Title

Hierarchical Prediction Error in Neuronal Responses along the Auditory Neuraxis

Authors

Javier Nieto-Diego[1,3], Guillermo V Carbajal[1,3], Gloria G Parras[1,3], Carles Escera[4,5] & Manuel S Malmierca[1,2,3]*

Affiliations

[1] Auditory Neuroscience Laboratory, Institute of Neuroscience of Castilla y León (INCYL), Salamanca, Spain.

[2] Department of Cell Biology and Pathology, Faculty of Medicine, University of Salamanca, Salamanca, Spain.

[3] The Salamanca Institute for Biomedical Research (IBSAL), Salamanca, Spain.

[4] Brainlab-Cognitive Neuroscience Research Group, Department of Clinical Psychology and Psychobiology, University of Barcelona, Barcelona, Catalonia-Spain.

[5] Institute of Neurosciences, University of Barcelona, Barcelona, Catalonia-Spain.

Contact Information

* msm@usal.es

Abstract

Current theories of brain function depict perception as a reciprocal interchange of predictions and prediction error signals between hierarchically organized processing stations. A growing family of large-scale brain responses to perceptual mismatches supports this postulate. However, the predictive activity of the brain and its hierarchical organization remains to be demonstrated at the neuronal level. We recorded single-neuron activity during oddball stimulation, and used novel control sequences to separate prediction error from adaptation effects. Our results reveal a hierarchical organization of prediction error along the central auditory system, present already at subcortical levels and gradually increasing towards the higher-order auditory cortex. We demonstrate that the predictive activity of sensory systems is detectable at the neuronal level and highlight the role of subcortical structures in perception.

Main Text

Unexpected events are likely to convey relevant information, and their prompt detection is fundamental for survival [1,2]. Brain responses to the perceptual mismatch between expected and actual sensory inputs have been extensively recorded in all sensory systems including auditory [3], visual [4], somatosensory [5] and olfactory [6] modalities, and are thought to underlie the brain's ability to resolve auditory objects [7], proving themselves a key to understanding perceptual processing [4,8,9]. Auditory mismatch responses are typically obtained with non-invasive brain recordings using *oddball* sequences [9], in which a repetitive (standard) tone is randomly replaced by a different (deviant) tone with a low probability. Over the past 40 years, a particular mismatch response recorded from the human scalp with electroencephalography, the so-called mismatch negativity (MMN) [10], has become a valuable

tool in cognitive and clinical neuroscience [11], especially as a reliable biomarker of schizophrenia and other brain disorders [12].

At the theoretical level, large-scale mismatch responses provide empirical support to the hierarchical predictive coding framework—a neurobiologically informed and unifying account of general brain function [13–15]—, seamlessly fitting it as the sum of thousands of neuronal prediction error signals [4,16–18]. According to this theory, the classical notion that brain activity evoked by a sensory event is a neuronal representation of the occurrence of that particular event, is only half of the story. This may be true for the first/lower processing stations of sensory systems. However, at the same time, higher stations are constantly trying to anticipate the future, and send descending signals to actively suppress this evoked, ascending neuronal activity. Therefore, as the sensory signal propagates up the hierarchy of sensory systems, neuronal responses progressively switch from representing the stimulus itself to represent sensory prediction error to that stimulus. This is why neuronal responses to standard tones show repetition suppression, or response attenuation with stimulus repetition [19,20], that propagates back from higher to lower stations [21], whereas deviant tones produce a large prediction error signal, which is relayed bottom-up, facilitating the task of automatic deviance detection [22,23].

However, at the cellular level, mismatch responses could also arise from a simpler neurophysiological mechanism [24,25], namely, stimulus-specific adaptation (SSA) [26], or response decrement with stimulus repetition [2] that leaves neuronal responses to different stimuli—e.g. the deviant—almost unaffected. SSA is a widespread property of auditory neurons, increasing from midbrain [27] through the thalamus [28] to primary [29,30] and higher-order [30] auditory cortex, and assumed to be due to synaptic depression [3,29,31]. Therefore, single neuron responses along the auditory pathway show a differential response to standard and deviant tones under oddball stimulation, just as MMN but at the cellular level [3,26]. Yet,

whereas it is now clear that large-scale mismatch responses indeed reflect the predictive activity of the auditory and other sensory systems [4,17], even at early processing stages [18] including subcortical midbrain and thalamus [3], and also in animal models [32–34], this predictive activity remains to be demonstrated at the neuronal level.

In this study, we recorded individual responses of subcortical and cortical neurons along the rat auditory pathway, using recently developed control sequences to separate repetition suppression from prediction error under oddball stimulation [29,35–37]. Our data show that differential responses to deviant and standard tones in oddball sequences indeed reflect active predictive activity, instead of a mere SSA in single neurons, and that this predictive activity emerges hierarchically from subcortical structures. These results unify three coexisting views of perceptual deviance detection at different levels of description: neuronal physiology, cognitive neuroscience and the theoretical predictive coding framework.

Results

Evidence of prediction error in single auditory neurons

The predictive coding framework assumes that the same operations (generation of predictions and prediction errors) would take place at every hierarchical level of sensory systems [13], and this could in principle include subcortical processing stations [19]. Unfortunately, there is a severe dearth of evidence for this, since research on predictive brain activity has until recently focused on cortical responses of varying source and latency [17,18], and the role of subcortical structures in cognition, albeit increasingly acknowledged [38,39], remains largely unexplored. In order to collect a representative sample from different processing stations along the auditory pathway, we recorded a total of 207 neurons (Table 1) from the auditory midbrain (IC), thalamus (MGB) and cortex (AC) of anesthetized rats, while stimulating the animal with sequences of pure tones (Fig. 1). Recorded neurons were further

grouped into "first-order" (*fo*) or "higher-order" (*ho*), depending on their particular location within each nuclei [3,30], thus leading to 6 different processing stations (*fo*-IC, *ho*-IC, *fo*-MGB, *ho*-MGB, *fo*-AC, *ho*-AC; Fig. 1B; see Methods). This distinction was made because higher-order (or non-primary) auditory regions represent a higher hierarchical level of processing [40] and are known to be more sensitive to acoustic change and contextual influences than first-order (or primary) ones [3,30,41].

For each recorded neuron, we presented a set of oddball sequences, using tones selected from the neuron's frequency-response area (FRA), and a "neuronal mismatch response" (nMM) was computed as the difference between responses to deviant (DEV) and standard (STD) conditions for each tone (Fig. 1D). To determine whether this difference (usually DEV > STD) reflected predictive activity, instead of (or in addition to) just SSA, we also presented two cascaded (CAS) sequences (ascending and descending) and one many-standards (MAS) sequence as controls [36,37] (Fig. 1C), containing all tones used in oddball sequences (see Methods). The main rationale behind this design is that, in the CAS/MAS control conditions, each tone has the same (low, 10%) probability of occurrence as a DEV tone in the oddball sequence, so it is not repetitive (as the STD), and therefore is free of repetition effects (e.g. repetition suppression), but it does not stand out from the statistical context (as the DEV), and therefore it is not perceived as a deviant [36,37]. Thus, responses to CAS/MAS control conditions are used as the reference yardstick with respect to which repetition suppression and prediction error effects can be discriminated (Fig. 1D). If the neuronal mismatch response (nMM = DEV – STD) is caused entirely by SSA to the STD tone, responses to DEV and CAS/MAS control conditions should remain comparable through all hierarchical levels, or if anything, the response to DEV tones should undergo a slightly stronger suppression than to the controls, due to cross-frequency adaptation [29] (Fig. 1E). By contrast, under the predictive coding framework, deviance detection is based on Bayesian

117 inference [15], such that stronger prediction errors will be produced as more sensory evidence
118 accumulates to increase the confidence and precision of current predictions [4,19,22]. Therefore,
119 stronger prediction errors should be elicited by DEV than by CAS/MAS tones, due to the
120 lack of sequential stimulus repetitions in the controls [4,36], and this effect should increase up
121 the hierarchy (Fig. 1E), since higher-order processing stations are able to code for more
122 complex regularities [3,18,23,42].

123 Individual responses of representative neurons are shown in Fig. 2. Responses of first-
124 order neurons are mostly dependent on tone frequency, with little sensitivity to the different
125 conditions, particularly at subcortical levels (Fig. 2A,B). However, in *fo*-AC (Fig. 2C), and
126 most clearly in higher-order neurons (Fig. 2E-F), strong response suppression to STD
127 condition is apparent, but also, a higher firing rate in response to DEV tones, as compared to
128 both MAS and CAS control conditions, was consistent across tested frequencies. This is, as
129 just explained, the signature of prediction error at the single neuron level [29,32].

130 In the following, we will present only the results using the cascaded sequence as
131 control, since it was designed as an improvement to the many-standards sequence that
132 controls for additional factors beyond presentation rate of the deviant tone [36,37] (see Materials
133 and Methods, *Experimental Design*). However, the results using either CAS/MAS condition
134 as a control were commensurable (Table 1), with no remarkable differences between them
135 (Wilcoxon signed-rank test, $z = -0.125$, $p = 0.9$).

136 *The contribution of prediction error to nMM increases along the auditory hierarchy*

137 Single neuron responses to the three conditions (DEV, STD, CAS) for all tones tested
138 in all neurons are represented in Fig. 3A-F, separately for each processing station. Each pair
139 of conditions, within each station, was tested for a difference in medians (Table 1). As
140 expected, responses to DEV condition were stronger than to STD condition within all stations

(Fig. 3A-F; Table 1). This is a well described neuronal behavior across the auditory pathway [3], which has been referred to as SSA in previous studies [26], even though it was postulated to be the neuronal mechanism underlying deviance detection [29]. Indeed, this nMM results mostly from suppression of the response to the repetitive STD condition (repetition suppression), since responses to STD were significantly weaker than to CAS condition within all stations (Table 1). Critically, responses to DEV tones were significantly higher than to CAS already within the *ho*-IC (Fig. 3D; Table 1), and this difference increased progressively in the *ho*-MGB, and *ho*-AC (Fig. 3E,F), where it was most apparent. Therefore, neuronal responses showed clear signs of prediction error at the population level, within all higher-order stations, but also within *fo*-AC (Fig. 3C; Table 1), consistent with the observed effects in individual cases (Fig. 2C-F).

To quantify the relative contribution of repetition suppression and prediction error to nMM in neuronal responses, and to facilitate comparisons between different neurons/stations, we normalized the neural responses to the three conditions (DEV, STD, CAS) for each neuron/tone combination. We applied Euclidean vector normalization (Fig. 3G), such that all normalized responses (DEV_N, STD_N, CAS_N) ranged between 0 and 1. Then, we computed three indices as the difference between normalized responses to pairs of conditions, ranging between −1 and +1 (Fig. 3G). The "index of neuronal mismatch", **iMM** = $DEV_N - STD_N$, is the relative difference in responses to STD and DEV tones in the oddball paradigm. The iMM is quantitatively equivalent to the typical "SSA index" [26], used in previous studies (Fig. S1). The "index of neuronal repetition suppression", **iRS** = $CAS_N - STD_N$, is the relative reduction of the response to a standard tone, as compared to the control. Thus, the iRS quantifies repetition effects [20]. Finally, and most importantly for this study, the "index of neuronal prediction error", **iPE** = $DEV_N - CAS_N$, is the relative increase in the response to a deviant tone, compared to the control. A positive iPE reflects predictive activity, as opposed

166 to SSA [36], and quantifies the proportion of prediction error accounting for nMM [29].
167 Therefore, the relation **iMM = iRS + iPE** provides a functional, quantitative decomposition
168 of nMM (Fig. 1D). The distribution of these indices across stations reveals that both iMM
169 and iPE increase along the auditory pathway, from *fo*-IC to *ho*-AC (Fig. 3G-L).

170 Summary statistics for these normalized responses and indices are shown in Fig. 4A
171 and 4B, respectively. The iPE shows a distinct increase in two ways: (1) from first- to higher-
172 order stations, and (2) from IC to MGB to AC (Fig. 4B). To validate these observations
173 statistically, we fitted a linear model for the iPE using *nucleus* (IC, MGB, AC) and *hierarchy*
174 (*fo*, *ho*) as categorical factors. The resulting model was:

175 **iPE** = 0.012 + 0.020****ho*** – 0.136***MGB** + 0.092***AC** + 0.185****ho******MGB** +
176 0.158****ho******AC**,

177 with a significant effect of *hierarchy* ($F=37.16$, $p=1.40\cdot10^{-9}$) and *nucleus* ($F=46.35$,
178 $p=3.15\cdot10^{-20}$), and a significant *hierarchy*nucleus* interaction ($F=3.48$, $p=0.031$). Therefore,
179 both trends are significant and robust from midbrain to cortex. In particular, the significant
180 *hierarchy* effect means that the small average iPE seen in *ho*-IC (iPE = 0.012 + 0.020 =
181 **0.032**) is nevertheless statistically significant (Fig. 4B), consistent with a significant
182 difference in absolute spike counts (DEV−CAS in Table 1; Fig. 3J). Overall, this analysis
183 demonstrates a gradual emergence of a prediction error component in responses of single
184 neurons as information progresses through the auditory pathway, both in bottom-up and in
185 first- to higher-order directions, with a mutual potentiation of these two effects.

186 According to previous modeling work, change-sensitivity in single neurons is
187 expected to be maximal for stimulus ranges where the firing rate of the neuron is below
188 saturation [43]. Consistent with this hypothesis, a common observation in the pool of recorded
189 neurons was that using low stimulation intensities it was easier to produce deviance-specific

responses, particularly for ascending deviants (e.g. Fig. 2D). To test these observations at the population level, we fitted a different model for the iPE, using *SPL* (in Bels = dB SPL/10) and *direction* (ascending, ASC, or descending, DSC) of deviant tones (see Fig. 1C) as predictors. The model showed a significant effect of *SPL* (F=4.59, p=0.03) and a *SPL*direction* interaction (F=6.66, p=0.01):

$$\mathbf{iPE} = 0.064 + 0.194*\mathbf{ASC} + 0.003*\mathbf{SPL} - 0.037*\mathbf{ASC}*\mathbf{SPL},$$

which indicates that the iPE is expected to be much higher for ascending deviants at intensities below 40 dB SPL (Fig. 4C). Indeed, we observed a distinct increase in the iPE within all stations, under these stimulation conditions (Fig. 4D), particularly in *ho*-AC, where prediction error accounted for around two thirds of the iMM. This effect could facilitate perception under challenging sensory conditions, by increasing the gain of prediction error responses at early processing stages [19]. These findings run parallel to previous observations in single neurons of the primary visual cortex, where cortical feedback improves figure-background discrimination of low-salience stimuli [44].

Prediction error in single neurons correlates with a large-scale mismatch response in the auditory cortex

We also recorded local field potentials (LFP), simultaneously to single neuron spikes, from the same electrode, to explore the direct correlation between prediction error in spike responses and large-scale mismatch responses (such as the MMN). We averaged LFP responses for each condition and station, as well as the difference between DEV and CAS conditions, which we called the "prediction error potential" [33,37]: $PEP = LFP_{DEV} - LFP_{CAS}$ (Fig. 5). A significant early PEP was already detectable within *ho*-IC and *ho*-MGB (Fig. 5D,E). In the auditory cortex, the PEP was strong and significant in both *fo*-AC and *ho*-AC, showing three major deflections (Fig. 5C,F): a fast negative deflection (N1; 35−50 ms after

change onset), a slower positive deflection (P2; 70−120 ms), and a third, late, negative deflection (N2; beyond 150 ms). Importantly, epidural MMN peaks between 60 and 120 ms in rats [32], the same range of the P2 recorded here for the PEP, and can be positive when recorded from inside the brain [45]. Then, the iPE was re-computed for 12 different time windows (20 ms width, from –50 to 190 ms respect to stimulus onset), for each neuron/tone combination separately, and averaged within each station (Fig. 5). The iPE showed a clear modulation over time in both *fo*-AC and *ho*-AC stations (Friedman test, not corrected for 6 independent tests). Each individual iPE value was also tested against zero, and this analysis revealed a significant iPE within *fo*-AC between 60−100 ms after change onset, and in *ho*-AC between 40−200 ms, and seemingly beyond (Fig. 5C,F). In summary, the highest iPE values, reflecting prediction error in single neuron responses, correlate in time and location (*ho*-AC) with a large-scale mismatch wave (the PEP), putatilvely corresponding to the MMN in the rat [32,33].

Discussion

This study provides evidence, hitherto unavailable, that the hierarchical predictive activity of perceptual systems is detectable at the cellular level, even subcortically. Specifically, oddball responses of individual neurons, from midbrain to cortex, reflect predictive processing and underlie large-scale electrophysiological indicators of deviance detection. After quantitatively decomposing neuronal mismatch responses (nMM; Fig. 1D) into repetition suppression (iRS) and prediction error (iPE), the data show a systematic increase in the proportion of prediction error accounting for nMM as the sensory signal propagates up the auditory hierarchy (Fig. 4B,D). The highest iPE values are reached within the higher-order auditory cortex, where they correlate with a simultaneously recorded, large-scale prediction error potential (Fig. 5F), and extend into late evoked potentials, suggesting

238 an influence from higher-association or prefrontal cortices [46]. These results are in total
239 agreement with the predictive coding account of mismatch responses, while at the same time
240 highlight the role of subcortical structures in perception [39], providing a novel extension of the
241 mostly corticocentric predictive coding literature [14,15,38].

242 Previous attempts to show predictive activity in auditory neurons were inconclusive
243 [29,45,47], and were limited to multi-unit activity recordings in primary auditory cortex (but see
244 [48,49] for compelling evidence in single visual neurons). However, a recent study in mouse A1
245 [50] and another in rat barrel cortex [51] showed deviance detection in late responses of single
246 units, using the MAS control sequence. Although the CAS sequence is arguably a better
247 control for repetition effects than the MAS sequence [36], only one animal study has previously
248 applied it, using epidural recordings, and yielding also inconclusive results [37]. Our results,
249 using single-unit recordings, were comparable or even more robust for the CAS than for the
250 MAS control (Table 1), in agreement with human studies [36]. Our finding that the contribution
251 of prediction error to nMM supersedes that of repetition suppression within the higher-order
252 auditory cortex (Fig. 4B,D), is consistent with studies of brain sources of MMN in animals
253 [33,41] and humans [42,46] using similar controls for repetition effects. This hierarchical
254 transformation of nMM, dominated by repetition suppression at lower hierarchical levels of
255 the auditory system, with a gradual emergence of prediction error at higher levels (Fig.
256 4B,D), confirms that lower levels are mostly sensitive to global stimulus probability, while
257 higher-order levels are more sensitive to local relationships between sounds (transitional
258 probabilities), exactly as observed in human MMN studies [52,53]. Thus, our data are consistent
259 with passive SSA (Fig. 1e) underlying oddball responses in first-order midbrain and thalamus
260 [29] (Fig. 4B). By contrast, they support a generative mechanism of Bayesian inference being at
261 play in auditory cortex and higher-order subcortical stations of perceptual processing [4]. The
262 contrast between first- and higher-order nMM is particularly clear within the auditory

thalamus (compare Fig. 1E and 4B). Thus, higher-order midbrain and thalamus behave like the auditory cortex with regard to prediction error, which is the novel extension of the predictive coding scholarship. Finally, asymmetries in the direction of frequency-change detection (ascending vs. descending) have also been found in both animal [37] and human [54] MMN studies.

In conclusion, our results demonstrate that prediction error is an intrinsic component of responses of single auditory neurons, emerging even from subcortical levels, and strengthen the case for the predictive coding theory of perceptual processing. In addition, we show that neuronal predictive activity underlies the generation of large-scale mismatch responses in animal models, and parallels important properties of human MMN. These are promising results for translational research into the cellular mechanisms that are disrupted in schizophrenia and other brain disorders characterized by reductions in large-scale mismatch responses, such as MMN.

Methods

Experimental Design

The goal of the present experiments was to test responses of single neurons of the central auditory system of the rat for signs of predictive activity under oddball stimulation. We recorded extracellular single neuron activity in response to sinusoidal tones in different auditory centers of the rat brain (Fig. 1a,b). Rats were deeply anesthetized prior to surgery preparation and during the whole recording session. One single neuron was recorded at a time, using one tungsten electrode inserted into the brain, and local field potential (LFP) activity was simultaneously recorded from the same electrode. Surgical, electrophysiological and histological procedures are detailed below.

An important part of our experimental design was to record a substantial sample of neurons from the major anatomical regions representing the hierarchical organization of the central auditory system, both at cortical and subcortical levels. The inferior colliculus (IC) of the midbrain is the main convergence hub of the subcortical auditory system [3,55]. The medial geniculate body (MGB) is the auditory section of the thalamus, and relies all ascending inputs to the auditory cortex (AC), considered the highest hierarchical level of the auditory system. All these auditory processing stations contain first- and higher-order divisions [56]. First-order divisions receive their main ascending input from the brainstem (central nucleus of the IC), or from first-order division of the preceding nucleus (ventral division of the MGB and cortical fields A1, VAF and AAF), and comprise the so-called "lemniscal" auditory pathway, where the auditory information is initially processed. Higher-order divisions are integration centers for more elaborate processing of abstract properties of the stimulation, and receive their main inputs from heterogeneous sources. The cortical regions of the IC are considered higher-order, as well as the dorsal and medial divisions of the MGB. Finally, the cortical fields SRAF and PAF receive their main ascending input from the higher-order MGB, and thus represent the highest level of the auditory hierarchy in the rat [30,57].

All stimuli presented were sinusoidal pure tones of 75 ms duration, including 5 ms raise/fall ramps. For each recorded neuron, the frequency-response area (FRA) was first computed, as the map of response magnitude for each frequency/intensity combination (Fig. 2). To obtain this FRA, a randomized sequence of tones was presented at a 4 Hz rate, randomly varying frequency and intensity of the presented tones (3-5 repetitions of all tones). Then, we selected 10 evenly-spaced tones (0.5 octave separation) at a fixed sound intensity (usually 20-30 dB above minimal response threshold), so that at least two of them fell within the FRA or close to its limits (see Fig. 1c and Fig. 2). These 10 frequencies were used to create the control sequences shown in Fig. 1c. Additionally, adjacent pairs of them were used to

311 present different oddball sequences. All sequences were 400 tones in length, at the same,
312 constant presentation rate of 3 Hz (for AC) or 4 Hz (for IC and MGB). A faster presentation
313 rate was used for subcortical recordings, to compensate for the relative slowing down of pre-
314 ferred repetition rates from brainstem to cortex [58].

315 To test the specific contribution of deviance to the neuronal responses, we used oddball
316 sequences [9,26] (Fig. 1c). An oddball sequence consisted of a repetitive tone (the standard),
317 occasionally replaced by a different tone (the deviant), with a p=0.1 probability, in a pseu-
318 dorandom fashion. The first 10 tones of the sequence were always the standard tone, and a
319 minimum of 3 standard tones always preceded each deviant. Oddball sequences were either
320 ascending or descending, depending on whether the deviant was of a higher or lower fre-
321 quency than the standard, respectively (Fig. 1c). To control for the overall presentation rate of
322 the target tone, as it reduces neuronal responses at high rates, we used two different control
323 sequences, namely, the many-standards and cascaded sequences [29,36] (Fig. 1c). The many-
324 standards control sequence was a random presentation of the 10 selected tones, such that each
325 of them appeared the same number of times in an unpredictable order, with the only con-
326 straint that a single tone was never repeated in a row. Two cascaded control sequences, as-
327 cending and descending, were built as a repetitive series of groups of the 10 tones, arranged
328 by ascending/descending frequency, respectively (Fig. 1c). Since all sequences were 400
329 stimuli long, at the same presentation rate, a tone appeared with the same overall presentation
330 rate in the DEV, MAS and CAS conditions, a total of 40 times along the 400-stimuli se-
331 quence. The cascaded sequence was recently designed as an improvement to the many-
332 standards, that controls for additional key factors beyond presentation rate of the deviant tone
333 [36,37]. First, the tone immediately preceding a deviant is the same in the oddball (a standard)
334 and cascaded sequences. This improves the estimation of the overall adaptation state of the
335 system by the time the deviant tone is played, and controls for the potential sensitivity of the

neuron to a rise or fall in frequency between two successive tones. Second, the cascaded sequence mimics the regular structure of the oddball sequence, with the important difference that now the target tone *conforms* to the rule, instead of being a deviant.

Thus, using this design, every tone presented as a deviant was also presented as a standard (in a different oddball sequence) and in the context of the many-standards and cascaded control sequences. These four conditions, and by extension also response measures to them, will be denoted DEV, STD, MAS and CAS, respectively. Note that there were two variants of the DEV condition (ascending/descending), which were compared with the corresponding ascending/descending CAS condition. The STD condition was averaged, for each frequency, across ascending/descending versions of the oddball sequence (as indicated in Fig. 1c). The order of presentation of these sequences was randomized across neurons, with a silent pause of ~30 seconds between sequences. If the neuron could be held for long enough, the same protocol was repeated at different sound intensities.

Surgical procedures

Experiments were performed on 36 adult, female Long-Evans rats with body weights between 200–250 g. The experimental protocols were approved by, and used methods conforming to the standards of, the University of Salamanca Animal Care Committee and the European Union (Directive 2010/63/EU) for the use of animals in neuroscience research. Each individual animal was used to record from only one auditory station, either IC, MGB or AC. The initial surgical procedures were identical in each case, and the electrophysiological procedures differed only in the location of the craniotomy, and placement/orientation of the recording electrode, for each different station.

Surgical anesthesia was induced and maintained with urethane (1.5 g/kg, i.p.), with supplementary doses (0.5 g/kg, i.p.) given as needed. Dexamethasone (0.25 mg/kg) and atropine

sulfate (0.1 mg/kg) were administered at the beginning of the surgery and every 10 h thereafter to reduce brain edema and the viscosity of bronchial secretions, respectively. After the animal reached a surgical plane of anesthesia, the trachea was cannulated for artificial ventilation and a cisternal drain was introduced to prevent brain hernia. The animal was then placed in a stereotaxic frame in which the ear bars were replaced by hollow specula that accommodated a sound delivery system. Corneal and hind-paw withdrawal reflexes were monitored to ensure that a moderately deep anesthetic plane was maintained as uniformly as possible throughout the recording procedure. Isotonic glucosaline solution was administered periodically (5–10 ml every 6–8 hours, s.c.) throughout the experiment to prevent dehydration. Body temperature was monitored with a rectal probe and maintained between 37–38°C with a homoeothermic blanket system (Cibertec).

For IC and MGB recordings, a craniotomy was performed in the left parietal bone to expose the cerebral cortex overlying the left IC/MGB. The dura was removed, and the electrode was advanced with an angle of 20º for the IC, and in a vertical direction for the MGB. For AC recordings, the skin and temporal muscles over the left side of the skull were reflected and a 6×5 mm craniotomy was made in the left temporal bone to expose the entire auditory cortex (see Figure 1 in ref. [30]). The dura was removed and the exposed cortex and surrounding area were covered with a transparent layer of agar to prevent desiccation and to stabilize the recordings. The electrode was positioned orthogonal to the pial surface, forming a 30º angle with the horizontal plane, to penetrate through all the cortical layers of one same cortical column.

Electrophysiological recording procedures

Experiments were performed inside a sound-insulated and electrically-shielded chamber. All sounds were generated using an RX6 Multifunction Processor (TDT) and delivered mon-

aurally (to the right ear) in a closed system through a Beyer DT-770 earphone (0.1–45 kHz) fitted with a custom-made cone and coupled to a small tube (12 gauge hypodermic) sealed in the ear. The sound system response was flattened with a finite impulse response (FIR) filter, and the output of the system was calibrated in situ using a ¼-inch condenser microphone (model 4136, Brüel & Kjær), a conditioning amplifier (Nexus, Brüel & Kjær) and a dynamic signal analyzer (Photon+, Brüel & Kjær). The output of the system had a flat spectrum at 76 dB SPL (±3 dB) between 500 Hz and 45 kHz, and the second and third harmonic components in the signal were ≤ 40 dB below the level of the fundamental at the highest output level (90 dB SPL). Prior to surgery and recording sessions, we recorded auditory brainstem responses (ABR) with subcutaneous electrodes to ensure the animal had normal hearing. ABRs were collected using TDT software (BioSig) and hardware (RX6 Multifunction Processor) following standard procedures (0.1 ms clicks presented at a 21/s rate, delivered in 10 dB ascending steps from 10 to 90 dB SPL).

Action potentials and local field potentials (LFP) were recorded with hand-manufactured, glass-coated tungsten electrodes [59] (1–4 MΩ impedance at 1 kHz). One individual electrode was used to record one single neuron at a time. The electrode was advanced using a piezoelectric micromanipulator (Sensapex) until we observed a strong spiking activity synchronized with the train of searching stimuli. The signal was amplified (1000×) and band-pass filtered (1 Hz to 3 kHz) with an alternate current differential amplifier (DAM-80, WPI). This analog signal was digitized at a 12K sampling rate and further band-pass filtered (with a second TDT-RX6 module) separately for action potentials (between 500 Hz and 3 kHz) and LFP (between 3 and 50 Hz). Stimulus generation and neuronal response processing and visualization were controlled online with custom software created with the OpenEx suite (TDT) and Matlab (Mathworks). A unilateral threshold for automatic action potential detection was manually set at about 2–3 standard deviations of the background noise. Spike waveforms

were displayed on the screen, and overlapped on each other in a pile-plot to facilitate isolation of single units. Only when all snippet waveforms were identical and clearly separable form other smaller units and the background noise, the recorded action potentials were considered to belong to a single unit.

Sounds used for stimulation were white noise bursts or pure tones with 5 ms rise-fall ramps. Sounds used for searching for neuronal activity were trains of noise bursts or pure tones (1–8 stimulus per second). We used short stimulus duration for searching (30 ms) to prevent strong adaptation. In addition, type (white noise, narrowband noise, pure tone) and parameters (frequency, intensity, presentation rate) of the search stimuli were varied manually when necessary to facilitate release from adaptation, and thus prevent overlooking responses with high SSA. Once a single neuron was isolated and confirmed to be stable, the whole stimulation protocol was applied, as described in the first section "Experimental Design".

Histological procedures and anatomical localization of recording sites

AC experiments. At the end of the surgery, a magnified picture (25×) of the exposed cortex was taken [30] with a digital SLR camera (D5100, Nikon) coupled to the surgical microscope (Zeiss) through a lens adapter (TTI Medical). The picture included a pair of reference points previously marked on the dorsal ridge of the temporal bone, indicating the absolute scale and position of the image with respect to bregma. This picture was displayed on a computer screen and a micrometric grid was overlapped to guide and mark the placement of the electrode for every recording made. Recording sites (250–500 μm spacing) were evenly distributed across the cortical region of interest while avoiding blood vessels. The vascular pattern was used as a local reference to mark the position of every recording site in the picture, but otherwise differed largely between animals. To confirm the actual depth and cortical lay-

er of the recorded neurons, at the end of the experiment we made electrolytic lesions at one to four of the recording sites, at the same depth that recordings were made. Five auditory cortical fields were identified according to tone frequency response topographies [30]. The limits and relative position of the auditory fields were determined for each animal at the end of the experiment, using the characteristic frequency (CF; the tone frequency that elicits a significant neuronal response at the lowest intensity) gradient as the main reference landmark [30,57]. We consistently observed distinct tonotopic gradients within the different fields, with a high-frequency reversal between VAF and AAF (rostrally), a low-frequency reversal between A1 and PAF (dorsocaudally) and a high-frequency reversal between VAF and SRAF (ventrally). We identified the boundary between A1 and VAF as a 90° shift in the CF gradient in the ventral low-frequency border of A1, and the boundary between A1 and AAF as an absence of tone-evoked responses in the ventral, high-frequency border of A1 [30]. We used these boundaries to assign each recording to a given field. The CF of each recording track was computed as the average CF of all neurons recorded in that track, including a fast multi-unit activity FRA recording made between 400-550 μm depth, corresponding to layers IIIb-IV of the auditory cortex.

IC and MGB experiments. Each recording track was marked with electrolytic lesions for subsequent histological localization of the neurons recorded. At the end of the experiment, the animal was given a lethal dose of sodium pentobarbital and perfused transcardially with phosphate buffered saline (0.5% $NaNO_3$ in PBS) followed by fixative (a mixture of 1% paraformaldehyde and 1% glutaraldehyde in rat Ringer's solution). After fixation and dissection, the brain tissue was cryoprotected in 30% sucrose and sectioned on a freezing microtome in the transverse or sagittal planes into 40 mm-thick sections. Sections were Nissl stained with 0.1% cresyl violet to facilitate identification of cytoarchitectural boundaries. Recording sites were marked on standard sections from a rat brain atlas (Paxinos and Watson, 6th Edition)

458 and neurons were assigned to one of the main divisions of the IC (central nucleus, dorsal,
459 lateral or rostral cortex) or the MGB (ventral, dorsal and medial division), respectively. The
460 stained sections with the lesions were used to localize each track mediolaterally, dorsoven-
461 trally and rostrocaudally in the Paxinos atlas. To determine the main IC or MGB subdivi-
462 sions, cytoarchitectonic criteria, i.e., cell shape and size, Nissl staining patterns and cell pack-
463 ing density, were used. This information was complemented and confirmed by the stereotaxic
464 coordinates used during the experiment to localize the IC/MGB. After assigning a section to
465 each track/lesion, the electrophysiological coordinates from each experiment and recording
466 unit, i.e., beginning and end of the IC/MGB, as well as the depth of the neuron, were used as
467 complementary references to localize each neuron within a track.

468 *Statistical Analysis*

469 All data analyses were performed with the Matlab^{TM} software, using the built-in func-
470 tions, the Statistics and Machine Learning toolbox, or custom scripts and functions developed
471 in our laboratory. Peri-stimulus time histograms (PSTH) were generated for each stimu-
472 lus/condition tested. Only the last STD tones preceding each DEV tone were used for the
473 analyses. A PSTH was a histogram of action potential density over time (in action potentials
474 per second, or Hz) from −75 to 250 ms around stimulus onset, using the 40 trials available for
475 each tone and condition. Every PSTH was smoothed with a 6 ms gaussian kernel ("ksdensi-
476 ty" function in Matlab) in 1 ms steps to estimate the spike-density function (SDF) over time,
477 and the baseline spontaneous firing rate (SFR) was determined as the average firing rate (in
478 Hz) during the 75 ms preceding stimulus onset. For any given time window, the excitatory
479 response was measured as the area below the SDF and above the baseline SFR (positive area
480 patches only, to avoid negative response values). This measure will be referred to as "base-
481 line-corrected spike count".

We used two types of sequences to control for repetition effects (*v.s. Experimental Design*), namely the many-standards and cascaded sequences (Fig. 1d). However, only one of them is required to decompose neuronal mismatch into repetition suppression and prediction error (Fig. 1d). In the following, we describe the analysis performed using the CAS condition as control, since the analysis using the MAS sequence is completely analogous. Baseline-corrected spike count responses of a neuron to the same tone in the three conditions (DEV, STD, CAS) were normalized using the formulas:

$DEV_N = DEV/N;$

$STD_N = STD/N;$

$CAS_N = CAS/N;$

Where

$$N = \sqrt{DEV^2 + STD^2 + CAS^2}$$

is the Euclidean norm of the vector (DEV, STD, CAS) defined by the three responses. This normalization procedure always results in a value ranging 0 to 1, and has a straightforward geometrical interpretation (Fig. 3b,h): Normalized values are the coordinates of a 3D unit vector (DEV_N, STD_N, CAS_N) with the same direction of the original vector (DEV, STD, CAS), and thus the same proportions between the three response measures. From these normalized responses, indices of neuronal mismatch (iMM), repetition suppression (iRS), and prediction error (iPE) were computed as:

iMM = $DEV_N - STD_N$,

iRS = $CAS_N - STD_N$,

iPE = $DEV_N - CAS_N$,

These indices, consequently, always range between -1 and 1, and provide the following quantitative decomposition of neuronal mismatch (Fig. 1d) into repetition suppression and prediction error:

iMM = iRS + iPE

As shown if Fig. S1, the iMM is largely equivalent to the typical SI, or "SSA index", commonly used in most previous studies of SSA in single units [26,29]:

SI = (DEV-STD)/(DEV+STD)

For the analysis of the LFP signal, we aligned the recorded wave to the onset of the stimulus for every trial, and computed the mean LFP for every recording site and stimulus condition (DEV, STD, CAS), as well as the "prediction error potential" (PEP = $LFP_{DEV} - LFP_{CAS}$). Then, grand-averages were computed for all conditions, for each auditory station separately. The p-value of the grand-averaged PEP was determined for every time point with a two-tailed t-test (Bonferroni-corrected for 200 comparisons, with family-wise error rate FWER<0.05), and we computed the time intervals where PEP was significantly different from zero (Fig. 5).

All statistical tests used were distribution-free tests (or "nonparametric", namely the Wilcoxon signed-rank test and Friedman test), given the non-normal nature of our dataset (baseline-corrected spike counts, normalized responses, indices of neuronal mismatch, repetition suppression and prediction error). Only the difference wave for the LFPs (PEP in Fig. 5) was tested using a t-test, since each LFP trace is itself an average of 40 waves, and thus approximately normal (according to the Central Limit Theorem). Linear models used to test significant average iPE within each auditory station (Fig. 4b,d) and significant effects of *nucleus*, *hierarchy*, *SPL*, *direction*, and interactions between them, were fitted using the 'fitlm' function in Matlab, with robust options.

References

1. Ranganath, C. & Rainer, G. Neural mechanisms for detecting and remembering novel events. *Nat Rev Neurosci* **4,** 193–202 (2003).

2. Whitmire, C. J. & Stanley, G. B. Rapid Sensory Adaptation Redux: A Circuit Perspective. *Neuron* **92,** 298–315 (2016).

3. Escera, C. & Malmierca, M. S. The auditory novelty system: an attempt to integrate human and animal research. *Psychophysiology* **51,** 111–123 (2014).

4. Stefanics, G., Kremláček, J. & Czigler, I. Visual mismatch negativity: a predictive coding view. *Front Hum Neurosci* **8,** 666 (2014).

5. Ostwald, D. *et al.* Evidence for neural encoding of Bayesian surprise in human somatosensation. *Neuroimage* **62,** 177–188 (2012).

6. Pause, B. M. & Krauel, K. Chemosensory event-related potentials (CSERP) as a key to the psychology of odors. *Int J Psychophysiol* **36,** 105–122 (2000).

7. Winkler, I., Denham, S. L. & Nelken, I. Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends Cogn Sci* **13,** 532–540 (2009).

8. Näätänen, R., Astikainen, P., Ruusuvirta, T. & Huotilainen, M. Automatic auditory intelligence: an expression of the sensory-cognitive core of cognitive processes. *Brain Res Rev* **64,** 123–136 (2010).

9. Dehaene, S., Meyniel, F., Wacongne, C., Wang, L. & Pallier, C. The Neural Representation of Sequences: From Transition Probabilities to Algebraic Patterns and Linguistic Trees. *Neuron* **88,** 2–19 (2015).

10. Näätänen, R., Paavilainen, P., Rinne, T. & Alho, K. The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin Neurophysiol* **118,** 2544–2590 (2007).

11. Näätänen, R. *et al.* The mismatch negativity (MMN)--a unique window to disturbed central auditory processing in ageing and different clinical conditions. *Clin Neurophysiol* **123,** 424–458 (2012).

12. Michie, P. T., Malmierca, M. S., Harms, L. & Todd, J. The neurobiology of MMN and implications for schizophrenia. *Biol Psychol* **116,** 90–97 (2016).

13. Friston, K. A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci* **360,** 815–836 (2005).

14. Bastos, A. M. *et al.* Canonical microcircuits for predictive coding. *Neuron* **76,** 695–711 (2012).

15. Friston, K. The free-energy principle: a rough guide to the brain? *Trends Cogn Sci* **13,** 293–301 (2009).

16. Schröger, E. *et al.* Predictive regularity representations in violation detection and auditory stream segregation: from conceptual to computational models. *Brain Topogr* **27,** 565–577 (2014).

17. Phillips, H. N. *et al.* Convergent evidence for hierarchical prediction networks from human electrocorticography and magnetoencephalography. *Cortex* **82,** 192–205 (2016).

18. Bendixen, A., SanMiguel, I. & Schröger, E. Early electrophysiological indicators for predictive processing in audition: a review. *Int J Psychophysiol* **83,** 120–131 (2012).

19. Auksztulewicz, R. & Friston, K. Repetition suppression and its contextual determinants in predictive coding. *Cortex* **80,** 125–140 (2016).

20. Baldeweg, T. Repetition effects to sounds: evidence for predictive coding in the auditory system. *Trends Cogn Sci* **10,** 93–94 (2006).

21. Garrido, M. I. *et al.* Repetition suppression and plasticity in the human brain. *Neuroimage* **48,** 269–279 (2009).

22. Garrido, M. I., Kilner, J. M., Stephan, K. E. & Friston, K. J. The mismatch negativity: a review of underlying mechanisms. *Clin Neurophysiol* **120,** 453–463 (2009).

23. Wacongne, C., Changeux, J.-P. & Dehaene, S. A neuronal model of predictive coding accounting for the mismatch negativity. *J Neurosci* **32,** 3665–3678 (2012).

24. May, P. J. C. & Tiitinen, H. Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. *Psychophysiology* **47,** 66–122 (2010).

25. Fishman, Y. I. The mechanisms and meaning of the mismatch negativity. *Brain Topogr* **27,** 500–526 (2014).

26. Ulanovsky, N., Las, L. & Nelken, I. Processing of low-probability sounds by cortical neurons. *Nat Neurosci* **6,** 391–398 (2003).

27. Malmierca, M. S., Cristaudo, S., Pérez-González, D. & Covey, E. Stimulus-specific adaptation in the inferior colliculus of the anesthetized rat. *J Neurosci* **29,** 5483–5493 (2009).

28. Antunes, F. M. & Malmierca, M. S. An overview of stimulus-specific adaptation in the auditory thalamus. *Brain Topogr* **27,** 480–499 (2014).

29. Taaseh, N., Yaron, A. & Nelken, I. Stimulus-specific adaptation and deviance detection in the rat auditory cortex. *PLoS One* **6,** e23369 (2011).

30. Nieto-Diego, J. & Malmierca, M. S. Topographic Distribution of Stimulus-Specific Adaptation across Auditory Cortical Fields in the Anesthetized Rat. *PLoS Biol.* **14,** e1002397 (2016).

31. Grill-Spector, K., Henson, R. & Martin, A. Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn. Sci. (Regul. Ed.)* **10,** 14–23 (2006).

32. Harms, L., Michie, P. T. & Näätänen, R. Criteria for determining whether mismatch responses exist in animal models: Focus on rodents. *Biol Psychol* **116,** 28–35 (2016).

33. Shiramatsu, T. I., Kanzaki, R. & Takahashi, H. Cortical mapping of mismatch negativity with deviance detection property in rat. *PLoS One* **8,** e82663 (2013).

34. Todd, J., Harms, L., Schall, U. & Michie, P. T. Mismatch negativity: translating the potential. *Front Psychiatry* **4,** 171 (2013).

35. Jacobsen, T. & Schröger, E. Is there pre-attentive memory-based comparison of pitch? *Psychophysiology* **38,** 723–727 (2001).

36. Ruhnau, P., Herrmann, B. & Schröger, E. Finding the right control: the mismatch negativity under investigation. *Clin Neurophysiol* **123,** 507–512 (2012).

37. Harms, L. *et al.* Mismatch negativity (MMN) in freely-moving rats with several experimental controls. *PLoS One* **9,** e110892 (2014).

38. Parvizi, J. Corticocentric myopia: old bias in new cognitive sciences. *Trends Cogn Sci* **13,** 354–359 (2009).

39. Güntürkün, O. & Bugnyar, T. Cognition without Cortex. *Trends Cogn Sci* **20,** 291–303 (2016).

40. Atiani, S. *et al.* Emergent selectivity for task-relevant stimuli in higher-order auditory cortex. *Neuron* **82,** 486–499 (2014).

41. Kraus, N., McGee, T., Littman, T., Nicol, T. & King, C. Nonprimary auditory thalamic representation of acoustic change. *J Neurophysiol* **72,** 1270–1277 (1994).

42. Opitz, B., Schröger, E. & von Cramon, D. Y. Sensory and cognitive mechanisms for preattentive change detection in auditory cortex. *Eur J Neurosci* **21,** 531–535 (2005).

43. Abbott, L. F., Varela, J. A., Sen, K. & Nelson, S. B. Synaptic depression and cortical gain control. *Science* **275,** 220–224 (1997).

44. Hupé, J. M. *et al.* Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature* **394,** 784–787 (1998).

45. Fishman, Y. I. & Steinschneider, M. Searching for the mismatch negativity in primary auditory cortex of the awake monkey: deviance detection or stimulus specific adaptation? *J Neurosci* **32,** 15747–15758 (2012).

46. Maess, B., Jacobsen, T., Schröger, E. & Friederici, A. D. Localizing pre-attentive auditory memory-based comparison: magnetic mismatch negativity to pitch change. *Neuroimage* **37,** 561–571 (2007).

47. Farley, B. J., Quirk, M. C., Doherty, J. J. & Christian, E. P. Stimulus-specific adaptation in auditory cortex is an NMDA-independent process distinct from the sensory novelty encoded by the mismatch negativity. *J Neurosci* **30,** 16475–16484 (2010).

48. Meyer, T. & Olson, C. R. Statistical learning of visual transitions in monkey inferotemporal cortex. *Proc. Natl. Acad. Sci. U.S.A.* **108,** 19401–19406 (2011).

49. Zmarz, P. & Keller, G. B. Mismatch Receptive Fields in Mouse Visual Cortex. *Neuron* **92,** 766–772 (2016).

50. Chen, I.-W., Helmchen, F. & Lütcke, H. Specific Early and Late Oddball-Evoked Responses in Excitatory and Inhibitory Neurons of Mouse Auditory Cortex. *J Neurosci* **35,** 12560–12573 (2015).

51. Musall, S., Haiss, F., Weber, B. & von der Behrens, W. Deviant Processing in the Primary Somatosensory Cortex. *Cereb Cortex* (2015). doi:10.1093/cercor/bhv283

52. Koelsch, S., Busch, T., Jentschke, S. & Rohrmeier, M. Under the hood of statistical learning: A statistical MMN reflects the magnitude of transitional probabilities in auditory sequences. *Sci Rep* **6,** 19741 (2016).

53. Winkler, I. & Schröger, E. Auditory perceptual objects as generative models: Setting the stage for communication by sound. *Brain Lang* **148,** 1–22 (2015).

54. Peter, V., McArthur, G. & Thompson, W. F. Effect of deviance direction and calculation method on duration and frequency mismatch negativity (MMN). *Neurosci Lett* **482,** 71–75 (2010).

55. Malmierca, M. S. The structure and physiology of the rat auditory system: an overview. *Int. Rev. Neurobiol.* **56,** 147–211 (2003).

56. Lee, C. C. & Sherman, S. M. On the classification of pathways in the auditory midbrain, thalamus, and cortex. *Hear. Res.* **276,** 79–87 (2011).

57. Polley, D. B., Read, H. L., Storace, D. A. & Merzenich, M. M. Multiparametric auditory receptive field organization across five cortical fields in the albino rat. *J Neurophysiol* **97,** 3621–3638 (2007).

58. Eggermont, J. J. Animal models of auditory temporal processing. *Int J Psychophysiol* **95,** 202–215 (2015).

59. Bullock, D. C., Palmer, A. R. & Rees, A. Compact and easy-to-use tungsten-in-glass microelectrode manufacturing workstation. *Med Biol Eng Comput* **26,** 669–672 (1988).

Acknowledgements

We thank Drs. Bernhard Englitz, Alan Palmer, Patrick May, Daniel Polley, José Luis Peña, Juanita Todd, Ryszard Auksztulewicz, Christoph Schreiner, Iria SanMiguel and Adrian Rees for their comments on a previous version of the manuscript and for their constructive criticisms.

Financial support was kindly provided by the Spanish MINECO (BFU2013-43608-P & SAF2016-75803-P) and JCYL (SA343U14) to MSM and a Explora-Ciencia grant (PSI2013-49348-EXPLORA) to MSM and CE. CE was also supported by the Generalitat de Catalunya (SGR2014-177) and by the Icrea Acadèmia Distinguished Professor Award. JND held a fellowship from the European Social Fund/Spanish JCYL (Operational Programme ESF Castilla y León 2007–2013). GGP held a fellowship from the Spanish MINECO (BES-2014-069113).

Author contributions

The experiments were performed at the Neurobiology of Hearing Laboratory, Institute of Neuroscience of Castilla y León- INCYL, University of Salamanca, Salamanca, Spain. The contribution of each author to the following aspects of the study is as stated: (1) collection of data: JND, GVC and GGP; (2) conception and design of experiments: JND and MSM; (3) analysis, interpretation of data and conceptual advice: JND, GVC; GGP; CE and MSM; (5) writing of the manuscript: JND and MSM. All authors approved the final version of the manuscript.

Figures

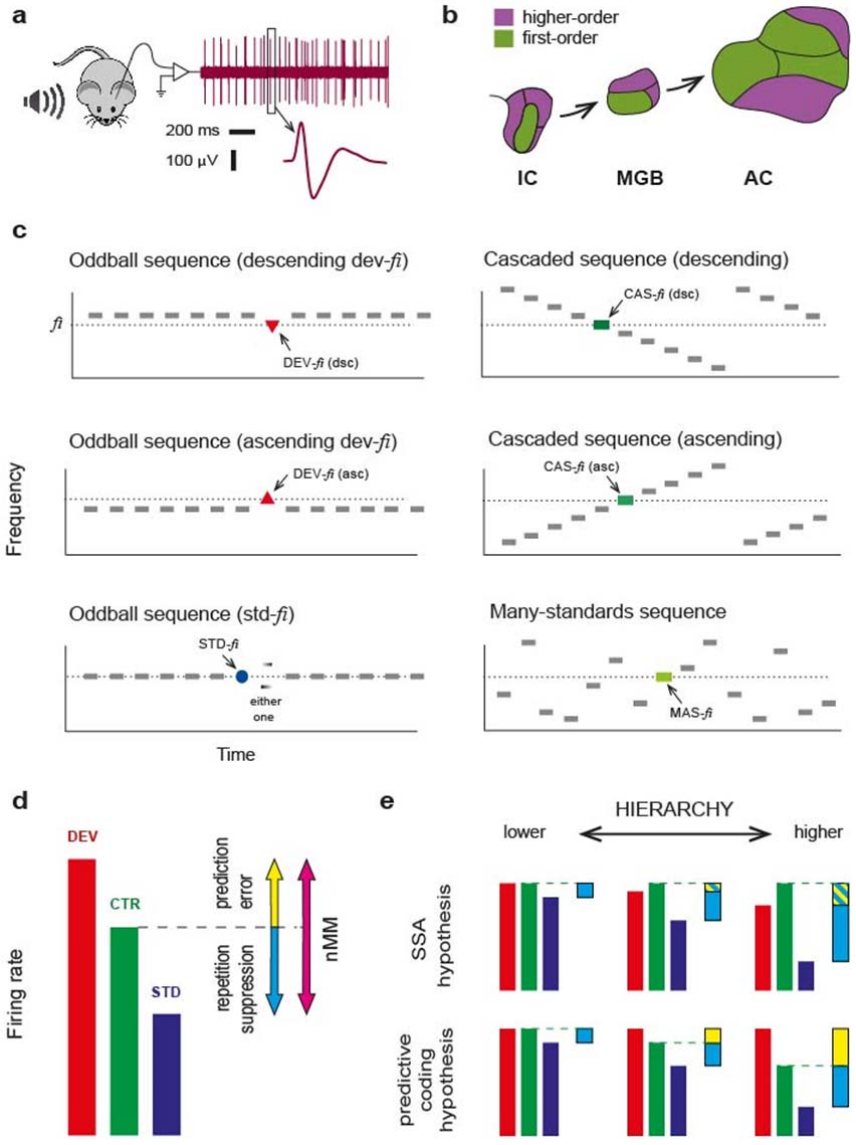


Figure 1: Experimental design. **a**. Sketch of experimental setup. Isolated neurons were recorded from different auditory nuclei of anesthetized rats, while stimulating with pure tones. **b.** Schematic representation of the major nuclei in the rat auditory pathway form midbrain to cortex [3,30], divided into first- and higher-order regions. **c.** Stimulation sequences. For each recorded neuron, 10 tones of evenly-spaced frequencies were selected to construct these stimulation sequences. Using this design, each tone f_i (i=1...10) lying inside the

690 neuron's receptive field could be presented in two experimental conditions (DEV and STD,
691 in separate oddball sequences), and two control conditions (CAS/MAS) for adaptation
692 effects. Note that ascending/descending DEV tones will be compared to the corresponding
693 version of the CAS condition (see Methods). **d.** Decomposition of neuronal mismatch
694 responses (nMM=DEV−STD) to the oddball sequence using either one of the control
695 conditions. **e.** Predicted scenarios under two competing mechanisms explaining nMM. If SSA
696 is the main mechanism underlying nMM, responses to STD tones will be more suppressed
697 the more synapses information traverses along the auditory hierarchy, and responses to
698 control (CAS/MAS) tones would be equal to, or stronger than, to DEV tones, since the
699 average intertonal distance is larger in the controls than in oddball sequences [29]. By contrast,
700 if nMM reflects Bayesian inference, responses to DEV tones would be progressively larger
701 than to the controls as the information propagates up the auditory hierarchy [4].

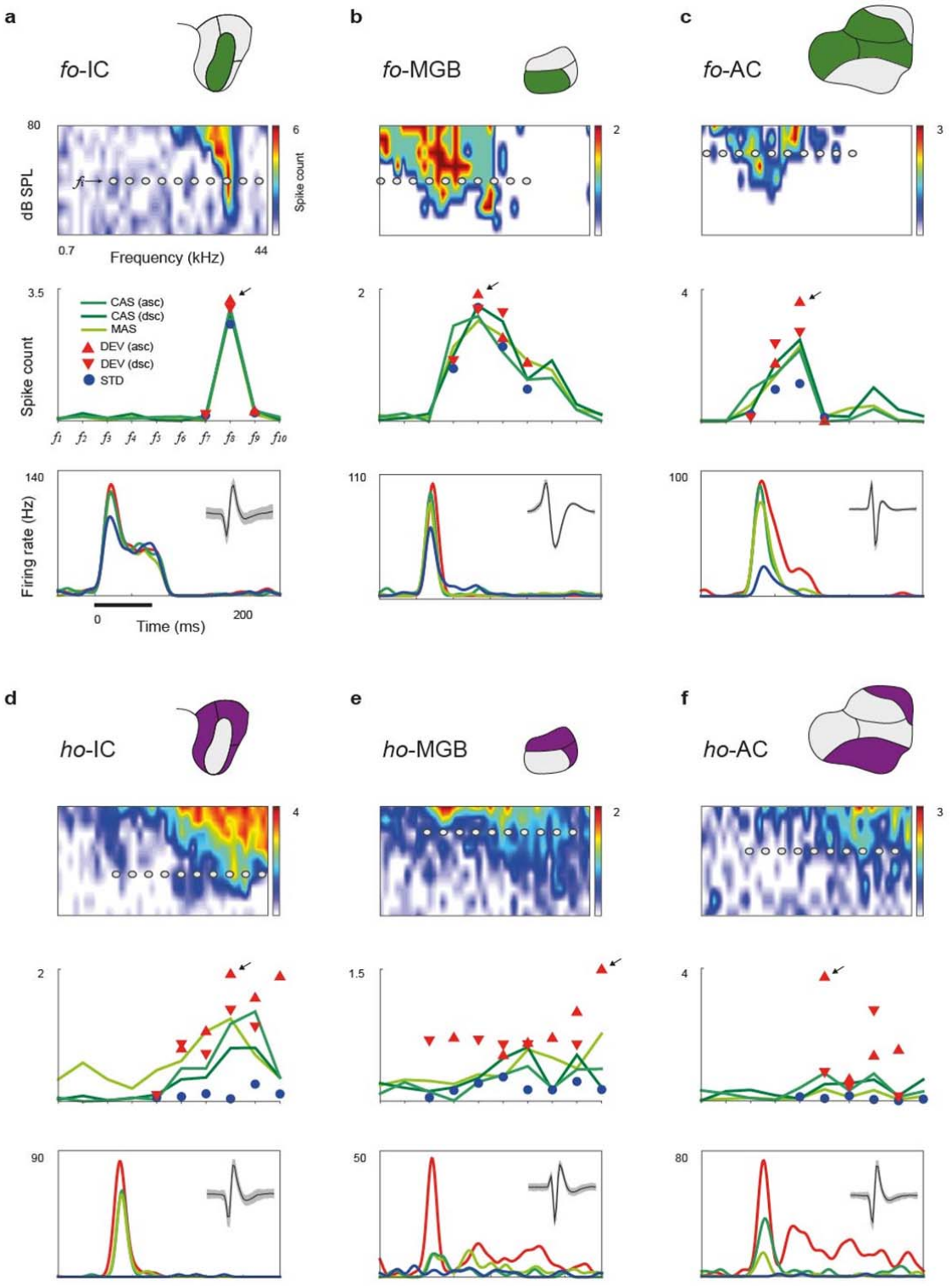
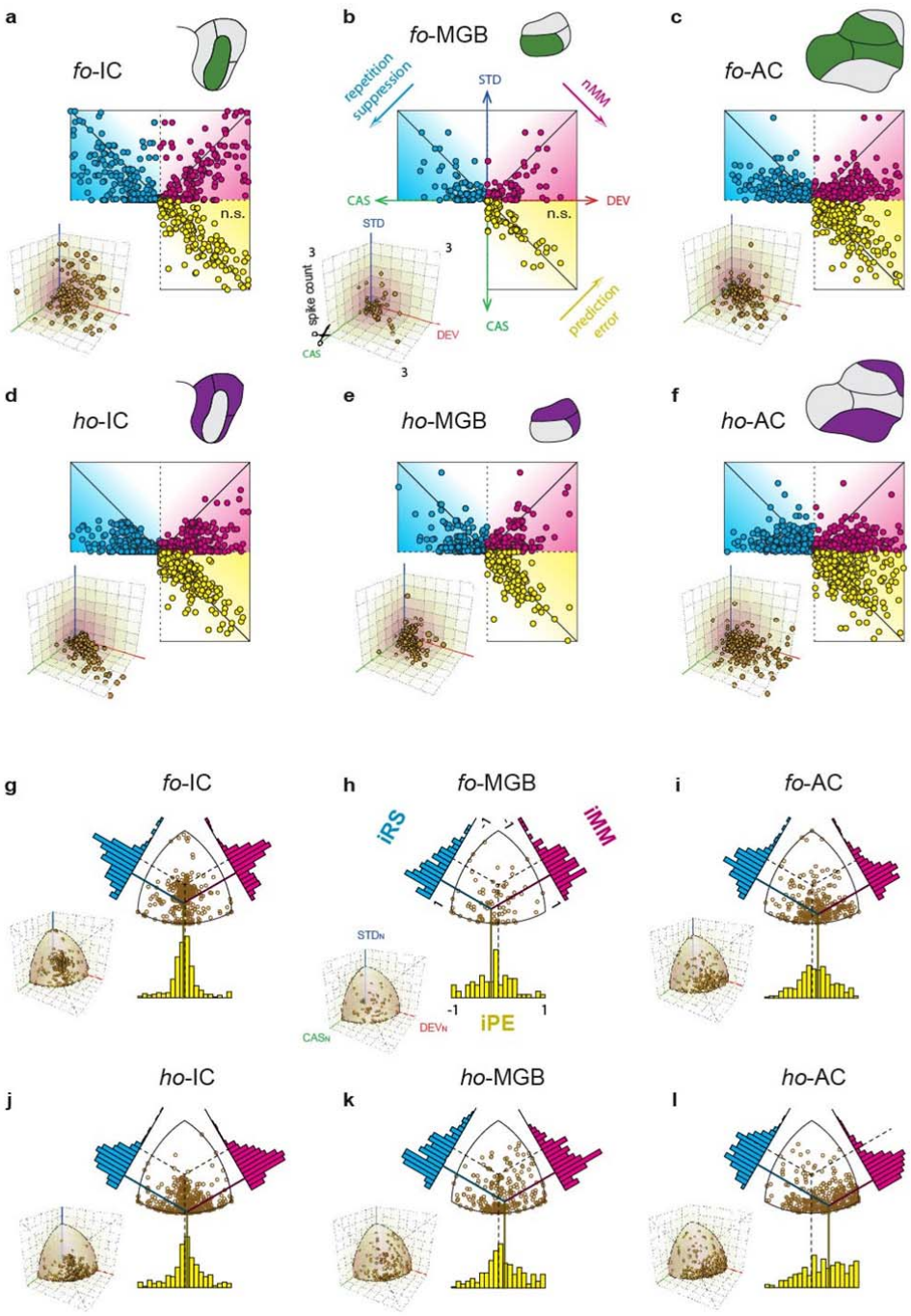


Figure 2: Prediction error in sample neuronal responses. a-f. Each panel shows responses of representative neurons within each station of the auditory pathway: (1) The FRA

(representation of neuronal sensitivity to different frequency/intensity combinations) and the 10 tones selected to create the control sequences for that particular neuron (see Methods). (2) Measured responses of the neuron to each tone (baseline-corrected spike counts, averaged within 0−180 ms after tone onset), for all conditions tested. (3) Sample PSTH for each condition, for the tone with the highest response (either ascending or descending; indicated with an arrow). Stimulus duration is represented by the thick, horizontal line, and the isolated spike (mean ± SEM) is shown in the small inset. Note that both repetition suppression (STD < CTR) and prediction error (DEV > CTR) can be observed in responses to some tones, and this is particularly consistent for higher-order neurons (panels D-F).



714

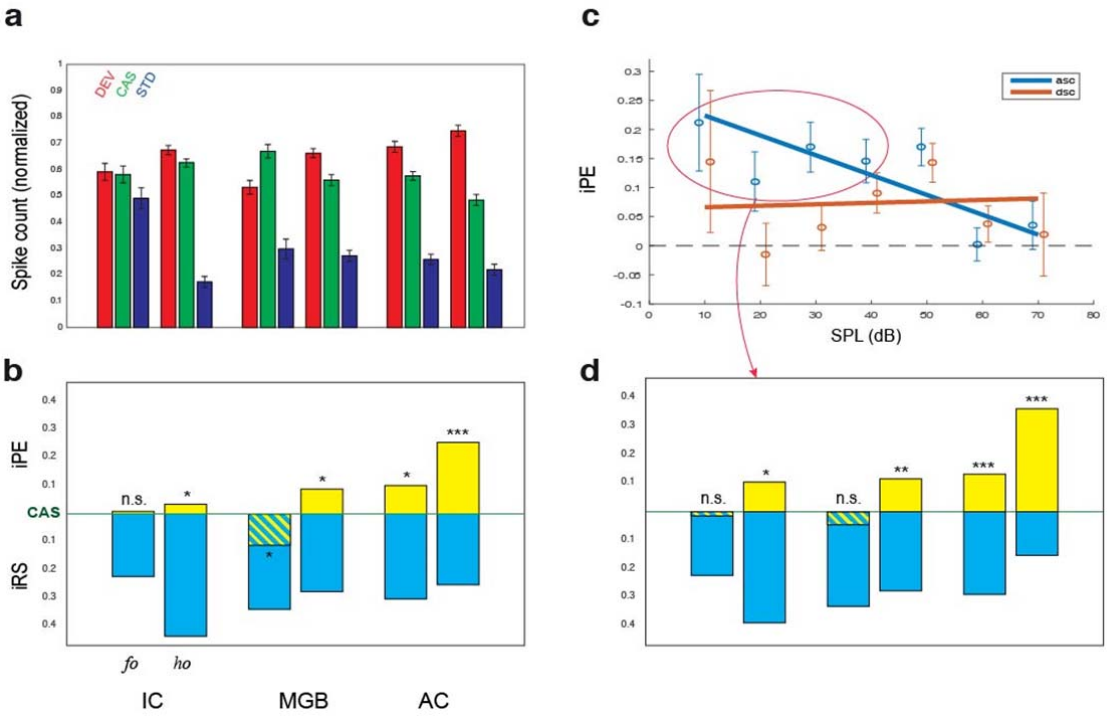
715 **Figure 3. Prediction error at the population level**. **a-f**. Responses to the three conditions

716 (DEV, STD, CAS; for all tones tested in all neurons) were represented on a 3D scatter,

717 separately for each station. These points were then orthogonally projected onto the three

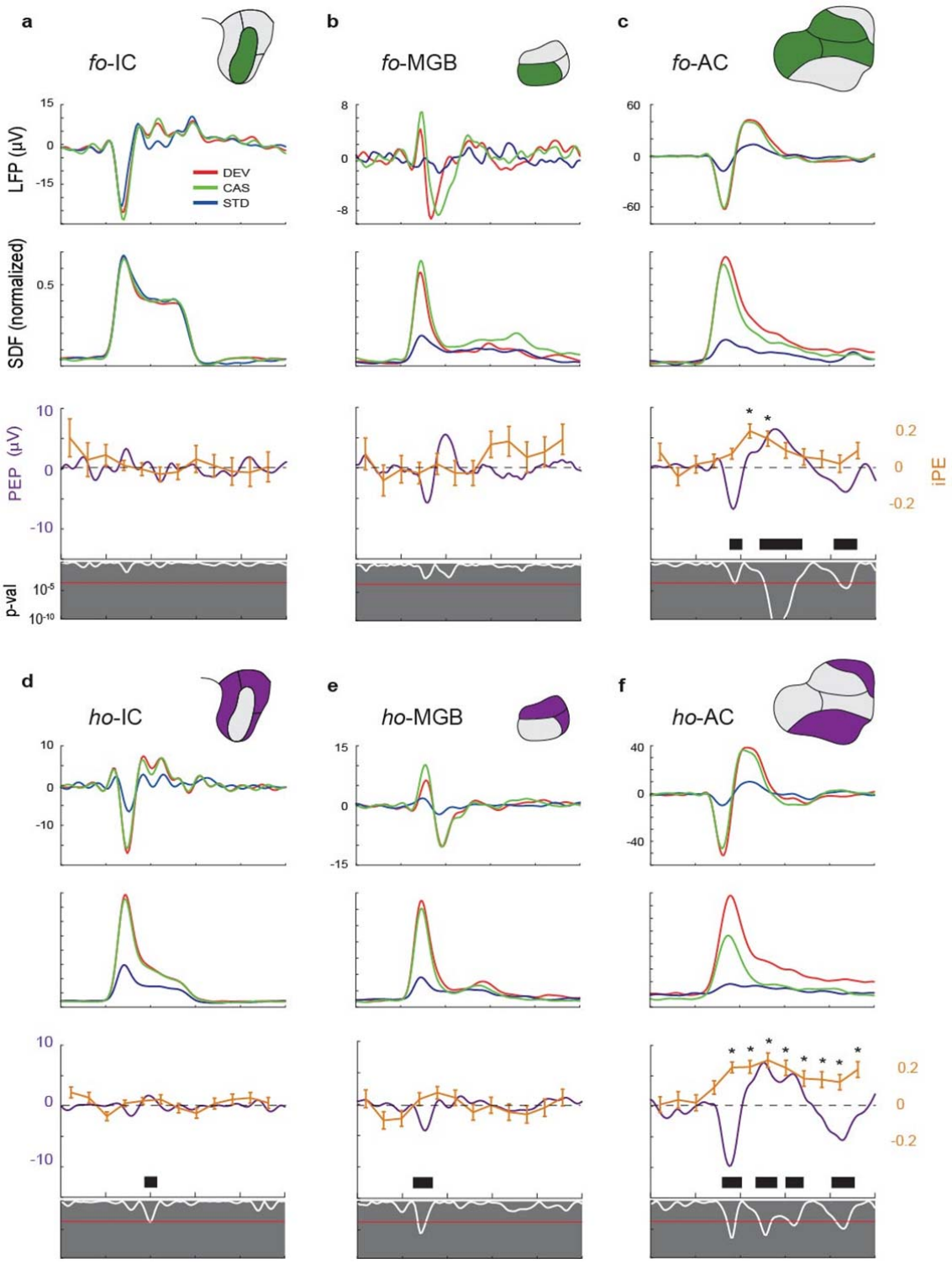
718 "walls", to compare two responses at a time, and then the "box" was unfolded (after "cutting"

along the CAS axes) to create the main, flat diagrams. Thus, each 2D point represents the response (baseline-corrected spike count) of a single neuron to one given tone for a pair of conditions. The clouds of magenta and blue points concentrate below the diagonal in all stations, indicating neuronal mismatch and repetition suppression, respectively, at the population level. The cloud of yellow points remains unbiased in lower stations (a,b), but is displaced above the diagonal in higher stations, especially in AC (c,f). This indicates an important contribution of prediction error to neuronal responses in these stations. **g-l**. Distribution of normalized responses and indices of neuronal mismatch (iMM), repetition suppression (iRS) and prediction error (iPE). Each point in the 3D scatters from panels a-f represents a vector in response space (DEV, STD, CAS). The normalization is just the radial projection of this point onto the unit sphere centered on the origin (small insets), so the resulting vector (DEV_N, STD_N, CAS_N,) is a scaled version of the former. The flat diagram is a zenith view of the 3D sphere. Each diagonal (dotted black lines) represents the line where the corresponding index is zero, and the index will increase or decrease as a projected point moves away from this line. Histograms represent index distributions, with their means indicated by colored lines. Note the overall shift of the mean iPE towards positive values, from IC through MGB to AC, and from first- to higher-order divisions.



736

737 **Figure 4: Emergence of iPE along the auditory hierarchy. a**. Average normalized
738 responses (mean ± SEM) to the three conditions (DEV_N, STD_N, CAS_N) within each station.
739 **b**. These same normalized responses are represented with respect to the CAS control
740 condition, so that the indices are represented by their differences (iPE is upwards-positive,
741 iRS is downwards-positive). Asterisks denote statistical significance of iPE against zero
742 median (Table 1) **c**. Linear model fitted for the iPE, using SPL and Direction
743 (ascending/descending) as predictors. Error bars denote mean and SEM for each SPL and
744 Direction. **d**. The same as in (b), but using only recordings for ascending deviant tones at
745 intensities ≤ 40 dB SPL.



746

747 **Figure 5: Correlation of iPE and the local-field prediction error signal (PEP).** Population

748 grand-averages for different response measures, computed for each processing station

separately: (1) Average local field potentials (LFP) across tested tones and recording sites for the different conditions. (2) Average firing rate profiles, as spike-density functions (SDF, normalized to better match the iPE traces shown below). (3) Average "local-field prediction error signal" (PEP = $LFP_{DEV} - LFP_{CAS}$; white trace: instantaneous p-value for the PEP, paired t-test against equal means; red horizontal line: critical threshold with Bonferroni correction for 200 comparisons, FWER<0.05; thick black bars: time intervals for which average PEP is significant). (4) Along with the PEP trace, the time course of the average iPE is plotted in orange (mean ± SEM, asterisks indicate a significant iPE for the corresponding time window; Wilcoxon signed rank test with Bonferroni correction for 12 comparisons, FWER<0.05). Highest iPE values are concurrent in time and location (auditory cortex; panels C-F) with the strongest PEP.

Tables

Table 1: Summary of principal dataset. For each auditory station: Number of recorded neurons and tested neuron/tone combinations (points). Median values for baseline-corrected spike counts (spk) to the different conditions. Median differences between the former measures, and associated p-values against zero (Friedman test with post-hoc multiple comparison, Fisher's Least Significant Difference method, uncorrected for 6 independent tests). Median indices of neuronal mismatch (iMM), repetition suppression (iRS) and prediction error (iPE), computed from each of the two control sequences (CAS or MAS), and their corresponding p-values (note that p-values are the same for absolute differences and normalized indices, since these indices are median differences between normalized responses, and the non-parametric test is independent of scaling). Values related to predictive neuronal activity are highlighted in bold case, since they represent the most significant result of this research.

773

	fo-IC	*ho*-IC	*fo*-MGB	*ho*-MGB	*fo*-AC	*ho*-AC
# Neurons	22	56	24	35	35	36
# Points	114	523	77	225	250	306
DEV (spk)	2.55	0.99	0.64	0.68	0.95	0.98
STD (spk)	1.93	0.22	0.20	0.14	0.24	0.21
CAS (spk)	2.37	0.97	0.71	0.55	0.77	0.59
MAS (spk)	2.51	0.95	0.90	0.65	0.85	0.52
DEV−STD (spk)	0.62	0.77	0.44	0.54	0.71	0.77
p-val	0.000	0.000	0.000	0.000	0.000	0.000
CAS−STD (spk)	0.44	0.76	0.51	0.40	0.53	0.38
p-val	0.000	0.000	0.000	0.000	0.000	0.000
DEV−CAS (spk)	0.18	**0.019**	−0.07	**0.13**	**0.18**	**0.39**
p-val	0.779	0.020	0.019	0.023	0.019	0.000
MAS−STD (spk)	0.57	0.73	0.70	0.50	0.60	0.31
p-val	0.003	0.000	0.000	0.000	0.000	0.000
DEV−MAS (spk)	0.04	0.04	−0.26	0.03	**0.11**	**0.46**
p-val	0.190	0.155	0.003	0.671	0.049	0.000
iMM_{CAS}	0.127	0.493	0.324	0.496	0.505	0.609
p-val	0.000	0.000	0.000	0.000	0.000	0.000
iRS_{CAS}	0.013	0.461	0.447	0.446	0.398	0.334
p-val	0.000	0.000	0.000	0.000	0.000	0.000
iPE_{CAS}	−0.002	**0.032**	−0.122	**0.050**	**0.107**	**0.275**
p-val	0.779	0.020	0.019	0.023	0.019	0.000
iMM_{MAS}	0.147	0.485	0.303	0.505	0.508	0.611
p-val	0.000	0.000	0.000	0.000	0.000	0.000
iRS_{MAS}	0.091	0.463	0.445	0.494	0.439	0.343
p-val	0.003	0.000	0.000	0.000	0.000	0.000
iPE_{MAS}	0.055	0.023	−0.143	0.010	**0.069**	**0.267**
p-val	0.190	0.155	0.003	0.671	0.049	0.000

774

775

Supplementary Materials

Figure S1: Quantitative comparison between iMM and the "classical" SI. The SI trace is plotted as a function of the DEV/STD ratio, since it does not take into account the control condition. Different iMM traces are plotted (dashed lines), as a function of the relative magnitude of the response to control condition with respect to DEV response (CTR/DEV), from low (CTR=0.2*DEV) to high (CTR=1.2*DEV) hypothetical responses to the control. The two indices (the SI and the iMM for different CTR response magnitudes) take values very close to each other under most conditions, except for very extreme and rare cases in which the response to the control condition is much larger than DEV of much smaller than STD.