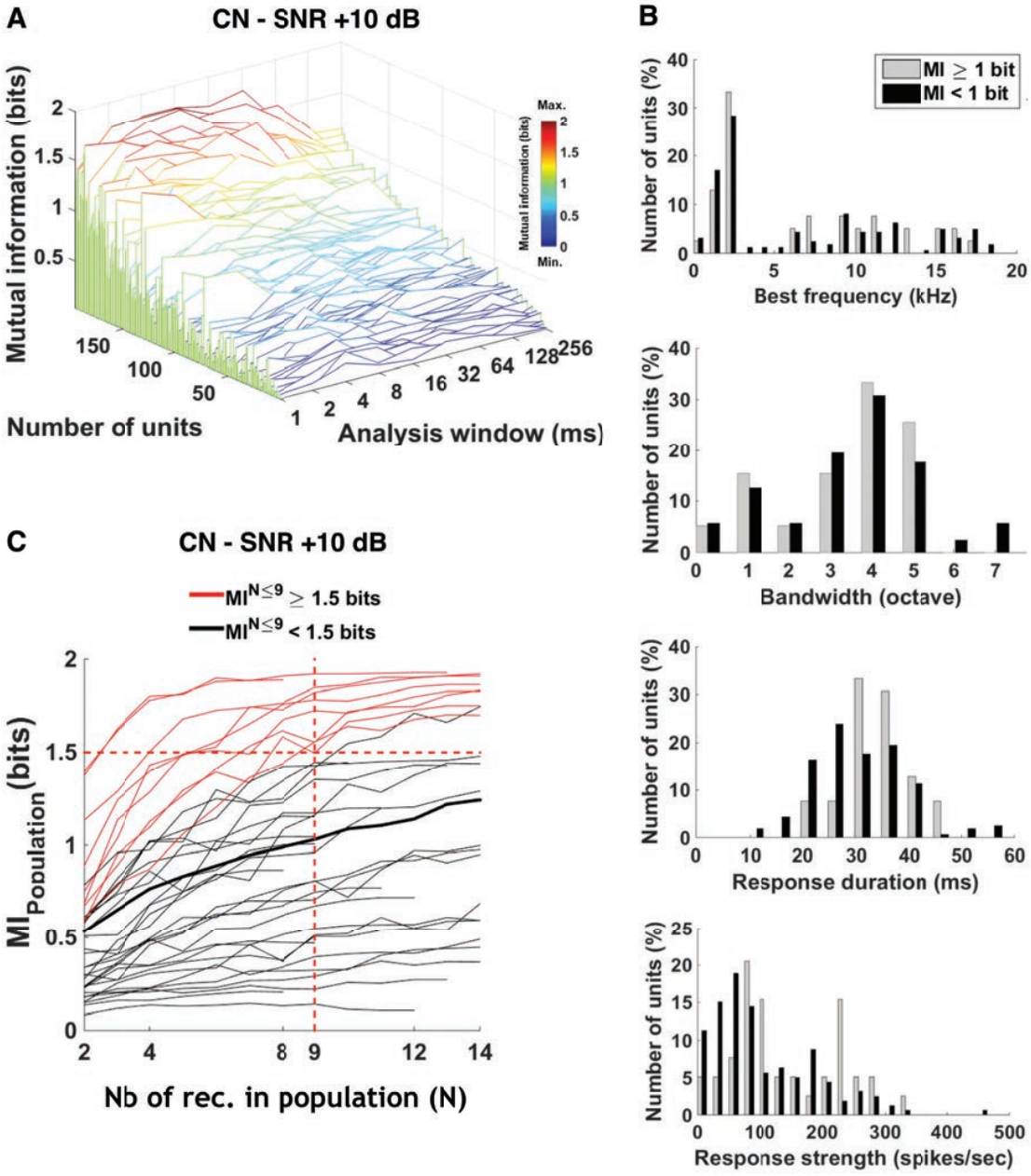

Figure 10.

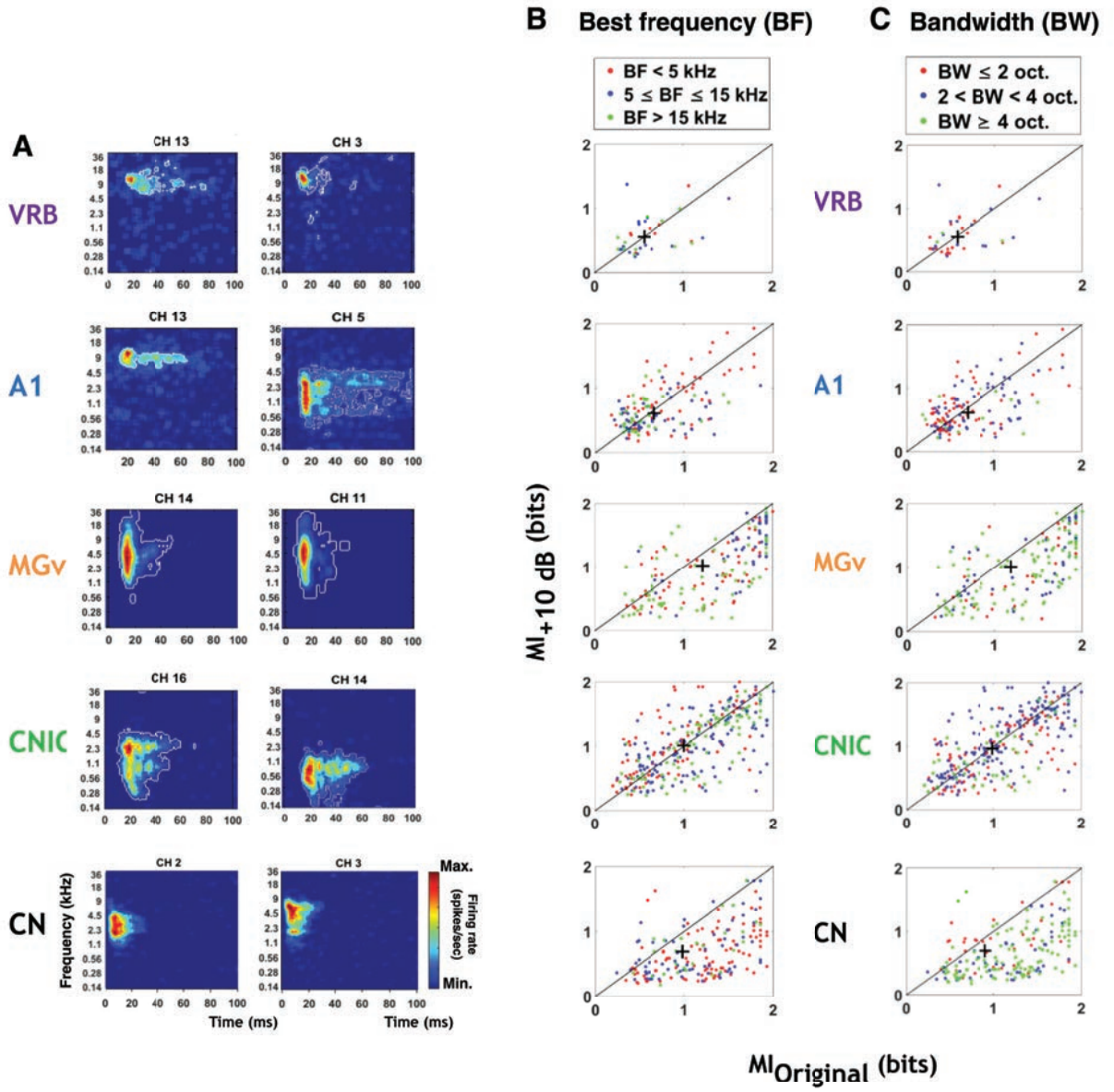

Figure 11.

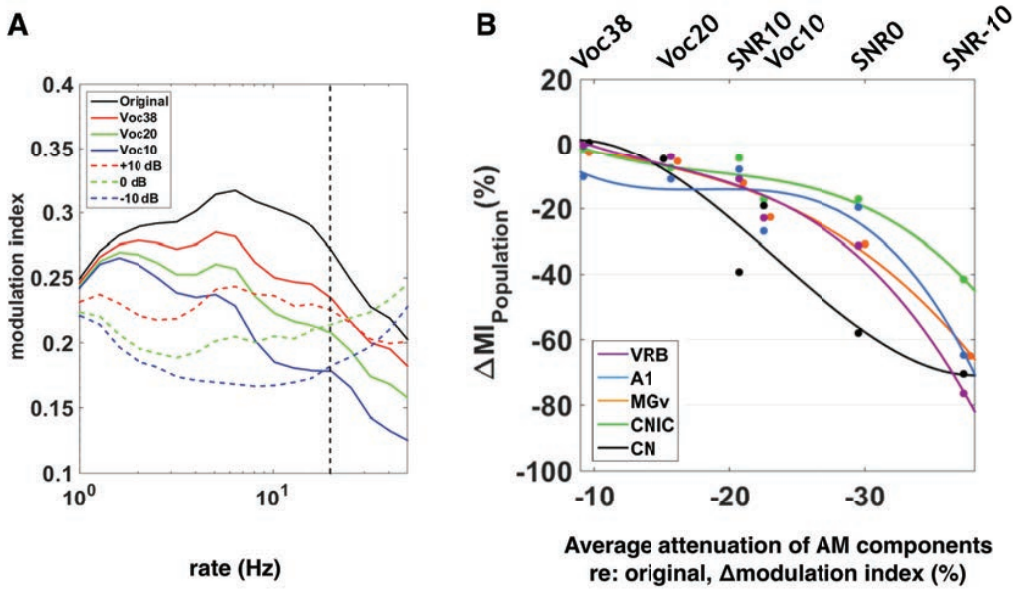

Figure 12.