The peer review history is available in the Supporting Information section of this article (https://doi.org/10.1113/JP283526#support-information-section).

Samira Souffi completed her PhD in France at the Institute of Neuroscience in Paris-Saclay University (NeuroPSI) under the supervision of Dr Jean-Marc Edeline. Her research was focused on understanding the effects of background noises on neuronal discrimination in the whole auditory system. Her postdoctoral research in Professor Israel Nelken's laboratory (ELSC, Hebrew University of Jerusalem, Israel) describes the neuronal networks involved in developing sound preferences (e.g. music preferences) on mice by recording the calcium activity in both the reward system and the auditory system.
Abstract  Humans and animals constantly face challenging acoustic environments, such as various background noises, that impair the detection, discrimination and identification of behaviourally relevant sounds. Here, we disentangled the role of temporal envelope tracking in the reduction in neuronal and behavioural discrimination between communication sounds in situations of acoustic degradations. By collecting neuronal activity from six different levels of the auditory system, from the auditory nerve up to the secondary auditory cortex, in anaesthetized guinea-pigs, we found that tracking of slow changes of the temporal envelope is a general functional property of auditory neurons for encoding communication sounds in quiet conditions and in adverse, challenging conditions. Results from a go/no-go sound discrimination task in mice support the idea that the loss of distinct slow envelope cues in noisy conditions impacted the discrimination performance. Together, these results suggest that envelope tracking is potentially a universal mechanism operating in the central auditory system, which allows the detection of any between-stimulus difference in the slow envelope and thus copes with degraded conditions.

(Received 29 June 2022; accepted after revision 8 November 2022; first published online 13 November 2022)

Corresponding author Jean-Marc Edeline: UMR 9197, Neuro-PSI (Institut des Neurosciences Paris-Saclay), CNRS – Université Paris-Saclay, Centre CEA Saclay, Bâtiment 151, 91400 Saclay cedex, France. Email: jean-marc.edeline@universite-paris-saclay.fr

Abstract figure legend  Methods: We simulated auditory nerve fibre (sANF) responses and recorded the neuronal activity in five auditory structures (from the cochlear nucleus to the secondary auditory cortex) in response to four vocalizations presented in quiet conditions and in two types of noise (a stationary noise and a chorus noise at three signal-to-noise ratios (SNRs): +10, 0 and −10 dB). In addition, we tested whether behaving animals can discriminate between whistles when engaged in a go/no-go task involving the discrimination between two of the four whistles used in our electrophysiological studies (W1 and W3). Licks to the $S^+$ were rewarded by a drop of water and licks to the $S^-$ were punished by a 5 s time-out period. Results: Subcortical and cortical auditory neurons track the slow changes of the temporal envelope ($<20$ Hz) with a high degree of fidelity in the original (positively correlated with the neuronal discrimination) and degraded conditions. Our results demonstrate that the between-stimulus envelope similarity, which increases in noisy conditions, is negatively correlated with both neuronal discrimination and behavioural performance.

Key points

• In quiet conditions, envelope tracking in the low amplitude modulation range ($<20$ Hz) is correlated with the neuronal discrimination between communication sounds as quantified by mutual information from the cochlear nucleus up to the auditory cortex.

• At each level of the auditory system, auditory neurons retain their abilities to track the communication sound envelopes in situations of acoustic degradation, such as vocoding and the addition of masking noises up to a signal-to-noise ratio of $-10$ dB.

• In noisy conditions, the increase in between-stimulus envelope similarity explains the reduction in both behavioural and neuronal discrimination in the auditory system.

• Envelope tracking can be viewed as a universal mechanism that allows neural and behavioural discrimination as long as the temporal envelope of communication sounds displays some differences.

Introduction

In humans, speech signals are characterized by rhythmic streams of amplitude and frequency modulations (AM and FM) that convey phoneme, syllable, word and phrase information (Ding et al., 2017; Rosen 1992; Varnet et al., 2017). It has been known for several decades that the low-frequency modulations of the temporal envelope carry essential cues for speech perception (Drullman et al., 1994a, 1994b; Shannon et al., 1995; Zeng et al., 2005). Even in challenging conditions (including in various types of noise), the human auditory system has the capacity to process highly degraded speech as long as the temporal envelope modulations below 20 Hz are preserved (Drullman et al., 1994a, 1994b; Shannon et al., 1995; Zeng et al., 2005). This is consistent with electroencephalographic (EEG) and magnetoencephalographic (MEG) studies in which cortical responses were found

© 2022 The Authors. The Journal of Physiology © 2022 The Physiological Society.
to be in phase with the temporal envelope of speech signals and strongly correlated with the average level of speech comprehension for both normal and compressed speech (Ahissar et al., 2001; Ding et al., 2014; Luo & Poeppel, 2007). In a recent study, Ortiz-Barajas and colleagues (2021) have found that newborns possess the neural capacity to track the amplitude and the phase of the speech envelope in their native language (French) as well as in rhythmically similar and different unfamiliar languages (Spanish and English). These results support the hypothesis that speech envelope tracking might be a necessary prerequisite, although not sufficient, for speech comprehension (Kösem & Van Wassenhove, 2017; Kösem et al., 2016).

In animals, the synchronization of auditory cortex responses with the temporal envelope of guinea-pig vocalizations has been observed in several studies (Grimsley et al., 2011, 2012; Wallace & Palmer, 2009; Wallace et al., 2005), some of them even suggesting that cortical responses could be isomorphic to the vocalization envelope (Grimsley et al., 2012: fig. 2A). Using speech stimuli with different levels of degradation (clear, conversational and compressed), Abrams and colleagues (2017) recorded responses of auditory cortex neurons in guinea-pigs and showed that populations of cortical neurons encode both the periodicity and compressed speech), and additional studies have pointed out that cortical neurons can still respond to target stimuli in important levels of noise (between −5 and 0 dB signal-to-noise ratio (SNR); Homma et al., 2020; Nagarajan et al., 2002; Narayan et al., 2007; Suta et al., 2003). At the subcortical level, several studies revealed in both mammals and birds that the average responses of inferior colliculus neurons can reflect the communication sound envelope (Rode et al., 2013; Suta et al., 2003; Woolley et al., 2006).

Here, we used acoustic degradations that differentially affected the similarities between acoustic envelopes: vocoders strongly altered the spectral cues but preserved most of the temporal information, whereas noise addition produced spectrottemporal degradations, reducing the temporal cues while introducing irrelevant envelope fluctuations and altering the spectral cues (Souffi et al., 2020: fig. 1). We used a stationary noise, which strongly increased the acoustic similarity between the envelopes, and a chorus noise, which differed between the four envelopes and therefore masked the vocalizations while not inducing an increase in the overall similarity of the stimuli. We showed previously that the addition of a stationary noise strongly impaired the neuronal discrimination performance at the subcortical and cortical levels, whereas the performance was less impaired in the vocoding conditions (Souffi et al., 2020: figs. 6–9).

To go further, our main goal in the present study was to determine whether the similarities between acoustic envelopes or the loss in envelope tracking ability by auditory neurons reduces or even prevents the neuronal and behavioural discrimination in situations of acoustic degradations. In a condition-independent scenario, the neurons keep the same intrinsic ability to track the stimulus envelopes whatever the acoustic conditions (in quiet and in degraded conditions): As long as the stimulus envelopes differ, the neurons will discriminate the stimuli. In contrast, in a condition-dependent scenario, the acoustic degradations reduce the ability of the neurons to track the stimulus envelopes. This deleterious effect can potentially occur when the neurons are strongly driven by the acoustic degradations (such as noise addition), leading to limited dynamic ranges for coding the target stimuli. This occurs, for example, for the responses of auditory nerve fibres (ANF) to tones in continuous noise. Although the responses to 120–300 Hz periodic AM stimuli were preserved at 0 and +6 dB SNR (Frisina et al., 1996: fig. 6), many studies have reported that the rate-level functions of ANF tested with pure tones were altered in noise; in many cases, the responses did not reach the same saturation level (Costalupes et al., 1984; Frisina et al., 1996; Geisler & Sinex, 1980; Rhode et al., 1978) or the whole curve was shifted toward the right (Costalupes et al., 1984; Rhode et al., 1978), indicating that the thresholds were higher and the dynamic ranges smaller than in quiet conditions. This was also a function of the bandwidth of the noise and the types of ANF (i.e. the effects differed between low, medium or high spontaneous rate fibres; e.g. see Reiss et al., 2011). Based on these studies, it seems that, as early as the auditory nerve, the detection of AM cues contained in target stimuli, and therefore the tracking abilities of central auditory neurons, can be reduced. Similar results have been observed in the inferior colliculus (Ramachandran et al., 2000).

In an attempt to distinguish between these two scenarios, we evaluated the relationship between the envelope tracking of sounds and the neuronal discrimination in the entire auditory system. We simulated auditory nerve fibre (sANF) responses (with a widely used model; Bruce et al., 2018) and recorded the neuronal activity in five auditory structures in response to four conspecific vocalizations presented in quiet conditions, using three tone-vocoders and two types of noise (a stationary noise and a chorus noise at three SNRs: +10, 0 and −10 dB) in anaesthetized guinea-pigs. We found that subcortical and cortical neurons track the envelopes in the low AM range (<20 Hz) with a high degree of fidelity in original and degraded conditions, suggesting that the auditory system maintains a robust temporal representation from the auditory nerve to
the auditory cortex. Behaving mice were also able to discriminate between these communication sounds and performed the assigned task above chance level in all noisy conditions. Overall, our results demonstrate that the between-stimulus envelope similarity, which increases in noisy conditions, is negatively correlated with both the neuronal discrimination and the behavioural performance.

**Methods**

Most of the methods are similar to those described by Souffi and colleagues (2020).

**Animals for the electrophysiological and behavioural experiments**

These experiments were performed under the national licence A-91-557 (project 2014-25, authorization 05202.02) and using the procedure numbers 32-2011 and 34-2012 validated by the Ethic committee number 59 [Comité d’Éthique pour l’Experimenation Animale (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 3–16 months old; 36 males and 11 females) at five different levels of the auditory system: the cochlear nucleus (CN), the inferior colliculus (IC), the medial geniculate body (MGB), the primary (A1) and secondary (area VRB) auditory cortex. Animals, weighing between 515 and 1100 g (mean 856 g), came from our own colony housed in a humidity- (50–55%) and temperature-controlled (22–24°C) facility on a 12 h–12 h light–dark cycle (light on at 07.30 h) with free access to food and water.

Two days before the electrophysiological experiment, the pure-tone audiogram of each animal was determined by testing auditory brainstem responses (ABR) under isoflurane anaesthesia (2.5%) as described by Gourévitch and colleagues (2009). A software package (RTLab; Echodia, Clermont-Ferrand, France) allowed averaging of 500 responses during the presentation of each pure-tone frequency and at each intensity (between 0.5 and 32 kHz; duration 10 ms; rise–fall time 2 ms) delivered by a speaker (Knowles Electronics) placed in the right ear canal of the animal. The threshold of each ABR was defined as the lowest intensity at which a small ABR wave could still be detected (usually wave III). For each frequency, the threshold was determined by gradually decreasing the sound intensity [from 80 down to −10 dB sound pressure level (SPL)]. There was a perfect agreement between the thresholds visually determined by two co-authors (S.S. and J.-M.E.). Based upon a large database of >250 guinea-pigs, we considered that all animals used in this study had normal pure-tone audiograms (Gourévitch & Edeline, 2011; Gourévitch et al., 2009).

Behavioural experiments were performed on nine 8-week-old C57Bl/6j female mice (for more details, see ‘Behavioural go/no-go discrimination task’ below).

**Acoustic stimuli**

The acoustic stimuli were the same as in the studies by Souffi and colleagues (2020, 2021). They were generated using MatLab, transferred to a RP2.1-based sound delivery system (TDT) and sent to a Fostex speaker (FE87E). The speaker was placed at 2 cm from the right ear of the guinea-pig, a distance at which the speaker produced a flat spectrum (±3 dB) between 140 Hz and 36 kHz. Calibration of the speaker was carried out using noise and pure tones recorded by a Bruel and Kjaer microphone (4133) coupled to a preamplifier (BandK 2169) and a digital recorder (Marantz PMD671).

Time–frequency response profiles (TFRPs) were determined using 129 pure-tone frequencies covering eight octaves (0.14–36 kHz) and presented at 75 dB SPL. The tones had a gamma envelope given by $\gamma(t) = (t/4)^2e^{-(t/4)}$, where $t$ is time (in milliseconds). At a given stimulus 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 13.6 ms, and at 50 ms the sound intensity was 6.7 dB SPL. 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 into Audition software (Adobe Audition 3), and four representative examples of whistles (W1–W4) were selected (Fig. 1A, left panel). As shown in Fig. 1B (left panel), their overall envelopes clearly differed, with W2 and W4 envelopes being the closest to each other. The four whistles were presented in two frozen noises ranging from 10 to 24,000 Hz. To generate these noises, audio-recordings were performed in the colony room where a large group of guinea-pigs were housed (30–40 animals; two to four animals per cage). Several 4 s segments of audio recordings were added up to generate the ‘chorus noise’, whose power spectrum was computed using a Fourier transform. The chorus noise masking each target vocalization was slightly different in terms of spectrottemporal
Envelope tracking of noisy sounds in the auditory system

Figure 1. Overall and filtered envelopes in three amplitude modulation ranges

A, spectrograms of original and degraded stimuli. The envelopes of the four original whistles (W1–W4) are presented in the left panel. Two whistles were used for a go/no-go behavioural discrimination task (see Fig. 7D): whistle 1 as the ‘go or S+’ stimulus and whistle 3 as the ‘no-go or S−’ stimulus. From left to right, the four envelopes of these stimuli are presented first, in the vocoding conditions (with 38, 20 and 10 frequency bands from top to bottom), then in stationary noise [at +10, 0 and −10 dB signal-to-noise ratio (SNR) from top to bottom] and in chorus noise conditions (at +10, 0 and −10 dB SNR from top to bottom). B, examples of the filtered envelopes for the original vocalizations using a bank of 35 gammatone filters with centre frequencies uniformly spaced along a guinea-pig-adapted equivalent rectangular bandwidth (ERB) scale ranging from 20 to 30,000 Hz. Three ranges of amplitude modulation (AM) have been investigated here: the low (<20 Hz), middle (between 20 and 100 Hz) and high (between 100 and 200 Hz) AM ranges. The red curves indicate the seven filtered envelopes selected along the signal for the subsequent analyses. [Colour figure can be viewed at wileyonlinelibrary.com]
content. The chorus noise spectrum was then used to shape the spectrum of a Gaussian white noise. The resulting 'vocalization-shaped stationary noise' therefore matched the 'chorus-noise' audio spectrum. Figure 1B displays the overall envelopes of the four whistles in the vocalization-shaped stationary noise (third panel) and in the chorus noise (fourth panel) with SNRs of +10, 0 and −10 dB.

The four selected whistles were also processed by three tone vocoders (Gnansia et al., 2009, 2010). In the following figures, the unprocessed whistles are referred to as the original versions, and the vocoded versions as Voc38, Voc20 and Voc10 using 38, 20 and 10 bands, respectively. 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 were found to introduce random (i.e. non-informative) intrinsic temporal-envelope fluctuations distorting the crucial spectrotemporal modulation features of communication sounds (Kates, 2011; Shamma & Lorenzi, 2013; Stone et al., 2011). Figure 1B displays the overall envelopes of the 38-band vocoded (first row, second panel), the 20-band vocoded (second row, second panel) and the 10-band vocoded (third row, second panel) versions of the four whistles. The three vocoders differed only in terms of the number of frequency bands (i.e. analysis filters) used to decompose the whistles (38, 20 or 10 bands). The 38-band vocoding process is described briefly 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 centre frequencies uniformly spaced along a guinea-pig-adapted equivalent rectangular bandwidth (ERB) scale ranging from 50 to 35,505 Hz (Sayles & Winter, 2010).

### Overall envelope extraction

In each frequency band, the temporal envelope was extracted using full-wave rectification and low-pass 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 centre 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 acoustic temporal fine structure across frequency channels (Hohmann, 2002). The modulated signals were finally weighted and summed over the 35 frequency bands (see section 'Quantification of the envelope tracking' below). The weighting compensated for imperfect superposition of the impulse responses of the bands at the desired group delay. The weights were optimized numerically to achieve a flat frequency response.

### Surgical procedures

All guinea-pigs were anaesthetized by an initial injection of urethane (1.2 g kg\(^{-1}\), i.p.) supplemented by additional doses of urethane (0.5 g kg\(^{-1}\), i.p.) when reflex movements were observed after pinching the hindpaw (usually two to four times during the recording session). A single dose of atropine sulphate (0.06 mg kg\(^{-1}\), s.c.) was given to reduce bronchial secretions, and a small dose of buprenorphine was administered (0.05 mg kg\(^{-1}\), s.c.) because urethane has no analgesic properties. After placing the animal in a stereotaxic frame, a craniotomy was performed, and a local anaesthetic (2% xylocaine) was injected into the wound.

For auditory cortex recordings (area A1 and VRB), a craniotomy was performed above the left temporal cortex. The dura above the auditory cortex was removed under binocular control, and the cerebrospinal fluid was drained through the cisterna to prevent the occurrence of oedema. For the recordings in MGB, a craniotomy was performed above the most posterior part of the MGB (8 mm posterior to bregma) to reach the left auditory thalamus at a location where the MGB is mainly composed of its ventral, tonotopic part (Anderson et al., 2007; Edeline et al., 1999; Redies et al., 1989; Wallace et al., 2007). For IC 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 was built from dental acrylic cement 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\(^{-1}\), i.p.) was administered to the animal.

### Recording procedures

Data from multi-unit recordings were collected in five 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 guinea-pig, neuronal recordings were collected in only one auditory structure.

Cortical extracellular recordings were obtained from arrays of 16 tungsten electrodes (TDT, TuckerDavis Technologies; 33 μm in diameter; <1 MΩ) composed of two rows of eight electrodes separated by 1000 μm (350 μm between electrodes of the same row). A silver wire, used as the earth, was inserted between the temporal bone and the dura mater on the contralateral side. The location
of the primary auditory cortex was estimated based on the pattern of vasculature observed in previous studies (Gaucher et al., 2013, 2020; Gaucher & Edeline, 2015; Wallace et al., 2000). The non-primary cortical area VRB was located ventral to A1 and distinguished by its longer response latencies to pure tones (Grimsley et al., 2012; Rutkowski et al., 2002). For each experiment, the position of the electrode array was set such that the two rows of eight electrodes sampled neurons responding from low to high frequency when progressing in the rostrocaudal direction [see examples in fig. 1 of Gaucher et al. (2012) and in fig. 6A of Occelli et al. (2016)].

In the MGB, IC and CN, the recordings were obtained using 16-channel multi-electrode arrays (NeuroNexus) composed of one shank (10 mm) of 16 electrodes spaced by 110 μm and with conductive site areas of 177 μm². The electrodes were advanced vertically (for MGB and IC) or with an angle of 40° to the CN surface until responses to pure tones could be detected on ≥10 electrodes.

All thalamic recordings were from the ventral part of the MGB (see above surgical procedures) and all displayed response latencies <9 ms. At the collicular level, we distinguished the lemniscal and non-lemniscal divisions of IC based on depth and the latencies of pure tone responses. We excluded the most superficial recordings (until a depth of 1500 μm) and those exhibiting latencies ≥20 ms, in an attempt to select recordings from the central nucleus of IC (CNIC). At the level of the cochlear nucleus, the recordings were collected from both 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 (sampling rate 25 kHz on each channel) was filtered (610 to 10,000 Hz) to extract multi-unit activity. The trigger level was set for each electrode to select the largest action potentials from the signal with a precision of 1 ms. Online and offline examination of the waveforms suggests that the multi-unit activity collected here was made of action potentials generated by a few neurons at the vicinity of the electrode.

However, given that we did not use tetrodes, the result of several clustering algorithms (Pouzat et al., 2004; Franke et al., 2015; Quiroga et al., 2004) based on spike waveform analyses were not reliable enough to isolate single units with confidence. Although these are not direct proofs, the facts that the electrodes were of similar impedance (0.5–1 MΩ) and that the spike amplitudes had similar values (100–300 μV) for the cortical and the subcortical recordings were two indications suggesting that the cluster recordings obtained in each structure included a similar number of neurons. Even if a similar number of neurons were recorded in the different structures, we cannot discard the possibility that the homogeneity of the multi-unit recordings (in terms of the number of cells contributing to each recording) differed between structures. By collecting several hundreds of recordings in each structure, these potential differences should be attenuated in the present study.

Simulations of auditory nerve fibre responses

A computational model of auditory nerve fibre responses was used to assess whether the envelope tracking properties measured in the central auditory system could be a mere consequence of the processing taking place at peripheral levels. For this purpose, we used a well-established and widely used model of the auditory periphery (Bruce et al., 2018). This model provides a phenomenological description of the major functional stages of the auditory periphery, from the middle ear up to the auditory nerve (Osses et al., 2022). The implementation used in the present study is available as the routine 'bruce2018' within the AMT toolbox (v.1.0) for MATLAB (Majdak et al., 2022).

In order to make the simulated data as comparable as possible to the neuronal responses collected in the electrophysiological experiments, the distribution of cochlear centre frequencies was chosen to be similar to the best frequencies obtained from the CN data. Default parameters were used for the later stages of the model. For each cochlear channel, five auditory-nerve fibres were simulated with different spontaneous rates (SRs): one low-SR fibre (SR = 0.1 spikes s⁻¹), one medium-SR fibre (SR = 4 spikes s⁻¹) and three high-SR fibres (SR = 100 spikes s⁻¹). The outcome of the model corresponds to the aggregated responses of these five simulated auditory nerve fibre fibres (sANF) in an attempt to keep the physiological ratio between low-, medium- and high-threshold fibres and to match roughly the number of cells contributing to the multi-unit activity collected in the central auditory structures (fewer than six neurons at the vicinity of the electrode).

The responses to 20 repetitions of each vocalization in the original and degraded conditions were simulated and analysed in the same way as recorded data.

Experimental protocol

Given that inserting an array of 16 electrodes into a brain structure unavoidably induces a deformation of this structure, a recovery time of 30 min was allowed for the structure to return to its initial shape, then the array was slowly lowered. Time–frequency response profiles were used to assess the quality of our recordings and to adjust electrode depth. For auditory cortex recordings (A1 and VRB), the recording depth was 500–1000 μm, which corresponds to layer III and the upper part of layer IV according to Wallace and Palmer (2008). For thalamic
recordings, the NeuroNexus probe was lowered ∼7 mm below the pia before the first responses to pure tones were detected. For the collicular and cochlear nucleus recordings, the NeuroNexus probe was inserted visually into the structure, and after a 15 min stabilization period, auditory stimuli were presented.

When a clear frequency tuning was obtained for ≥10 of the 16 electrodes, the stability of the tuning was assessed. We required that the recorded neurons displayed at least three (each lasting 6 min) successive similar TFRPs (i.e. with similar best frequencies) before starting the protocol. When the stability was satisfactory, the protocol was started by presenting the acoustic stimuli in the following order. We first presented the four whistles at 75 dB SPL in their original versions (in quiet conditions), then the vocoded whistles (Voc38, Voc20 and Voc10 versions) were presented at 75 dB SPL, followed by the vocalizations presented against the chorus noise and then against the vocalization-shaped stationary noise at 65, 75 and 85 dB SPL. Thus, the level of the original vocalizations was kept constant (75 dB SPL) and the noise level increased (65, 75 and 85 dB SPL). In all cases, each vocalization was repeated 20 times, and all the loudness levels are root mean square (RMS) values. Presentation of this entire stimulus set lasted 45 min. The protocol was restarted either after moving the electrode array on the cortical map or after lowering the NeuroNexus probe by ≥300 μm for subcortical structures.

**Behavioural go/no-go discrimination task**

Nine eight-week-old C57Bl/6J mice were water deprived (33 μl g⁻¹ day⁻¹) and trained daily for 200–300 trials in a go/no-go task involving two of the guinea-pig whistles (W1 and W3 in Fig. 1), one (the S+) signalling the reward (a drop of water) and the other not (the S−). The training procedures were similar to those described in previous studies (Ceballo et al., 2019; Deneux et al., 2016). Mice were head-fixed and held in a plastic tube on aluminium foil. Mice first performed one to three habituation sessions to learn to obtain a water reward (∼5 μl) by licking a stainless-steel water spout at least eight times after the positive stimulus, S+. A trial started only when the mice were not licking the spout for ≥3 s. Licks were detected by changes in resistance between the aluminium foil and the water spout. After habituation, the fraction of collected rewards was ∼80%.

The learning protocol then started, in which mice received the S−, for which they had to lick fewer than three times to avoid a 5 s time-out. One of the two whistles (the S+ or the S−) was presented every 10–20 s (uniform distribution), followed by a 1 s test period during which the mouse had to lick at least five to eight times to receive the reward. Positive and negative stimuli were played in a pseudorandom order, with the constraint that exactly four positive and four negative sounds must be played every eight trials.

Once a mouse showed ≥80% of correct discrimination between the S+ and the S− for two successive days in the original conditions, it was trained in noisy conditions, first with the stationary noise (successively at +10, 0 and −10 dB SNR), then with the chorus noise (successively at +10, 0 and −10 dB SNR). Each mouse had to perform ≥1 day at 80% in a given SNR to be tested on the following day at a lower SNR. Behavioural analyses were all automated, hence no animal randomization or experimenter blinding was used.

**Data analysis**

All the analyses were performed in MATLAB 2021 (MathWorks).

**Quantification of responses to pure tones.** The TFRPs were obtained by constructing peristimulus time histograms for each frequency with 1 ms time bins. The firing rate evoked by each frequency was quantified by summation all the action potentials from the tone onset up to 100 ms after this onset. Thus, TFRPs were matrices of 100 bins in abscissa (time) multiplied by 129 bins in ordinate (frequency). All TFRPs were smoothed with a uniform 5 × 5 bin window for visualization (not for the data analyses). For each TFRP, the best frequency (BF) was defined as the frequency at which the highest firing rate was recorded. Peaks of significant response were automatically identified using the following procedure. A positive peak in the TFRP was defined as a contour of firing rate above the average level of the baseline activity (100 ms of spontaneous activity taken before each tone onset) plus six times the standard deviation of the baseline activity. Recordings without a significant peak of responses or with inhibitory responses (decreases in firing rate 3 standard deviations below spontaneous activity) were excluded from the data analyses.

**Quantification of the envelope tracking.** We first extracted the envelope as explained in a previous section (see above, ‘Overall envelope extraction’), then filtered them (original, vocoded and noisy vocalizations) using a bank of 35 gammatone filters with centre frequencies uniformly spaced along a guinea-pig-adapted ERB (equivalent rectangular bandwidth) scale ranging from 20 to 30,000 Hz. Then, three ranges of amplitude modulation (AM) were investigated: the low (L, <20 Hz), middle (M, between 20 and 100 Hz) and high (H, between 100 and 200 Hz) AM ranges. For all the AM filtering, we used Butterworth filters at −6 dB per octave. Second, the envelopes were downsampled to a
resolution of 1 ms to match the sampling rate of the peristimulus time histograms (PSTHs). Third, we applied a half-wave rectification followed by a normalization with the corresponding RMS value.

The neuronal responses (i.e. the PSTHs) were also filtered with the same three frequency bands as the envelopes, followed by a normalization with the corresponding RMS value. The rationale for this filtering step was that we wanted to isolate and quantify the correspondence between temporal aspects of the stimuli in particular frequency ranges and PSTHs.

Next, we performed normalized cross-correlations between the filtered envelopes and PSTHs for each AM range. We selected seven gammatones, as a trade-off between accurately representing the envelopes along the audio spectrum and minimizing redundancy between envelopes. Maximal values in the correlograms were automatically detected in each structure to account for propagation delays in the auditory system. The lags were selected according to the distributions of the latencies obtained in response to pure tones at 75 dB SPL. The different lags identified were 1–10 ms for CN, 5–20 ms for CNIC, 6–15 ms for ventral part of the medial geniculate (MGv), 9–30 ms for A1 and 9–40 ms for VRB. In all analyses, we decided to keep the maximal correlation coefficient out of the seven selected gammatone filters (RmaxE-PSTH).

**Evaluation of the correlation significance by shuffling the evoked activity.** It is known that a significant correlation between neuronal events and sensory stimuli can be obtained by chance (for a review, see Harris, 2020). Therefore, it was crucial to run drastic controls to reduce the probability that the correlations detected here were spurious.

To determine a significance threshold for the correlation, we shuffled only the evoked activity in the original conditions on a time scale of 1 ms, in order to preserve the global shape of the whole response (i.e. the four response peaks attributable to the starting of each stimulus, separated by a period of silence). Specifically, for the original conditions, we shuffled only the spikes obtained during the presentation of each whistle to avoid adding spikes in the silent period. The shuffled PSTHs obtained were then processed using the same procedure as for unshuffled PSTHs (i.e. filtering in the three AM ranges and half-wave rectification followed by a normalization with the corresponding RMS value). Then, we computed the cross-correlation (\( R_{\text{Random}} \)) between each shuffled PSTH and each envelope. We performed this procedure 1000 times and set, for each correlation value PSTH-Envelope (PSTH-E), a significance threshold of the R value that is the mean of the \( R_{\text{Random}} \) values plus two times the standard deviations \([\mu(R_{\text{Random}}) \pm 2\sigma]\).

Based upon this criterion, percentages of recordings were discarded in each structure and for each AM range as follows: in VRB, 30, 10 and 47% of recordings were discarded in the L, M and H range, respectively; in A1, 51, 38 and 73% of recordings were discarded in the L, M and H range, respectively; in MGv, 61, 49 and 63% of recordings were discarded in the L, M, and H range, respectively; in CNIC, 29, 26 and 33% of recordings were discarded in the L, M and H range, respectively; in CN, 33, 43 and 50% of recordings were discarded in the L, M and H range, respectively; and in sANF, 35, 86 and 77% of recordings were discarded in the L, M, and H range, respectively. Although this drastic procedure discarded a non-negligible proportion of recordings, it reduced the probability that the correlations described here were obtained by chance.

**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 contained in the responses to vocalizations obtained with natural, vocoded or noisy stimuli. This method allows quantification of how well the identity of the vocalization can be inferred from neuronal responses. Neuronal responses were represented using different time scales ranging from the duration of the whole response (total spike count) to a 1 ms precision (precise temporal patterns), which allows analysis of how much the spike timing contributes to the information. Given that this method is described exhaustively by Schnupp and colleagues (2006) and Gaucher and colleagues (2013), below we present only 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 subcortical or cortical site to the different stimuli, a single response (test pattern) is extracted and represented as a PSTH with a given bin size. Then, a mean response pattern is computed from the remaining responses 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, in which 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 four whistles and selected the first 280 ms of these responses in order 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 four 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 fig. 2B in the paper by Souffi and colleagues (2020) for the evolution of MI with temporal precisions ranging from 1 to 40 ms]. The value of 8 ms was selected for the data analysis because in each structure the MI reached its maximum at this value of temporal precision.

The MI estimates are subject to non-negligible positive sampling biases. Therefore, as in the study by 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 (\( \text{MI}_{\text{bias}} \)). These \( \text{MI}_{\text{bias}} \) estimates were then used as estimators for computation of the statistical significance of the MI estimate for the real (unshuffled) data sets. The real estimate was considered significant if its value was statistically different from the distribution of \( \text{MI}_{\text{bias}} \) shuffled estimates. Significant MI estimates were computed for MI calculated from neuronal responses under one electrode and for each of the conditions. Therefore, there was an \( \text{MI}_{\text{bias}} \) value for each MI estimate. The range of \( \text{MI}_{\text{bias}} \) values was very similar between brain structures; depending on the conditions (original, vocoded and noisy vocalizations), it ranged from 0.102 to 0.107 bits in the CN, from 0.107 to 0.110 bits in the IC, from 0.105 to 0.114 bits in the MGB, from 0.107 to 0.111 bits in the A1 and from 0.106 to 0.116 bits in VRB. There was no significant difference between the mean values of \( \text{MI}_{\text{bias}} \) in the different structures (Student’s unpaired t test, all \( P > 0.25 \)).

Quantification of acoustic envelope similarity. For each of the acoustic conditions and each AM range, we quantified the acoustic similarity between each pair of stimuli as the correlation between their envelopes across the seven selected gammatones. Then, we averaged the six correlation values (related to all possible combinations with the four stimuli) to obtain an estimate of the similarity between the four stimuli for each of the conditions (original, vocoding and noisy conditions) and each AM range (see Fig. 7A, dark lines). More precisely, we averaged Fisher z-transformed coefficients and reported the back-transformed averages in Fig. 7A. In order to confirm that there was no bias in our gammatone selection, we carried out the same analysis on the output of the 35 gammatones and obtained similar results (see Fig. 7A, light lines).

Statistical analysis

We used an ANOVA for multiple factors to reveal the main effects in the whole data set (vocoding conditions, three levels; masking noise conditions, three levels for each noise; auditory structures, six levels; and AM ranges, three levels). Post hoc pairwise tests were performed between the original conditions and the vocoding or noisy conditions, or between structures to assess the significance of the multiple comparisons. They were corrected for multiple comparisons using Bonferroni corrections and were considered as significant if their \( P \)-value was <0.05.

Results

We simulated auditory nerve fibre (sANF) responses and collected neuronal recordings from five auditory structures: the cochlear nucleus (CN, 10 animals), the central nucleus of the inferior colliculus (CNIC, 11 animals), the ventral part of the medial geniculate (MGv, 10 animals), the primary auditory cortex (A1, 11 animals) and a secondary auditory area (VRB, 5 animals).

All analyses were performed on a set of recordings (or simulated recordings) selected using stringent criteria, and \( n \) values correspond to the number of selected recordings. Note that all the R values presented below are considered significant (for more details, see the Methods section).

Figure 1A and B illustrates the spectrograms and the overall envelopes of all stimuli in the original, vocoded and noisy conditions. In the following, the term 'stimulus' refers either to the four original or vocoded whistles or to the four whistles embedded in noise. The four overall envelopes of the stimuli were clearly different between each other in the original and vocoded conditions; however, they became progressively more similar in noisy conditions as the SNR decreased, especially in stationary noise.

Auditory neurons track the envelopes in the low AM range better than in middle and high AM ranges

We first determined which ranges of amplitude modulations are tracked by the subcortical and cortical neurons. To address this question, we filtered both the envelopes and the neuronal responses in three AM ranges:
the low (<20 Hz), middle (between 20 and 100 Hz) and high (between 100 and 200 Hz) AM ranges. Figure 1C presents the seven selected Butterworth-filtered envelopes (among the 35) of the four whistles after first having been filtered using a gammatone filterbank and indicates that the low AM range contained larger envelope fluctuations than the middle and high AM ranges. Figure 2A and B presents individual examples (Fig. 2A) and populations (Fig. 2B) of PSTHs constructed from the responses to presentation of the original vocalizations in each structure. Based on these PSTHs, it appears that the evoked responses tended to be more phasic in the two cortical areas (A1 and VRB) than in the subcortical structures.

Figure 2C–E shows the PSTHs from individual recordings (in black) and stimulus envelopes (E, in red) obtained in each structure (and for the sANF) in the original conditions, both filtered in the same AM ranges. For each example of E-PSTH, we indicate the cross-correlation value (R) at the top left of each panel. In the following results, the correlation value selected for each recording at a given AM range was the maximum over the seven gammatone filters (RmaxE-PSTH). Note that similar results were obtained when we used the correlation value obtained with the gammatone filter closest to the best frequency of each neuronal recording (data not shown). Whatever the structure, in these individual recordings, the higher R values were in the low AM range rather than in the middle and high AM ranges (Fig. 2C–E). Figure 2F presents the distribution, the mean and the interquartile range of the RmaxE-PSTH values for each structure in the three AM ranges (L, M and H) in the original conditions. Overall, we found a statistically significant difference in average RmaxE-PSTH values for both the three AM ranges and the six structures (two-way ANOVA, \( P < 0.05 \)), with a significant interaction between these two factors. For all structures, the mean RmaxE-PSTH values were much higher in the L range compared with the M and H ranges. In the low AM range, sANF, CN and CNIC recordings displayed significantly higher mean RmaxE-PSTH values than MGv and cortical recordings [one-way ANOVA, \( P < 0.0001 \), \( F_{(5,1165)} = 138.25 \) with Student’s unpaired \( t \) test, sANF vs. MGv, A1 or VRB \( P < 0.0001 \), CN vs. MGv, A1 or VRB \( P < 0.0001 \), CNIC vs. MGv, A1 or VRB \( P < 0.0001 \); mean (±SD) RmaxE-PSTH values: \( R_{\text{MGv}}(n = 217) = 0.81 \pm 0.03 \), \( R_{\text{CN}}(n = 336) = 0.80 \pm 0.08 \), \( R_{\text{CNIC}}(n = 274) = 0.78 \pm 0.11 \), \( R_{\text{MGv}}(n = 102) = 0.68 \pm 0.13 \), \( R_{\text{A1}}(n = 171) = 0.60 \pm 0.13 \) and \( R_{\text{VRB}}(n = 66) = 0.63 \pm 0.14 \)]. In the middle and high AM ranges, the differences between the structures were less clear, but the CNIC recordings still exhibited slightly higher mean RmaxE-PSTH values compared with the other structures [mean (±SD) RmaxE-PSTH values in the middle AM range: \( R_{\text{sANF}}(n = 77) = 0.31 \pm 0.07 \), \( R_{\text{MGv}}(n = 220) = 0.27 \pm 0.05 \) and \( R_{\text{VRB}}(n = 85) = 0.29 \pm 0.06 \); mean (±SD) RmaxE-PSTH values in the high AM range: \( R_{\text{sANF}}(n = 77) = 0.31 \pm 0.03 \), \( R_{\text{CN}}(n = 249) = 0.31 \pm 0.05 \), \( R_{\text{CNIC}}(n = 257) = 0.33 \pm 0.05 \), \( R_{\text{MGv}}(n = 97) = 0.29 \pm 0.04 \), \( R_{\text{A1}}(n = 196) = 0.27 \pm 0.05 \) and \( R_{\text{VRB}}(n = 50) = 0.30 \pm 0.05 \)]. This poor ability to follow fast AM changes was expected for auditory cortex neurons but not expected for subcortical neurons and for sANF (which can synchronize at higher AM rates when tested with periodic artificial stimuli; reviewed by Joris et al., 2004). This suggests that only a partial encoding of high AM rates contained in complex natural sounds is performed by subcortical neurons.

To summarize, in the original conditions, the PSTHs of the neurons were more strongly correlated with the stimulus envelope in the low AM range than in the middle and high AM ranges, at both subcortical and cortical levels.

**In the original conditions, the better the cortical and subcortical neurons track the slow envelope (<20 Hz), the higher the value of mutual information**

Does envelope tracking allow auditory neurons to discriminate the four vocalizations in the original conditions? To address this question, we examined whether there is a relationship between the neuronal discrimination performance and the abilities of neurons to follow the stimulus envelope (Fig. 3). The distribution, the mean and the interquartile range of the neuronal discrimination (quantified by the MI) are presented for each structure in Fig. 3A. As previously reported (Souffi et al., 2020), subcortical neurons (CN, CNIC and MGv neurons) were better at discriminating the original whistles compared with cortical neurons (A1 and VRB neurons), and here we extended this result to sANF [one-way ANOVA \( P < 0.0001 \), \( F_{(5,1538)} = 266.46 \) with Student’s unpaired \( t \) test, sANF vs. A1 or VRB \( P < 0.0001 \), CN vs. A1 or VRB \( P < 0.0001 \), CNIC vs. A1 or VRB \( P < 0.0001 \); mean (±SD) MI values: \( M_{\text{SANF}}(n = 77) = 1.84 \pm 0.21 \) bits, \( M_{\text{CNIC}}(n = 249) = 0.92 \pm 0.47 \) bits, \( M_{\text{MGv}}(n = 257) = 1.00 \pm 0.5 \) bits, \( M_{\text{A1}}(n = 196) = 0.68 \pm 0.37 \) bits and \( M_{\text{VRB}}(n = 50) = 0.55 \pm 0.29 \) bits; Fig. 3A]. The scattergrams presented in Fig. 3B display the RmaxE-PSTH values as a function of the MI values in each structure and AM range.

Figure 3C summarizes the correlation values between RmaxE-PSTH and MI parameters, in each structure and in the three AM ranges. All significant correlation values between these two variables are reported in red. In all but one case (in CNIC in the middle AM range), significant positive correlations between RmaxE-PSTH and MI values were obtained in all AM ranges in subcortical structures.
Figure 2. Correlations in the original conditions between peristimulus time histograms of subcortical and cortical recordings and the stimulus envelope, both filtered in the three selected amplitude modulation ranges.

A, individual examples of original peristimulus time histograms (PSTHs) obtained in each structure (from bottom to top: sANF, simulated auditory nerve fibres; CN, cochlear nucleus; CNIC, central nucleus of the inferior colliculus; MGv, ventral division of the medial geniculate; A1, primary auditory cortex; VRB, ventrorostral belt). B, population responses ranked from the lowest to the highest best frequencies, with the colour code representing the normalized firing rate. At the bottom of each panel, the population firing rate represents the instantaneous summed activity of the whole virtual population, and on the right, the total firing rate along the different best frequencies. C–E, examples of correlations between the PSTH (in black) and the envelope (in red). In each panel, the PSTHs and the stimulus envelopes are filtered in the same frequency range. For each recording, the correlation value between the PSTH and the envelope is shown at the top left. In a given amplitude modulation (AM) range, the stimulus envelopes differ between examples because we selected the gammatone envelope (out of seven gammatones).
that induced the highest correlation. Note that the PSTHS are not lagged compared with the envelopes as during the analysis, $F$, box plots showing the distributions of the maximal correlation coefficient out of the seven selected gammatone filters (RmaxE-PSTH) values for the six auditory structures (sANF–VRB) in the three AM ranges. The red dots in the box plots correspond to the mean RmaxE-PSTH values, and the boxes correspond to the interquartile ranges. Note the higher RmaxE-PSTH values in the low (L) AM range compared with the middle (M) and high (H) AM ranges. The black lines represent significant differences between the mean RmaxE-PSTH values. In the low AM range, sANF, CN and CNIC recordings displayed significantly higher mean RmaxE-PSTH values than MGv and cortical recordings [low AM range: one-way ANOVA, $P < 0.0001$, $F_{(5,1165)} = 138.25$. At Student's unpaired t test, sANF vs. MGv, A1 or VRB $P < 0.0001$, CN vs. MGv, A1 or VRB $P < 0.0001$; CNIC vs. MGv, A1 or VRB $P < 0.0001$; mean ($\pm$SD) RmaxE-PSTH values: RANF($n = 217$) = 0.81 $\pm$ 0.03, RCN($n = 336$) = 0.80 $\pm$ 0.08, RHCN($n = 274$) = 0.78 $\pm$ 0.11, RGGv($n = 102$) = 0.68 $\pm$ 0.13, RA1($n = 171$) = 0.60 $\pm$ 0.13 and RVBB($n = 66$) = 0.63 $\pm$ 0.14. In the middle and high AM ranges, the CNIC recordings still exhibited slightly higher mean RmaxE-PSTH values compared with the other structures [mean ($\pm$SD) RmaxE-PSTH values in the middle AM range: $R_{\text{ANF}}(n = 44) = 0.32 \pm 0.03$, $R_{\text{CN}}(n = 285) = 0.31 \pm 0.07$, $R_{\text{HCN}}(n = 285) = 0.34 \pm 0.07$, $R_{\text{MGv}}(n = 133) = 0.31 \pm 0.07$, $R_{\text{A1}}(n = 220) = 0.27 \pm 0.05$ and $R_{\text{VRB}}(n = 85) = 0.29 \pm 0.06$; mean ($\pm$SD) RmaxE-PSTH values in the high AM range: $R_{\text{ANF}}(n = 77) = 0.31 \pm 0.03$, $R_{\text{CN}}(n = 249) = 0.31 \pm 0.05$, $R_{\text{HCN}}(n = 257) = 0.33 \pm 0.05$, $R_{\text{MGv}}(n = 97) = 0.29 \pm 0.04$, $R_{\text{A1}}(n = 196) = 0.27 \pm 0.05$ and $R_{\text{VRB}}(n = 50) = 0.30 \pm 0.05$]. Note that $n$ values correspond to the selected simulations or recordings. [Colour figure can be viewed at wileyonlinelibrary.com]

The acoustic degradations decreased the values of mutual information but did not affect the envelope tracking performed by the auditory neurons

In almost all situations of acoustic degradations, the ability of neurons to discriminate between the four vocimations was decreased. Figure 4 presents the distributions of the MI values for the original conditions and the three levels of degradation conditions: the three tone-vocoders (38, 20 and 10 frequency bands) or the three SNRs (+10, 0 and −10 dB SNR) in the two types of noise (stationary noise and chorus noise). In general, there were modest effects on the MI values in the vocoding and chorus noise conditions compared with the stationary noise conditions for all structures except sANF. The decrease was significant only for the 10-band vocoded vocalizations in MGv and A1 [Fig. 4D, one-way ANOVA, $P = 0.05$, $F_{(3,811)} = 2.58$ with Student's paired t test, MGvOri vs. MGvVoc38 $P < 0.0001$, mean ($\pm$SD) MI values: $M_{\text{Ori}} = 1.19 \pm 0.55$ bits, $M_{\text{Voc38}} = 1.13 \pm 0.57$ bits, $M_{\text{Voc20}} = 1.15 \pm 0.55$ bits and $M_{\text{Voc10}} = 1.03 \pm 0.53$ bits; Fig. 4E, one-way ANOVA, $P = 0.001$, $F_{(3,722)} = 3.73$ with Student's paired t test, A1Ori vs. A1Voc10 $P < 0.0001$, mean ($\pm$SD) MI values: $M_{\text{Ori}} = 0.68 \pm 0.37$ bits, $M_{\text{Voc38}} = 0.62 \pm 0.33$ bits, $M_{\text{Voc20}} = 0.61 \pm 0.33$ bits and $M_{\text{Voc10}} = 0.56 \pm 0.30$ bits], and no significant difference was detected in VRB [Fig. 4F, one-way ANOVA, $P = 0.75$, $F_{(3,186)} = 0.41$, mean ($\pm$SD) MI values: $M_{\text{Ori}} = 0.55 \pm 0.29$ bits, $M_{\text{Voc38}} = 0.55 \pm 0.28$ bits, $M_{\text{Voc20}} = 0.55 \pm 0.29$ bits and $M_{\text{Voc10}} = 0.61 \pm 0.31$ bits]. However, the decrease was already significant with 38-band vocoded vocalizations in sANF or with 20-band vocoded vocalizations in CN and CNIC [Fig. 4A, one-way ANOVA, $P < 0.0001$, $F_{(3,1302)} = 111.3$ with Student's paired t test, sANFOri vs. sANFVoc38 $P < 0.0001$, mean ($\pm$SD) MI values: $M_{\text{Ori}} = 1.84 \pm 0.21$ bits, $M_{\text{Voc38}} = 1.61 \pm 0.32$ bits, $M_{\text{Voc20}} = 1.39 \pm 0.52$ bits and $M_{\text{Voc10}} = 1.29 \pm 0.52$ bits; Fig. 4B, one-way ANOVA, $P < 0.0001$, $F_{(3,1424)} = 12.42$ with Student's paired t test, CNICOri vs. CNICVoc20 $P < 0.0001$, mean ($\pm$SD) MI values: $M_{\text{Ori}} = 0.92 \pm 0.47$ bits, $M_{\text{Voc38}} = 0.96 \pm 0.46$ bits, $M_{\text{Voc20}} = 0.86 \pm 0.45$ bits and $M_{\text{Voc10}} = 0.75 \pm 0.43$ bits; Fig. 4G, one-way ANOVA, $P < 0.0001$, $F_{(3,1331)} = 13.17$ with Student's paired t test, CNICOri vs. CNICVoc20 $P < 0.0001$, mean ($\pm$SD) MI values: $M_{\text{Ori}} = 1.00 \pm 0.49$ bits, $M_{\text{Voc38}} = 0.99 \pm 0.46$ bits, $M_{\text{Voc20}} = 0.90 \pm 0.46$ bits and $M_{\text{Voc10}} = 0.79 \pm 0.39$ bits]. Note that there was also a significant increase in MI values with the 38-band vocoded vocalizations in CN.
Figure 3. In the original conditions, the better that cortical and subcortical neurons track the slow envelope (<20 Hz), the higher the value of mutual information.

A, box plots showing the distributions of the mutual information (MI) values obtained in the six levels of the auditory system in the original conditions. The red dots in the box plots correspond to the mean MI values, and the boxes correspond to the interquartile ranges. Note the lower significant values obtained at the cortical level in the primary auditory cortex (A1) and ventrorostral belt (VRB) compared with those obtained in simulated auditory nerve fibres (sANF) and subcortical structure (one-way ANOVA, $P < 0.0001$, $F_{(5,1538)} = 266.46$ with Student’s unpaired $t$ test, sANF vs. A1 or VRB $P < 0.0001$, cochlear nucleus (CN) vs. A1 or VRB $P < 0.0001$, central nucleus of the inferior colliculus (CNIC) vs. A1 or VRB $P < 0.0001$, ventral division of the medial geniculate (MGV) vs. A1 or VRB $P < 0.0001$; mean ($±$SD) MI values: $MI_{sANF}$($n = 77$) = 1.84 ± 0.21 bits, $MI_{CN}$($n = 249$) = 0.92 ± 0.47 bits, $MI_{CNIC}$($n = 257$) = 1.19 ± 0.55 bits, $MI_{A1}$($n = 196$) = 0.68 ± 0.37 bits and $MI_{VRB}$($n = 50$) = 0.55 ± 0.29 bits). Note that $n$ values correspond to the selected simulations or recordings. The black lines represent the first significant differences between the mean values. Note that for the sake of clarity, not all significant differences are indicated by black lines. For example, the difference between sANF and MGV was the first one to be significant, but it was also significant between sANF and the two cortical areas, A1 and VRB.

B, scattergrams showing the relationships between the maximal correlation coefficient out of the seven selected gammatone filters ($R_{max,E-PSTH}$) and MI values for the six structures in the three AM ranges. Black lines represent the first significant differences between the mean values.

C, matrix summarizing the mutual information (MI) values between $R_{max,E-PSTH}$ and MI.

|        | L   | M   | H   |
|--------|-----|-----|-----|
| VRB    | 0.19| 0.10| 0.00|
| A1     | 0.27| 0.19| 0.14|
| MGV    | 0.29| 0.42| 0.27|
| CNIC   | 0.30| 0.07| 0.18|
| CN     | 0.29| 0.35| 0.17|

Mean ($±$SD) MI values: $MI_{sANF}$($n = 77$) = 1.84 ± 0.21 bits, $MI_{CN}$($n = 249$) = 0.92 ± 0.47 bits, $MI_{CNIC}$($n = 257$) = 1.19 ± 0.55 bits, $MI_{A1}$($n = 196$) = 0.68 ± 0.37 bits and $MI_{VRB}$($n = 50$) = 0.55 ± 0.29 bits.
correlation coefficients between $R_{max\_PSTH}$ and MI in each structure and AM range. The values in red indicate that the correlation was significant. In all but one case (in CNIC in the middle AM range), significant positive correlations between $R_{max\_PSTH}$ and MI values were observed in all AM ranges in subcortical structures ($F^{CNIC} < 0.0001$, $P^{CNIC} < 0.0001$, $P^{CNIC} = 0.24$, $P^{CNIC} = 0.005$, $P^{MGv} = 0.006$, $P^{MGv} < 0.0001$ and $P^{MGv} = 0.01$). For the sANF, the range of MI values was too limited to compute reliable correlations (most MI values were close to the maximum of 2 bits). At the subcortical level, the highest correlation values between $R_{max\_PSTH}$ and MI values were found in MGv. At the cortical level, significant correlations between $R_{max\_PSTH}$ and MI values were detected in the low and middle AM ranges in A1 ($P^{A1} = 0.01$, $P^{A1} = 0.05$ and $P^{A1} = 0.32$), and there was no significant correlation in VRB ($F^{VRB} = 0.31$, $P_{VRB} = 0.51$ and $P_{VRB} = 0.99$). [Colour figure can be viewed at wileyonlinelibrary.com]

(Fig. 4B, one-way ANOVA, $P < 0.0001$, $F_{(3,1424)} = 12.42$ with Student's paired $t$ test, CNIC vs. CNVoc38 $P = 0.0073$).

In noise, in CN and CNIC, there was no significant decrease in mean MI values [Fig. 4B, one-way ANOVA, $P = 0.05$, $F_{(3,1176)} = 2.65$, mean (±SD) MI values: $MIOri = 0.92 ± 0.47$ bits, $MIOri = 0.85 ± 0.48$ bits, $MIOri = 0.83 ± 0.50$ bits and $MIOri = 0.83 ± 0.49$ bits; Fig. 4C, one-way ANOVA, $P = 0.36$, $F_{(3,1188)} = 1.06$, $MIOri = 1.00 ± 0.49$ bits, $MIOri = 1.05 ± 0.50$ bits, $MIOri = 1.06 ± 0.52$ bits], whereas in sANF and MGv, the mean MI values significantly decreased at +10 or 0 dB SNR, respectively [Fig. 4A, one-way ANOVA, $P < 0.0001$, $F_{(3,1333)} = 232.86$ with Student's paired $t$ test, sANF vs. sANF $P < 0.0001$, mean (±SD) MI values: $MIOri = 1.84 ± 0.21$ bits, $MIOri = 1.55 ± 0.37$ bits, $MIOri = 1.20 ± 0.45$ bits and $MIOri = 1.09 ± 0.50$ bits; Fig. 4D, one-way ANOVA, $P < 0.0001$, $F_{(3,753)} = 7.3$ with Student's paired $t$ test, MGv vs. MGv $P < 0.0001$, mean (±SD) MI values: $MIOri = 1.19 ± 0.55$ bits, $MIOri = 1.08 ± 0.52$ bits, $MIOri = 0.99 ± 0.50$ bits and $MIOri = 0.98 ± 0.48$ bits]. At the cortical level, there was a significant decrease in A1 at 0 dB SNR and no significant change of mean MI values in VRB [Fig. 4E, one-way ANOVA, $P = 0.0039$, $F_{(3,679)} = 4.5$ with Student's paired $t$ test, A1 vs. A1 $P < 0.0001$, mean (±SD) MI values: $MIOri = 0.68 ± 0.37$ bits, $MIOri = 0.64 ± 0.36$ bits, $MIOri = 0.55 ± 0.30$ bits and $MIOri = 0.60 ± 0.31$ bits; Fig. 4F, one-way ANOVA, $P = 0.31$, $F_{(3,179)} = 1.19$, mean (±SD) MI values: $MIOri = 0.55 ± 0.29$ bits, $MIOri = 0.63 ± 0.32$ bits, $MIOri = 0.54 ± 0.27$ bits and $MIOri = 0.53 ± 0.24$ bits].

Stationary noise strongly reduced the MI values in comparison to the vocoding and the choroid noise addition. The mean MI value in sANF, CN and MGv was significantly reduced already at +10 dB SNR [Fig. 4A, one-way ANOVA, $P < 0.0001$, $F_{(3,1153)} = 767.64$ with Student's paired $t$ test, sANF vs. sANF $P < 0.0001$, mean (±SD) MI values: $MIOri = 1.84 ± 0.21$ bits, $MIOri = 1.09 ± 0.39$ bits, $MIOri = 0.68 ± 0.37$ bits and $MIOri = 0.66 ± 0.39$ bits; Fig. 4B, one-way ANOVA, $P < 0.0001$, $F_{(3,812)} = 61.22$ with Student's paired $t$ test, CN vs. CN $P < 0.0001$, mean (±SD) MI values: $MIOri = 0.92 ± 0.47$ bits, $MIOri = 0.68 ± 0.38$ bits, $MIOri = 0.53 ± 0.26$ bits and $MIOri = 0.41 ± 0.17$ bits; Fig. 4D, one-way ANOVA, $P < 0.0001$, $F_{(3,630)} = 62.03$ with Student's paired $t$ test, MGv vs. MGv $P < 0.0001$, mean (±SD) MI values: $MIOri = 1.19 ± 0.55$ bits, $MIOri = 1.00 ± 0.50$ bits, $MIOri = 0.74 ± 0.35$ bits and $MIOri = 0.46 ± 0.19$ bits], whereas the mean MI value in CNIC was significantly reduced at 0 dB SNR [Fig. 4C, one-way ANOVA, $P < 0.0001$, $F_{(3,1078)} = 32.08$ with Student's paired $t$ test, CNIC vs. CN $P < 0.0001$, mean (±SD) MI values: $MIOri = 1.00 ± 0.49$ bits, $MIOri = 1.00 ± 0.49$ bits, $MIOri = 0.58 ± 0.44$ bits and $MIOri = 0.63 ± 0.28$ bits]. At the cortical level, noise significantly reduced the mean MI value in A1 only at −10 dB SNR [Fig. 4E, one-way ANOVA, $P < 0.0001$, $F_{(3,669)} = 13.99$ with Student's paired $t$ test, A1 vs. A1 $P < 0.0001$, mean (±SD) MI values: $MIOri = 0.68 ± 0.37$ bits, $MIOri = 0.62 ± 0.34$ bits, $MIOri = 0.59 ± 0.31$ bits and $MIOri = 0.42 ± 0.18$ bits], whereas the mean MI values in VRB remained unchanged in all conditions [Fig. 4F, one-way ANOVA, $P = 0.26$, $F_{(3,164)} = 1.34$, mean (±SD) MI values: $MIOri = 0.55 ± 0.29$ bits, $MIOri = 0.55 ± 0.28$ bits, $MIOri = 0.51 ± 0.22$ bits and $MIOri = 0.42 ± 0.17$ bits].

What can be the scenarios explaining the decrease in neuronal discrimination and involving the envelope tracking in situations of acoustic degradations? At least two scenarios can be envisioned. First, in a condition-independent envelope tracking scenario, a neuron retains the same intrinsic capacity to track the stimulus envelopes whatever the acoustic conditions (i.e. in both quiet conditions and in conditions of acoustic degradations). In that case, as long as the stimulus envelopes present some differences, the neuron will detect these differences and will discriminate the stimuli. Second, in a condition-dependent scenario, the acoustic degradations reduce the ability of neurons to track the stimulus envelope. In that case, despite differences between the stimulus envelopes, the intense activity occurring when the neurons are strongly driven by the noise prevents the recorded neuron from tracking the stimulus envelopes. To determine which of these two scenarios operates, we investigated whether the $R_{max\_PSTH}$ values were changed in the conditions of acoustic degradations, such as the vocoding or the noise addition (Figs. 5 and 6). Figure 5 shows, for individual recordings, the superpositions of the PSTH
Figure 4. Neuronal discrimination performance in all degraded conditions along the auditory system
A–F, neuronal discrimination performance [quantified by the mutual information (MI), in bits] in the original conditions (Ori) and in the three situations of acoustic degradation (top panels, vocoding; middle panels, stationary noise; and bottom panels, chorus noise). In each box plot, the horizontal line corresponds to the median value and the boxes correspond to the interquartile ranges. For all structures except in simulated auditory nerve fibres (sANF), note the largest decrease in the MI value in the stationary noise in the subcortical structures compared with the relative stability of these values in vocoding and chorus noise. Note also the much smaller decreases observed at the cortical level in the three situations of acoustic alterations. The asterisks represent the first significant differences between the mean original values and those obtained in degraded conditions. Note that for the sake of clarity, not all significant differences are indicated by asterisks. The decrease was significant only for the 10-band vocoded vocalizations (Voc10) in the ventral division of the medial geniculate (MGv) and primary auditory cortex (A1) (D, one-way ANOVA, \( F_{(3,668)} = 13.99 \), with Student’s paired \( t \)-test, MGvOri vs. MGvVoc10, \( P < 0.0001 \); E, one-way ANOVA, \( F_{(3,669)} = 3.37 \) with Student’s paired \( t \)-test, A1Ori vs. A1Voc10, \( P < 0.0001 \)), and no significant difference was detected in VRB (\( F_{(3,812)} = 1.06 \)).

In stationary noise, the mean MI value in sANF, CN and MGv was significantly reduced already at +10 dB SNR (A, one-way ANOVA, \( P < 0.0001 \); \( F_{(3,1153)} = 2.58 \) with Student’s paired \( t \)-test, sANFOri vs. sANFVoc10, \( P < 0.0001 \); B, one-way ANOVA, \( P < 0.0001 \); \( F_{(3,7722)} = 3.73 \) with Student’s paired \( t \)-test, A1Ori vs. A1Voc10, \( P < 0.0001 \)), and no significant difference was detected in VRB (\( F_{(3,186)} = 0.41 \)). The decrease was already significant with 38-band vocoded vocalizations (Voc38) in simulated auditory nerve fibres (sANF) or with 20-band vocoded vocalizations in the cochlear nucleus (CN) and the central nucleus of the inferior colliculus (CNIC) (A, one-way ANOVA, \( P < 0.0001 \); \( F_{(3,1302)} = 111.3 \) with Student’s paired \( t \)-test, sANFOri vs. sANFVoc38, \( P < 0.0001 \); B, one-way ANOVA, \( P < 0.0001 \); \( F_{(3,1424)} = 12.42 \) with Student’s paired \( t \)-test, CNICOri vs. CNICVoc20, \( P < 0.0001 \); C, one-way ANOVA, \( P < 0.0001 \); \( F_{(3,1231)} = 13.17 \) with Student’s paired \( t \)-test, CNICOri vs. CNICVoc20, \( P < 0.0001 \)). Note that there was also a significant increase in MI values with the 38-band vocoded vocalizations in CN (\( F_{(3,124)} = 12.42 \) with Student’s paired \( t \)-test, CNICOri vs. CNICVoc38, \( P = 0.0073 \)). In chorus noise, in CN and CNIC there was no significant decrease in mean MI values (B, one-way ANOVA, \( P = 0.05 \); \( F_{(3,1176)} = 2.65 \); C, one-way ANOVA, \( P = 0.36 \); \( F_{(3,1188)} = 1.06 \)), whereas in sANF and MGv, the mean MI values decreased significantly at +10 or 0 dB signal-to-noise ratio (SNR), respectively (A, one-way ANOVA, \( P < 0.0001 \); \( F_{(3,1331)} = 23.82 \) with Student’s paired \( t \)-test, sANFOri vs. sANFVoc10, \( P < 0.0001 \); D, one-way ANOVA, \( P < 0.0001 \); \( F_{(3,753)} = 7.3 \) with Student’s paired \( t \)-test, MGvOri vs. MGvVoc10, \( P < 0.0001 \)). At the cortical level, there was a significant decrease in A1 at 0 dB SNR and no significant change of mean MI values in VRB (\( F_{(3,697)} = 4.5 \) with Student’s paired \( t \)-test, A1Ori vs. A1Voc10, \( P < 0.0001 \); F, one-way ANOVA, \( P = 0.31 \); \( F_{(3,179)} = 1.19 \)). In stationary noise, the mean MI value in sANF, CN and MGv was significantly reduced already at +10 dB SNR (A, one-way ANOVA, \( P < 0.0001 \); \( F_{(3,1153)} = 767.64 \) with Student’s paired \( t \)-test, sANFOri vs. sANFVoc10, \( P < 0.0001 \); B, one-way ANOVA, \( P < 0.0001 \); \( F_{(3,31)} = 61.22 \) with Student’s paired \( t \)-test, CNICOri vs. CNICVoc10, \( P < 0.0001 \); D, one-way ANOVA, \( P < 0.0001 \); \( F_{(3,630)} = 62.03 \) with Student’s paired \( t \)-test, MGvOri vs. MGvVoc10, \( P < 0.0001 \)), whereas the mean MI value in CNIC was significantly reduced at 0 dB SNR (C, one-way ANOVA, \( P < 0.0001 \); \( F_{(3,1078)} = 32.08 \) with Student’s paired \( t \)-test, CNICOri vs. CNICVoc10, \( P < 0.0001 \)). At the cortical level, stationary noise significantly reduced the mean MI value in A1 only at −10 dB SNR (\( F_{(3,669)} = 13.99 \) with Student’s paired \( t \)-test, A1Ori vs. A1−10db, \( P < 0.0001 \)), whereas the mean MI values in VRB remained unchanged in all conditions (\( F_{(3,164)} = 1.34 \)). [Colour figure can be viewed at wileyonlinelibrary.com]
and the envelope (E) in the low AM range for the original conditions and all the degraded conditions (vocoding, stationary and chorus noise) in each auditory structure. The $\text{R}_{\text{maxE-PSTH}}$ values are indicated at the top left of each panel. In these individual recordings, the $\text{R}_{\text{maxE-PSTH}}$ values presented very little changes in all degraded conditions compared with the original conditions.

We next quantified, for each recording, the $\Delta \text{R}_{\text{maxE-PSTH}}$ variations compared with the original conditions ($\Delta \text{R}_{\text{maxE-PSTH}}$). This was quantified in each structure, for all the degraded conditions and each AM range (Fig. 6). Compared with the $\text{R}_{\text{maxE-PSTH}}$ values obtained in the original conditions, there was little or no change in the degraded conditions for all structures. More precisely, in sANF and CN, we observed a maximal increase in mean (±SD) $\text{R}_{\text{maxE-PSTH}}$ values of 0.16 (±0.03) [and 0.11 (±0.13) for CN] and a maximal decrease of 0.10 (±0.04) [and 0.05 (±0.06) for CN] depending on the degraded conditions and the AM range (Fig. 6). In CNIC and MGv, the mean (±SD) $\Delta \text{R}_{\text{maxE-PSTH}}$ changes in degraded conditions were very small [Fig. 6; between $-0.06$ (±0.03) and 0.006 (±0.06) for CNIC].

![Figure 5. Individual examples of correlations between neuronal responses and envelopes in all acoustic conditions and all structures. Note that for the subcortical and cortical structures, we present the results from the low amplitude modulation (AM) range. The correlation value between the peristimulus time histogram (PSTH; in black) and the envelope (E; in red) is shown at the top left of each panel. In all structures, the correlation values for these individual recordings remained similar between the acoustic conditions. [Colour figure can be viewed at wileyonlinelibrary.com]](image-url)
and between $-0.08$ (±0.08) and $-0.0002$ (±0.15) for MGv. In A1, the changes in mean (±SD) $R_{\text{maxE-PSTH}}$ values varied between $-0.09$ (±0.11) and $0.07$ (±0.15), and in VRB they varied between $-0.13$ (±0.13) and $0.07$ (±0.15).

These results clearly provide evidence that the abilities of neurons for tracking the temporal envelope cues were preserved at each level of the auditory system and in all the situations of acoustic degradations.

The increase in between-envelope similarity explains the decrease in neuronal discrimination

If the neurons are still able to track the stimulus envelopes in all conditions of acoustic degradations, what could explain the pronounced decrease in MI in these situations? The most parsimonious explanation is that the addition of noise increases the similarity between stimulus envelopes, which in turn reduces the neuronal discriminative efficiency based on the envelope tracking.

Figure 6. The envelope tracking is only slightly affected by the different situations of acoustic degradation

Mean (±SD) changes of the maximal correlation coefficient out of the seven selected gammatone filters ($R_{\text{maxE-PSTH}}$) quantified from the original conditions in the different situations of acoustic degradations [vocoding, stationary noise (SN) and chorus noise (ChN)] for all recordings obtained in the six structures. Note that, in each structure and amplitude modulation (AM) range, the $R_{\text{maxE-PSTH}}$ values were changed only slightly between the original and the degraded conditions. In simulated auditory nerve fibres (sANF) and the cochlear nucleus (CN), we observed a maximal increase in mean (±SD) $R_{\text{maxE-PSTH}}$ values of 0.16 (±0.03) [and 0.11 (±0.13) for CN] and a maximal decrease of 0.10 (±0.04) [and 0.05 (±0.06) for CN] depending on the degraded conditions and the AM range. In the central nucleus of the inferior colliculus (CNIC) and the ventral division of the medial geniculate (MGv), the mean (±SD) $R_{\text{maxE-PSTH}}$ changes in degraded conditions were very small [between $-0.06$ (±0.03) and 0.006 (±0.06) for CNIC and between $-0.08$ (±0.08) and $-0.0002$ (±0.15) for MGv]. In the primary auditory cortex (A1), the changes in mean (±SD) $R_{\text{maxE-PSTH}}$ values varied between $-0.09$ (±0.11) and 0.07 (±0.15), and in the ventrorostral belt (VRB) they varied between $-0.13$ (±0.13) and 0.07 (±0.15). [Colour figure can be viewed at wileyonlinelibrary.com]
We thus quantified the acoustic similarity between the stimulus envelopes in the original conditions and in all situations of acoustic degradations. The changes of the between-envelope similarity in each AM range are presented in Fig. 7A. In the M and H ranges, the between-stimulus envelope similarity was low and remained low in all degraded conditions. In contrast, large changes occurred in the low AM range; in these frequency ranges, the envelope similarity increased progressively with the acoustic degradations. In the following results, we will focus on this AM range.

In the vocoding conditions, the similarity between the four whistle envelopes was relatively constant, except for the 10-band vocoded conditions, where this similarity was slightly higher. In the stationary noise, the four stimulus envelopes became similar and reached a correlation value $>0.8$ in the $-10$ dB SNR conditions (which is very close to the maximal value of the acoustic similarity). In the chorus noise conditions, the four stimulus envelopes remained different (because spectrotemporal differences were present in the frozen chorus noise), with the highest similarity in the $-10$ dB SNR conditions.

Figure 7B illustrates that in the conditions where the between-stimulus envelope similarity was higher (at $-10$ dB SNR in stationary noise), the envelope tracking remained similar (the $\Delta R_{max E-PSTH}$ values remained stable), whereas the neuronal discrimination decreased compared with the original conditions (most of the $\Delta MI$ values were largely negative). This clearly demonstrates the dissociation between changes in the value $R_{max}$ and those in the value of MI. Figure 7C highlights the close relationship between the acoustic similarity of the four stimulus envelopes and the abilities of auditory neurons to discriminate between them. In both subcortical and cortical structures, as the acoustic distance between the four stimulus envelopes in the low AM range progressively decreased, the neuronal discrimination decreased (Fig. 7C).

Together, these results indicate that it is not a loss in neuronal envelope tracking that leads to a reduction of the neuronal discriminative abilities in the degraded conditions. Instead, it is the increase in envelope similarity in situations of acoustic degradations that is one of the important factors responsible for the decrease in discrimination abilities. Thus, the between-stimulus envelope similarity in the lower AM range ($<20$ Hz) can predict the evolution of discrimination in the entire auditory system.

The increase in between-envelope similarity is also correlated with the behavioural performance in noise

To examine whether the discrimination performance of auditory neurons might provide a neuronal basis for behavioural performance, we tested whether behaving animals can discriminate between whistles when engaged in an operant conditioning task involving the same stimuli. We opted to train mice rather than guinea-pigs in a behavioural task for two main reasons: (i) guinea-pigs are poor and slow learners in instrumental tasks; and (ii) this avoided the possibility that the stimuli used for the behavioural task have innate particular meanings, because whistles are alert signals for guinea-pigs.

The behavioural task was a go/no-go task involving the discrimination between two of the four whistles used in our electrophysiological studies (W1 and W3; see Fig. 1): Licks to the S+ were rewarded by a 5 $\mu$l drop of water and licks to the S− were punished by a 5 s time-out period. Mice were first trained for 5–10 initial sessions to perform the discrimination in the original conditions until they reached 80% of correct responses for two successive days ($n = 9$). Then, the mice were sequentially trained in the stationary noise at the +10, 0 and $-10$ dB SNR for at least four sessions. The performance in the last four sessions at each SNR is displayed in Fig. 7D. For all mice, the average performance decreased at the 0 and $-10$ dB SNR; although two mice were still at 80% of correct performance, the others were slightly above the chance level. In the chorus noise, the performance of most of the mice was relatively stable, which can be explained by the fact that acoustically, the chorus noise surrounding the two target vocalizations differed between the two whistles, meaning that there were more acoustic cues to discriminate between the target stimuli in these conditions (note that this could also stem from the fact that the mice were already extensively trained to perform the discrimination task in stationary noise when they started the chorus noise). Despite this pitfall, the main result of this behavioural study was that mice can discriminate the target vocalizations above chance level even at $-10$ dB SNR in stationary noise. Furthermore, the decrease in behavioural performance was strongly related to the reduction of the differences between the two temporal envelopes in the low AM range (inset in Fig. 7D). These results provide evidence that the behavioural performance of mice is correlated with the changes in the slow temporal envelope cues.

Discussion

Our first major result is that the neuronal discrimination performance in the original conditions was correlated with the capacity for tracking the envelopes in the low AM range for both subcortical and cortical neurons, except in the secondary auditory cortex (VRB; Fig. 3C). Our second major result is that, in acoustically degraded conditions and in each structure, the ability for envelope tracking changed only slightly compared with the original
Figure 7. In all situations of acoustic alteration, the decrease in neuronal discrimination performance can be explained by the increase in envelope similarity in the low range.

A, acoustic similarity (R_{Env}) between the envelopes of the four whistles in the original conditions (Ori) and in the three situations of acoustic alterations (vocoding, stationary noise and chorus noise) for the low (L, red lines), middle (M, yellow lines) and high (H, purple lines) amplitude modulation (AM) ranges. Dark lines correspond to the R_{Env} values based on the seven selected gammatones, whereas the light lines correspond to the R_{Env} values based on 35 gammatones. Note that in the stationary noise, the correlation between the stimulus envelopes largely increased in the L range, indicating that the stimuli tended to be similar to each other in these AM ranges, which was not the case in the middle and high (M and H) ranges. This between-stimuli increase in correlation in the L range was much weaker in the vocoding and chorus noise situations.

B, scattergrams showing the variation of the maximal correlation (ΔR_{max,PSTH}) in the low AM range as a function of the variation of mutual information (ΔMI) in the −10 dB signal-to-noise ratio (SNR) conditions compared with the original conditions in each structure. C, mean changes (ΔMI, as a percentage) of mutual information in simulated auditory nerve fibres (sANF), cochlear nucleus (CN), central nucleus of the inferior colliculus (CNIC), ventral division of the medial geniculate (MGv), primary auditory cortex (A1) and ventrorostral belts (VRB) as a function of the variation (ΔR_{Env}, as a percentage) of the acoustic similarity in the low AM range relative to the original conditions. Each dot represents neuronal data (ΔMI) in sANF (in dark red), CN (in black), CNIC (in green), MGv (in orange), A1 (in blue) and VRB (in purple). From left to right, all degraded acoustic conditions were organized according to the acoustic distance of the envelopes (R_{Env}) between the four whistles quantified in A (+10 dB SNR, chorus noise, Voc38, Voc20, +10 dB SNR, stationary noise, Voc0).
middle- and high-SR fibres could have changed our results. Therefore, a different ratio between low-, middle- and high-SR fibres that have higher tones. Low- (and middle-) SR fibres that have higher eliminates the ability of the fibre to respond to tested approaches the saturation rate of the fibre and ultimately noise. Thus, as the noise level increases, the discharge rate reduces in performance were observed for 0 and −10 dB SNRs in the stationary noise. The inset shows that the decrease in behavioural performance (average across sessions and animals) was strongly related to the reduction in the differences between the two temporal envelopes (W1 and W3) in the low AM range. Colour figure can be viewed at wileyonlinelibrary.com

conditions (Figs. 5, 6 and 7B). Finally, our findings reveal that the increased similarity between the stimulus envelopes in the low AM range (<20 Hz; Fig. 7C–D) is one of the important factors responsible for the decrease in neuronal and behavioural discrimination.

Slow envelope tracking: a general property of auditory neurons

At the level of the auditory nerve, previous studies reported conflicting results concerning noise resistance. Frisina and colleagues (1996) found that all AN units partially preserve their AM coding, even in the presence of loud (0 or +6 dB SNR) background noise. However, many other electrophysiological studies, and the present simulated data, showed a low resistance to noise in the auditory nerve (Costalupes, 1985; Costalupes et al., 1984; Geisler & Sinex, 1980; Palmer & Evans, 1982; Rhode et al., 1978; Young & Barta, 1986). This can be explained by several factors, including: (i) the type of target stimuli (artificial vs. natural stimuli); (ii) the noise type (stationary vs. non-stationary noise); (iii) the type of auditory nerve fibres (low-, middle- and high-SR fibres); and (iv) the noise levels that have been tested. For example, electrophysiological studies have shown that low- and medium-SR fibres with best frequencies around the frequency of a pure tone exhibited tone-evoked rate changes in the presence of a stationary noise at positive SNRs (Costalupes, 1985; Costalupes et al., 1984; Geisler & Sinex, 1980; Palmer & Evans, 1982; Rhode et al., 1978; Young & Barta, 1986). High-SR fibres, in contrast, exhibited much weaker tone-evoked rate changes at positive SNRs limited by the high-rate response to noise. Thus, as the noise level increases, the discharge rate approaches the saturation rate of the fibre and ultimately eliminates the ability of the fibre to respond to tested tones. Low- (and middle-) SR fibres that have higher thresholds and a wider dynamic range are significantly more resistant to saturation by high noise levels than high-SR fibres. Therefore, a different ratio between low-, middle- and high-SR fibres could have changed our results in such a way that more sANF responses showed a higher resistance to noise. Here, given that we wanted to be as close as possible to the multi-unit activity recorded in the auditory structures (we assumed that we recorded about five neurons under each electrode), we decided to choose five fibres with a classical ratio of one low-SR, one medium-SR and three high-SR fibres. However, we should bear in mind that all these previous studies have used tones (or amplitude-modulated tones) in noise at positive SNRs, but natural sounds can potentially trigger more complex encoding as early as the auditory nerve. Recently, using a similar model of the auditory nerve to the one in the present study, Rabinowitz and colleagues (2013) showed poor adaptation to the noise statistics by simulated fibres when natural environmental sounds were used as target stimuli, whereas auditory cortex and IC neurons showed a better adaptation to noise. In our study, we also found a high sensitivity to noise for sANF as early as +10 dB SNR for both noises (see examples in Fig. 5) and also in the vocoding conditions, potentially owing to a higher sensitivity to the spectrotemporal alterations compared with the other structures (see Fig. 4A).

A few electrophysiological studies have shown that subcortical neurons can display responses very close to the envelopes of natural stimuli (inferior colliculus: Rode et al., 2013; Suta et al., 2003; MGB: Philibert et al., 2005; Suta et al., 2007; Tanaka & Taniguchi, 1991). Rode and colleagues (2013) found that between 15 and 60% of collicular neurons displayed high correlations for at least one of the three vocalization envelopes, and a subset of collicular neurons even followed the envelopes of the three guinea-pig vocalizations with high correlations (>0.85). A similar range of correlations (between 0.6 and 0.9) in CNIC was obtained in the present study and, as in their study, we also did not find a relationship between the gammatone filter eliciting the highest RE-PSTH value and the best frequency of the neurons.

Unlike other previous cortical studies (Abrams et al., 2017; Bar-Yosef et al., 2002; Grimsley et al., 2012; Nagarajan et al., 2002; Wang et al., 1995), we filtered envelopes and neuronal responses in the same frequency bands [from low (<20 Hz) to high (100 and 200 Hz) ranges] to obtain a direct quantification of the envelope tracking abilities in particular frequency ranges. Furthermore, we compared the degree of envelope
tracking performed by subcortical and cortical neurons in challenging situations where the envelope is either relatively well preserved or strongly degraded. Nagarajan and colleagues (2002) found that the synchronization between A1 responses and the temporal envelope of vocalizations was highly significant and, interestingly, this property was underestimated based on responses to amplitude-modulated tones. In addition, they pointed out that A1 responses were fairly resistant to spectral degradations (generated by a noise-vocoder) and to noise addition up to 0 dB SNR. More importantly, the responses were similar when the vocalization envelope was preserved between 2 and 30 Hz, whereas the responses were strongly reduced when the envelope was low-pass filtered at 4 or 10 Hz. We confirmed these cortical results in several aspects: (i) the highest correlation coefficients were detected in the lower AM range (<20 Hz) for each acoustic condition; and (ii) we showed that the envelope tracking ability was little affected by the presence of noise addition or by the vocoding. We extended these results to a non-primary cortical area (VRB), to each subcortical level and even to sANF. Note that the envelope tracking ability is not specific to the processing of conspecific vocalizations; similar results were found with speech in noise in the auditory cortex of guinea-pigs (Abrams et al., 2017).

Together, these results highlight that subcortical and cortical auditory neurons maintain their capacity to track the slow envelope of natural sounds both when they are composed of noise-free vocalizations or of a mixture of noise and vocalizations, suggesting that this property is immutable and unchanged by the acoustic degradations.

In the low AM range (<20 Hz), we noticed a decrease in mean correlation ($R_{maxE-PSTH}$) values from mid-brain to thalamus to cortex (Fig. 2F), reflecting that the further away from the periphery, the less precise is the phase-locking ability on the AM cues. For higher AM rates, we expected higher correlations between the neuronal responses and the envelopes for subcortical structures (Creutzfeldt et al., 1980; Frisina et al., 1990; Neuert et al., 2001; Rhode & Greenberg, 1994; for review, see Joris et al., 2004). Surprisingly, such a hierarchy was not detected in our results, with the mean correlations in higher AM rates (>20 Hz) being similarly low for each structure, including the sANF. These lower correlation coefficients obtained for the middle and high AM frequency bands for all structures might result because the envelopes have much lower amplitudes in these bands than in the low AM frequency band (see Fig. 1C). Another hypothesis is that shorter segments of neuronal responses could be highly correlated with the higher AM ranges of the envelopes. If so, reducing the time window in which the correlation is computed should increase the correlations in the higher AM ranges. We computed the cross-correlation for each whistle (∼300 ms) and still found low correlations in higher AM ranges (data not shown). This suggests that if higher correlations exist in higher AM ranges, smaller temporal windows (less than several hundreds of milliseconds) are required to reveal them. The fact that Abrams and colleagues (2017) found some residues of the fundamental frequency (between 100 and 120 Hz, relative to the pitch) in segments of A1 responses no longer than 100 ms argues in favour of this possibility.

The main hypothesis of our study was that the tracking abilities of auditory neurons is one of the mechanisms explaining neuronal discrimination. Another possibility is that for higher AM cues, some auditory neurons respond by increasing their firing rate. This hypothesis relies on the existence of a rate–place code for periodicity. The temporal tracking abilities decrease along the auditory pathway, whereas higher periodicities can be encoded by a rate–place code, because this has been demonstrated using amplitude-modulated sounds in different species (Langner, 1992; Langner & Schreiner, 1988; Liang et al., 2002; Lu & Wang, 2004; Lu et al. 2001a, 2001b; Schreiner & Langner, 1988). According to this possibility, neurons increasing their firing rate for coding higher AM cues should be located at particular locations in the IC and auditory cortex (Langner et al., 2009; Schnupp et al., 2015). However, in order to explain the neuronal discrimination, it seems necessary that each of the four whistles activated different locations in the periodicity maps of the different auditory structures. As shown in Fig. 1C, each of the four whistles contained about the same energy in low, middle and high AM modulations and, as a consequence, similar locations should be activated in these periodicity maps, leading to a low discrimination level. Thus, although we cannot discard this hypothesis, the possibility that the neural discrimination relies on a rate–place code for particular AM cues seems unlikely. At the cortical level (both in A1 and in VRB), it is also possible that despite the fact individual neurons cannot keep tracking the detailed envelope fluctuations (because of their low-pass properties regarding AM cues and their prominent onset responses), they might, as a large population, track the envelope changes if each neuron is sensitive to a particular rate of change of the stimulus envelope (a particular rate of transients). Note that according to this hypothesis, which was formulated almost 20 years ago (Heil, 2003), this tracking mechanism would also lose accuracy with increasing levels of background noise.

The decrease in neuronal discrimination can be explained by the increase of between-envelope similarities in the low AM range

In the original conditions, the better neurons track the slow envelope (<20 Hz), the higher the neuronal
discrimination performance for all structures (Fig. 3B and C). In situations of acoustic degradation, the envelopes of the original stimuli were altered, leading to situations where the envelopes were mostly dominated by the noise envelopes. However, the three situations of acoustic degradations used here differed notably. In the tone-vocoder situation, the spectral content is strongly degraded but the slow temporal envelope is relatively well preserved (Kates, 2011; Shannon et al., 1995; Souffi et al., 2020). In the chorus noise, there was only a small increase in acoustic similarity in the low AM range (Fig. 7A), because the chorus noise itself contains strong temporal variations that differ from one whistle to another. As a consequence of this pitfall, when the target vocalizations were inserted into the chorus noise, specific regions in the spectrotemporal domain were dominated by the target vocalizations, whereas in other regions it was dominated by the chorus noise. Consequently, the target vocalizations embedded in the chorus noise generated stimuli that can be discriminated at all SNRs either based on the vocalization envelopes or based on the chorus noise envelope itself. In all structures, the neuronal discrimination showed little decrease in the chorus noise (see Fig. 4), as did the behavioural performance (Fig. 7D). Only in the stationary noise did the four slow envelopes become closer as the level of degradation increased (<20 Hz; see Fig. 7A). This would be detrimental for discriminating vocalizations in noisy conditions, in which envelope tracking becomes inefficient and, worse, could strongly reduce the neuronal discrimination along the auditory system. Therefore, reducing or increasing the envelope differences in the low AM range would constrain or facilitate the neuronal discrimination in subcortical and cortical levels. Furthermore, the behavioural performance of mice revealed that they could discriminate the target vocalizations in quiet conditions (with >90% correct performance) and could even discriminate the vocalizations up to 0 dB SNR in stationary noise (with 70–80% correct performance), suggesting that the between-stimulus envelope differences could explain the behavioural performance during a discrimination task. Previous studies have reported good behavioural discrimination performance in conditions of acoustic degradations such as vocoded consonants or vowels (Ranasinghe et al., 2012), consonants in various levels of background noise (Shetake et al., 2011), bird songs embedded in stationary noise and chorus noise (Narayan et al., 2007) or in broadband dynamic moving ripples (Hommma et al., 2020). In all these studies, the discrimination performance of auditory cortex neurons, based upon spike timing, has been found to match the behavioural performance relatively well (Hommma et al., 2020; Narayan et al., 2007; Ranasinghe et al., 2012) and sometimes even with performance of human subjects (Walker et al., 2008). Altogether, our results indicate that it is not a loss in neuronal envelope tracking that leads to a reduction of the neuronal discriminative abilities in the degraded conditions; instead, it is the direct consequence of the changes in acoustic distance between stimulus envelopes.

Comparison with human studies: the case of newborn infants

The speech envelope corresponds to the slow amplitude fluctuations of the signal over time, with peaks occurring roughly at the syllabic rate. The two pioneer results supporting the view that the envelope plays a key role in speech comprehension are: (i) that comprehension is impaired when the speech envelope is filtered out (Druyllman et al. 1994a, 1994b); and (ii) that adult listeners readily understand degraded speech in which only the envelope is preserved, at least when speech is presented in silence (Shannon et al., 1995). Additionally, studies have shown that when adults listen to speech, their neuronal activity synchronizes with specific features of the envelope, a phenomenon known as speech envelope tracking (Abrams et al., 2008; Ahissar et al., 2001; Luo & Poeppel, 2007; Nourski et al., 2009). Several recent electrophysiological results have provided new insights into this putative speech envelope tracking mechanism. First, oscillations whose frequency corresponds to the modulation frequency of the speech envelope (4–5 Hz) have been found to be independent of comprehension. Brain responses in the theta band track the speech envelope even when speech is time compressed at a rate that renders it incomprehensible for adult listeners (Kösem & Van Wassenhove, 2017; Kösem et al., 2016; Pefkou et al., 2017; Zoefel & Van Rullen, 2016). Second, results from newborns and young infants have also provided new insights. For example, combining haemodynamic (near-infrared spectroscopy) and EEG recordings, Cabrera and Gervain (2020) showed that infants (9–10 months old) detect consonant changes on the basis of envelope cues (without the temporal fine structure) and they can even do so on the basis of the slow temporal variation alone (AM <8 Hz). More recently, Ortiz-Barajas and colleagues (2021) found that the cortical networks of newborns (exclusively exposed to French before birth) have the capacity to track the amplitude and the phase of the speech envelope in their native languages and in unfamiliar languages (Spanish and English). Altogether, these results suggest that amplitude and phase tracking take place in the absence of attention and comprehension.

Thus, envelope tracking can be viewed as a universal mechanism used in all species to discriminate between communication sounds in a large diversity of acoustic situations ranging from quiet to adverse, challenging conditions.
References

Abrams, D. A., Nicol, T., Zecker, S., & Kraus, N. (2008). Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. *Journal of Neuroscience, 28*(15), 3958–3965.

Abrams, D. A., Nicol, T., White-Schwoch, T., Zecker, S., & Kraus, N. (2017). Population responses in primary auditory cortex simultaneously represent the temporal envelope and periodicity features in natural speech. *Hearing Research, 348*, 31–43.

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., & Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America, 98*(23), 13367–13372.

Anderson, L. A., Wallace, M. N., & Palmer, A. R. (2007). Identification of subdivisions in the medial geniculate body of the guinea pig. *Hearing Research, 228*(1–2), 156–167.

Aushana, Y., Souffi, S., Edeline, J. M., 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. *Journal of the Association for Research in Otolaryngology, 19*(2), 163–180.

Bar-Yosef, O., Rotman, Y., & Nelken, I. (2002). Responses of neurons in cat primary auditory cortex to bird chirps: Effects of temporal and spectral context. *Journal of Neuroscience, 22*(19), 8619–8632.

Bruce, I. C., Erfani, Y., & Zilany, M. S. A. (2018). A phenomenological model of the synapse between the inner hair cell and auditory nerve: Implications of limited neurotransmitter release sites. *Hearing Research, 360*, 40–54.

Cabrera, L., & Gervain, J. (2020). Speech perception at birth: The brain encodes fast and slow temporal information. *Science Advances, 6*(30), eaba7830.

Creutzfeldt, O., Hellweg, F. C., & Schreiner, C. (1980). Thalamocortical transformation of responses to complex auditory stimuli. *Experimental Brain Research, 39*(1), 87–104.

Ding, N., Melloni, L., Yang, A., Wang, Y., Zhang, W., & Poeppel, D. (2017). Characterizing neural entrainment to hierarchical linguistic units using electroencephalography (EEG). *Frontiers in Human Neuroscience, 11*, 481.

Ding, N., Chatterjee, M., & Simon, J. Z. (2014). Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. *Neuroimage, 88*, 41–46.

Drullman, R., Festen, J. M., & Plomp, R. (1994a). Effect of reducing slow temporal modulations on speech reception. *Journal of the Acoustical Society of America, 95*(5), 2670–2680.

Drullman, R., Festen, J. M., & Plomp, R. (1994b). Effect of temporal envelope smearing on speech reception. *The Journal of the Acoustical Society of America, 95*(2), 1053–1064.

Edeline, J. M., 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*(1–2), 135–152.

Franke, F., Quian, Q. R., Hierlemann, A., & Obermayer, K. (2015). Bayes optimal template matching for spike sorting - combining fisher discriminant analysis with optimal filtering. *Journal of Computational Neuroscience, 38*(3), 439–459.

Frisina, R. D., Karcich, K. J., Tracy, T. C., Sullivan, D. M., Walton, J. P., & Colombo, J. (1996). Preservation of amplitude modulation coding in the presence of background noise by chinchilla auditory-nerve fibers. *Journal of the Acoustical Society of America, 99*(1), 475–490.

Frisina, R. D., Smith, R. L., & Chamberlain, S. C. (1990). Encoding of amplitude modulation in the gerbil cochlear nucleus: I. A hierarchy of enhancement. *Hearing Research, 44*(2–3), 99–122.

Gaucher, Q., & Edeline, J. M. (2015). Stimulus-specific effects of noradrenaline in auditory cortex: Implications for the discrimination of communication sounds. *The Journal of Physiology, 593*(4), 1003–1020.

Gaucher, Q., Edeline, J. M., & Gourévitch, B. (2012). How different are the local field potentials and spiking activities? Insights from multi-electrodes arrays. *Journal of Physiology, Paris, 106*(3–4), 93–103.

Gaucher, Q., Huetz, C., Gourévitch, B., & Edeline, J. M. (2013). Cortical inhibition reduces information redundancy at presentation of communication sounds in the primary auditory cortex. *Journal of Neuroscience, 33*(26), 10713–10728.

Geisler, C. D., & Sinex, D. G. (1980). Responses of primary auditory fibers to combined noise and tonal stimuli. *Hearing Research, 3*(4), 317–334.

Geisler, C. D., & Sinex, D. G. (1980). Responses of primary auditory fibers to combined noise and tonal stimuli. *Hearing Research, 3*(4), 317–334.

Gnansia, D., Péan, V., Meyer, B., & Lorenzi, C. (2009). Effects of spectral smearing and temporal fine structure degradation on speech masking release. *Journal of the Acoustical Society of America, 125*(6), 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*(1–2), 46–53.
Gourévitch, B., Doisy, T., Avillac, M., & Edeline, J. M. (2009). Age-related changes in the guinea pig auditory cortex: Relationship with brainstem changes and comparison with tone-induced hearing loss. European Journal of Neuroscience, 34(12), 1953–1965.

Gourévitch, B., Doisy, T., Avillac, M., & Edeline, J. M. (2009). Follow-up of latency and threshold shifts of auditory brainstem responses after single and interrupted acoustic trauma in guinea pig. Brain Research, 1304, 66–79.

Grimsley, J. M., Palmer, A. R., & Wallace, M. N. (2011). Different representations of tooth chatter and purr call in guinea pig auditory cortex. NeuroReport, 22(12), 613–616.

Grimsley, J. M. S., Shanbhag, S. J., Palmer, A. R., & Wallace, M. N. (2012). Processing of communication calls in guinea pig auditory cortex. Plos One, 7(12), e51646.

Harris, K. D. (2020). Nonsense correlations in neuroscience. bioRxiv. 2020.11.29.402719; doi: https://doi.org/10.1101/2020.11.29.402719. bioRxiv.

Heil, P. (2003). Coding of temporal onset envelope in the auditory system. Speech Communication, 41(1), 123–134.

Hohmann, V. (2002). Frequency analysis and synthesis using a Gammatone filterbank. Acta Acustica United with Acustica, 88, 433–442.

Homma, N. Y., Hullett, P. W., Atencio, C. A., & Schreiner, C. E. (2020). Auditory cortical plasticity dependent on environmental noise statistics. Cell Reports, 30(13), 4445–4458.e5.

Joris, P. X., Schreiner, C. E., & Rees, A. (2004). Neural processing of amplitude-modulated sounds. Physiological Reviews, 84(2), 541–577.

Kates, J. M. (2011). Spectro-temporal envelope changes caused by temporal fine structure modification. The Journal of the Acoustical Society of America, 129(6), 3981–3990.

Kösem, A., & Van Wassenhove, V. (2017). Distinct contributions of low- and high-frequency neural oscillations to speech comprehension. Language, Cognition and Neuroscience, 32(5), 536–544.

Kösem, A., Basirat, A., Azizi, L., & van Wassenhove, V. (2016). High-frequency neural activity predicts word parsing in ambiguous speech streams. Journal of Neurophysiology, 116(6), 2497–2512.

Langner, G., & Schreiner, C. E. (1988). Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. Journal of Neurophysiology, 60(6), 1799–1822.

Langner, G. (1992). Periodicity coding in the auditory system. Hearing Research, 60(2), 115–142.

Langner, G., Dinse, H. R., & Godde, B. (2009). A map of periodicity orthogonal to frequency representation in the cat auditory cortex. Frontiers in Integrative Neuroscience, 3, 27.

Liang, L., Lu, T., & Wang, X. (2002). Neural representations of sinusoidal amplitude and frequency modulations in the primary auditory cortex of awake primates. Journal of Neurophysiology, 87(5), 2237–2261.

Lu, T., Liang, L., & Wang, X. (2001a). Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. Nature Neuroscience, 4(11), 1131–1138.

Lu, T., Liang, L., & Wang, X. (2001b). Neural representations of temporally asymmetric stimuli in the auditory cortex of awake primates. Journal of Neurophysiology, 85(6), 2364–2380.

Luo, T., & Wang, X. (2004). Information content of auditory cortical responses to time-varying acoustic stimuli. Journal of Neurophysiology, 91(1), 301–313.

Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron, 54(6), 1001–1010.

Majdak, P., Hohomey, C., & Baumgartner, R. (2022). AMT 1.0: The toolbox for reproducible research in auditory modeling. Acta Acustica, 6, 19.

Malmiera, M. S., Rees, A., Le Beau, F. E., & Bjaalie, J. G. (1995). Laminar organization of frequency-defined local axons within and between the inferior colliculi of the guinea pig. Journal of Comparative Neurology, 357(1), 124–144.

Malmiera, M. S., Le Beau, F. E., & Rees, A. (1996). The topographical organization of descending projections from the central nucleus of the inferior colliculus in guinea pig. Hearing Research, 93(1–2), 167–180.

Nagarajan, S. S., Cheung, S. W., Bedenbaugh, P., Beitel, R. E., Schreiner, C. E., & Merzenich, M. M. (2002). Representation of spectral and temporal envelope of twitter vocalizations in common marmoset primary auditory cortex. Journal of Neurophysiology, 87(4), 1723–1737.

Narayan, R., Best, V., Ozmeral, E., McClaine, E., Dent, M., Shinn-Cunningham, B., & Sen, K. (2007). Cortical interference effects in the cocktail party problem. Nature Neuroscience, 10(12), 1601–1607.

Neuert, V., Pressnitzer, D., Patterson, R. D., & Winter, I. M. (2001). The responses of single units in the inferior colliculus of the guinea pig to damped and ramped sinusoids. Hearing Research, 159(1–2), 36–52.

Nourski, K. V., Reale, R. A., Oya, H., Kawasaki, H., Kovach, C. K., Chen, H., Howard, M. A. 3rd, & Brugge, J. F. (2009). Temporal envelope of time-compressed speech represented in the human auditory cortex. Journal of Neuroscience, 29(49), 15564–15574.

Occelli, F., Suied, C., Pressnitzer, D., Edeline, J. M., & Gourévitch, B. (2016). A neural substrate for rapid timbre recognition? Neural and behavioral discrimination of very brief acoustic vowels. Cerebral Cortex, 26(6), 2483–2496.

Ortiz Barajas, M. C., Guevara, R., & Gervain, J. (2021). The origins and development of speech envelope tracking during the first months of life. Developmental Cognitive Neuroscience, 48, 100915.

Osses, V. A., Varnet, L., Carney, L. H., Dau, T., Bruce, I. C., Verhulst, S., & Majdak, P. (2022). A comparative study of eight human auditory models of monaural processing. Acta Acustica, 6, 17.

Palmer, A. R., & Evans, E. F. (1982). Intensity coding in the auditory periphery of the cat: Responses of cochlear nerve and cochlear nucleus neurons to signals in the presence of bandstop masking noise. Hearing Research, 7(3), 305–323.

Paraouty, N., Stasiak, A., Lorenzi, C., Varnet, L., & Winter, I. M. (2018). Dual coding of frequency modulation in the ventral cochlear nucleus. Journal of Neuroscience, 38(17), 4123–4137.

Patterson, R. D. (1987). A pulse ribbon model of monaural phase perception. Journal of the Acoustical Society of America, 82(5), 1560–1586.
Pefkou, M., Arnal, L. H., Fontolan, L., & Giraud, A. L. (2017). $\theta$-band and $\beta$-band neural activity reflects independent syllable tracking and comprehension of time-compressed speech. *Journal of Neuroscience*, 37(33), 7930–7938.

Philibert, B., Laudanski, J., & Edeline, J. M. (2005). Auditory thalamus responses to guinea-pig vocalizations: A comparison between rat and guinea-pig. *Hearing Research*, 209(1–2), 97–103.

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. *Journal of Neurophysiology*, 91(6), 2910–2928.

Quiroga, R. Q., Nadasy, Z., & Ben-Shaul, Y. (2004). Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Computation*, 16(8), 1661–1687.

Rabinowitz, N. C., Willmore, B. D., King, A. J., & Schnupp, J. W. (2013). Constructing noise-invariant representations of sound in the auditory pathway. *Plos Biology*, 11(11), e1001710.

Ramachandran, R., Davis, K. A., & May, B. J. (2011). Rate representations coexist with robust cortical encoding of natural vocalizations. *Frontiers in Neural Circuits*, 5(2), 2874–2881.

Reiss, L. A., Sarbazi, A., Malmierca, M. S., & Le Beau, F. E. N. (1997). Regularity of firing of neurons in the inferior colliculus. *Journal of Neurophysiology*, 77(6), 2945–2965.

Redies, H., Brandner, S., & Creutzfeldt, O. D. (1989). Anatomy of the auditory thalamocortical system of the guinea pig. *The Journal of Comparative Neurology*, 282(4), 489–511.

Ranasinghe, K. G., Vrana, W. A., Matney, C. J., & Kilgard, M. P. (2012). Neural mechanisms supporting robust discrimination of spectrally and temporally degraded speech. *Journal of the Association for Research in Otolaryngology*, 13(4), 527–542.

Rhode, W. S., & Greenberg, S. (1994). Encoding of amplitude modulation in the cochlear nucleus of the cat. *Journal of Neurophysiology*, 71(5), 1797–1825.

Rhode, W. S., Geisler, C. D., & Kennedy, D. T. (1978). Auditory nerve fiber response to wide-band noise and tone combinations. *Journal of Neurophysiology*, 41(3), 692–704.

Rode, T., Hartmann, T., Hubka, P., Scheper, V., Lenarz, M., Lenarz, T., Kral, A., & Lim, H. H. (2013). Neural representation in the auditory midbrain of the envelope of vocalizations based on a peripheral ear model. *Frontiers in Neural Circuits*, 7, 166.

Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences*, 336(1278), 367–373.

Rutkowski, R. G., Shackleton, T. M., Schnupp, J. W. H., Wallace, M. N., & Palmer, A. R. (2002). Spectrottemporal receptive field properties of single units in the primary, dorsocaudal and ventrorostral auditory cortex of the guinea pig. *Audiology and Neurotology*, 7(4), 214–227.

Sayles, M., & Winter, I. M. (2010). Equivalent-rectangular bandwidth of single units in the anaesthetized guinea-pig ventral cochlear nucleus. *Hearing Research*, 262(1–2), 26–33.

Shamma, S., & Lorenzi, C. (2013). On the balance of envelope and temporal fine structure in the encoding of speech in the early auditory system. *Journal of the Acoustical Society of America*, 133(5), 2818–2833.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, 270(5234), 303–304.

Shetake, J. A., Wolf, J. T., Cheung, R. J., Engineer, C. T., Ram, S. K., & Kilgard, M. P. (2011). Cortical activity patterns predict robust speech discrimination ability in noise. *European Journal of Neuroscience*, 34(11), 1823–1838.

Schreiner, C. E., & Langner, G. (1988). Periodicity coding in the inferior colliculus of the cat. II. Topographical organization. *Journal of Neurophysiology*, 60(6), 1823–1840.

Schnupp, J. W., Garcia-Lazaro, J. A., & Lesica, N. A. (2015). Periodotopy in the gerbil inferior colliculus: Local clustering rather than a gradient map. *Frontiers in Neural Circuits*, 9, 37.

Schnupp, J. W. H., Hall, T. M., Kokelaar, R. F., & Ahmed, B. (2006). Plasticity of temporal pattern codes for vocalization stimuli in primary auditory cortex. *Journal of Neuroscience*, 26(18), 4785–4795.

Souffi, S., Lorenzi, C., Huetz, C., & Edeline, J. M. (2021). Robustness to noise in the auditory system: A distributed and predictable property. *eNeuro*, 8(2), ENEURO.0043-21.2021.

Souffi, S., Lorenzi, C., Varnet, L., Huetz, C., & Edeline, J. M. (2020). Noise-sensitive but more precise subcortical representations coexist with robust cortical encoding of natural vocalizations. *Journal of Neuroscience*, 40(27), 5228–5246.

Stone, M. A., Füllgrabe, C., Mackinnon, R. C., & Moore, B. C. J. (2011). The importance for speech intelligibility of random fluctuations in ‘steady’ background noise. *Journal of the Acoustical Society of America*, 130(5), 2874–2881.

Suta, D., Popelár, J., Kvasnák, E., & Syká, J. (2007). Representation of species-specific vocalizations in the medial geniculate body of the guinea pig. *Experimental Brain Research*, 183(3), 377–388.

Suta, D., Kvasnák, E., Popelár, J., & Syká, J. (2003). Representation of species-specific vocalizations in the inferior colliculus of the guinea pig. *Journal of Neurophysiology*, 90(6), 3794–3808.

Tanaka, H., & Taniguchi, I. (1991). Responses of medial geniculate neurons to species-specific vocalized sounds in the guinea pig. *Japanese Journal of Physiology*, 41(6), 817–829.

Ter-Mikaelian, M., Semple, M. N., & Sanes, D. H. (2013). Effects of spectral and temporal disruption on cortical encoding of gerbil vocalizations. *Journal of Neurophysiology*, 110(5), 1190–1204.
Varnet, L., Ortiz-Barajas, M. C., Erra, R. G., Gervain, J., & Lorenzi, C. (2017). A cross-linguistic study of speech modulation spectra. *The Journal of the Acoustical Society of America*, **142**(4), 1976–1989.

Walker, K. M., Ahmed, B., & Schnupp, J. W. (2008). Linking cortical spike pattern codes to auditory perception. *Journal of Cognitive Neuroscience*, **20**(1), 135–152.

Wallace, M. N., & Palmer, A. R. (2008). Laminar differences in the response properties of cells in the primary auditory cortex. *Experimental Brain Research*, **184**(2), 179–191.

Wallace, M. N., Shackleton, T. M., Anderson, L. A., & Palmer, A. R. (2005). Representation of the purr call in the guinea pig primary auditory cortex. *Hearing Research*, **204**(1–2), 115–126.

Wallace, M. N., Rutkowski, R. G., & Palmer, A. R. (2000). Identification and localisation of auditory areas in guinea pig cortex. *Experimental Brain Research*, **132**(4), 445–456.

Wallace, M. N., & Palmer, A. R. (2009). Functional subdivisions in low-frequency primary auditory cortex (AI). *Experimental Brain Research*, **194**(3), 395–408.

Zeng, F. G., Nie, K., Stickney, G. S., Kong, Y. Y., Vongphoe, M., Bhargave, A., Wei, C., & Cao, K. (2005). Speech recognition with amplitude and frequency modulations. *Proceedings National Academy of Science USA*, **102**(7), 2293–2298.

Zoefel, B., & VanRullen, R. (2016). EEG oscillations entrain their phase to high-level features of speech sound. *Neuroimage*, **124**(Pt A), 16–23.

**Author contributions**

This work was performed at the Paris-Saclay Institute of Neurosciences (NeuroPSI). S.S. and J.-M.E. designed the experiments. S.S. and M.Z. performed the experiments. S.S., L.V., B.B. and C.H. participated in the acquisition, analysis or interpretation of data. All authors were involved in drafting the work or revising it critically for important intellectual content. All authors approved the final version of the manuscript and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. All persons designated as authors qualify for authorship, and all those who qualify for authorship are listed.

**Funding**

J.-M.E. was supported by grants from the French Agence Nationale de la Recherche (ANR; grant number ANR-14-CE30-0019-01). S.S. was supported by the Fondation pour la Recherche Médicale (FRM; grant number ECO20160736099) and by the Entendre Foundation.

**Acknowledgements**

We thank Nihaad Paraouty for training us in the cochlear nucleus surgery. We thank Professor Christian Lorenzi and Professor Shihab Shamma for helpful comments on a previous version of the manuscript, and we are particularly grateful to Dr Virginie van Wassenhove for suggesting several improvements to the manuscript. We thank Céline Dubois, Mélanie Dumont and Aurélie Bonilla for taking care of the guinea-pig colony.

**Author’s present address**

Samira Souffi: Edmond and Lily Safra Centre for Brain Sciences, The Hebrew University of Jerusalem, Edmond J. Safra Campus, Givat Ram, Jerusalem 9190401, Israel.

**Keywords**

auditory system, degraded acoustic conditions, neuronal and behavioural discrimination, temporal envelope tracking

**Supporting information**

Additional supporting information can be found online in the Supporting Information section at the end of the HTML view of the article. Supporting information files available:

Statistical Summary Document
Peer Review History

© 2022 The Authors. The Journal of Physiology © 2022 The Physiological Society.