Title:

Prediction errors explain mismatch signals of neurons in the medial prefrontal cortex

Authors

Lorena Casado-Román[1,2], *David Pérez-González*[1,2] *& Manuel S. Malmierca*[1,2,3,*]

Affiliations:

[1]Cognitive and Auditory Neuroscience Laboratory, Institute of Neuroscience of Castilla y León (INCYL), University of Salamanca, Salamanca 37007, Castilla y León, Spain

[2]The Salamanca Institute for Biomedical Research (IBSAL), Salamanca 37007, Castilla y León, Spain

[3]Department of Cell Biology and Pathology, Faculty of Medicine, Campus Miguel de Unamuno, University of Salamanca, Salamanca 37007, Castilla y León, Spain

[*]**Correspondence**: msm@usal.es

Abstract

According to predictive coding theory, perception emerges through the interplay of neural circuits that generate top-down predictions about environmental statistical regularities and those that generate bottom-up error signals to sensory deviations. Prediction error signals are hierarchically organized from subcortical structures to the auditory cortex. Beyond the auditory cortex, the prefrontal cortices integrate error signals to update prediction models. Here, we recorded neuronal activity in the medial prefrontal cortex of the anesthetized rat while presenting oddball and control stimulus sequences, designed to separate prediction errors from repetition suppression effects of mismatch responses. Robust mismatch signals were mostly due to prediction errors. The encoding of a regularity representation and the repetition suppression effect over the course of repeated stimuli were fast. Medial prefrontal cells encode stronger prediction errors than lower levels in the auditory hierarchy. These neurons may, therefore, represent the neuronal basis of a fundamental mechanism of hierarchical inference.

Keywords: auditory processing, predictive coding, prediction error, mismatch negativity, medial prefrontal cortex, anaesthesia, neuronal activity

Introduction

As a fundamental mechanism for survival, organisms must constantly extract regular patterns of sensory input and use that information to form predictions about future input to rapidly detect the occurrence of irregularities and unexpected events[1]. Brains generate perceptual mismatch signals which reflect deviance detection mechanisms from predictions to unexpected sensory inputs[2–5]. Human scalp-recorded mismatch negativity (MMN) can be evoked by an oddball paradigm in which a deviant sound is embedded in a sequence of repeated standard sounds. MMN appears without conscious attention, even during sleep, coma, or anesthesia[6–9]. Mismatch responses are widely regarded as an indicator of prediction error (PE) signals in terms of predictive coding, a neurobiologically informed theory of perceptual inference[2,10–12]. Under the predictive coding framework, brains are considered hierarchically organized systems within distributed neuronal networks. Each level compares the bottom-up sensory input from the level below with the top-down predictions of the higher level. During perceptual learning, when an incoming input is repeated, statistical regularities of the natural world are learned such that forthcoming inputs are predicted. PE signals are suppressed through a process of repetition suppression (RS). Conversely, if a new stimulus arrives, such as a deviant event, there will be a failure to predict the bottom-up input and to suppress the PE signal. This PE signal is used to optimize the encoding of sensory causes at higher levels[8,10,11]. Thus, oddball mismatch signals are the result of the two components: RS and PE[13]. This generative system of RS and PE is hierarchically organized from subcortical structures to the auditory cortex as demonstrated in awake and anesthetized rodents[14].

The MMN has interacting generators in a frontal-temporal network[15–17], where the prefrontal cortex (PFC) is believed to play an essential role in contextual processing and thus, in predictive coding[18]. Studies in humans have shown evidence of PE- and even prediction signals in frontal cortices (FCs)[19–23]. Some of those studies were carried out using human electrocorticography (ECoG). ECoG reflects the population network dynamics because it records high-gamma activity (80–150 Hz) and measures multiunit- but also synaptic activity[24]. Additionally, the clinical electrode placement is heterogeneous and limited by the patient´s requirements, making comparisons of time course and signal magnitude difficult or impossible. Thus, for obvious methodological and ethical reasons, these experiments are hampered in their fine temporal and spatial resolution to disentangle the neuronal correlates of predictive signals.

To date, spike events have not been studied to estimate whether predictive activity independent from attention emerges at the cellular level in PFC. Animal studies using oddball sequences are scarce and most of them have been limited to the mesoscopic scale. ECoG recordings in the macaque PFC demonstrated actual predictive activity[25] and local field potentials (LFP) in awake rats showed prefrontal mismatch responses to passive oddball sounds[26–28]. However, one study in the PFC of the macaque with deep electrode recording failed to replicate the strong mismatch signals found in primates[29]. Also, these animal studies lacked appropriate controls to interpret the generative system of PE[13,30]. Under the oddball paradigm, two components contribute to the extraction of the standard-repetition rule: the RS, or attenuation of the evoked response to a specific repeated stimulus feature, and more complex forms of predictive activity, such as the generation of statistical inferences. Hence, the details and neuronal contribution of RS and PE at the PFC are currently unknown.

In the present account, we recorded neuronal activity in the rat medial prefrontal cortex (mPFC) under an oddball stimulus paradigm and adequate control sequences that allowed us to separate

mismatch signals into RS and PE[14]. We found robust mismatch signals that were explained by maximal indices of PE across the mPFC. Our findings support the notion that mPFC neurons detect unpredictable deviations from the auditory background. These cells may, therefore, represent the neuronal basis of predictive activity in FCs.

Results

Context-dependent responses across fields in mPFC.

To seek experimental evidence for predictive coding signals in the mPFC, we recorded neuronal activity across all fields in 33 urethane-anesthetized rats. We recorded 83 multiunits (AGM [medial agranular cortex]: 25; ACC [anterior cingulate cortex]: 20; PL [prelimbic cortex]: 20; IL [infralimbic cortex]: 18) and tested 384 tones (AGM: 13; ACC: 90; PL: 81; IL: 81) as part of oddball paradigms and suitable control sequences, namely, cascade and many-standards conditions (Fig. 1). Figure 2 illustrates 5 examples of electrolytic lesions in three Nissl-stained sections at different recording sites across the mPFC. Regardless of their anatomical location in all mPFC fields, we found sound-driven neuronal activity to pure tones (0.6–42.5 kHz; 25–70 dB SPL). We assessed significantly increased responses to sound by comparing baseline-corrected spike counts after stimulus presentation against a simulated null peristimulus time histograms (PSTH) with a firing rate equal to the baseline (detailed in Methods). We tested the neuronal response to each pure tone as either being deviant, standard, or part of a control sequence and only included those frequency tones that demonstrated a significant response to any condition. Thus, 86.2% (331/384) of deviant stimuli evoked a significant neuronal discharge, while 26.6% (102/384) of standard stimuli, 26.6% (102/384) of many-standards stimuli and 32.3% (124/384) of cascade stimuli evoked a firing rate significantly larger than the baseline in the mPFC population. This finding highlights a profound reduction in the evoked responses to standard, cascade, and many-standards conditions, as compared to deviant sounds of the

same frequency and intensity (Fig. 3a). While typically the deviant sounds evoked the largest responses, in some cases other conditions also evoked robust firing rates, as the cascade descending control example from the PL field (Fig. 3a; cyan line). The latency and amplitude of mPFC responses differed per multiunit. mPFC multiunits frequently presented a peak of maximum firing followed by a decay (e.g., the AGm, ACC and PL examples in Fig. 3a), whereas some cells showed a period of sustained firing as the IL example in Fig. 3a. For each multiunit, we tested 1–8 pairs of deviant-standard combinations among the 10 frequencies comprising the control sequences. We calculated the baseline-corrected spike counts per trial from 100 to 600 ms after sound onset (see Methods, Fig. 1c). Overall, within the same multiunit, spike counts were larger for deviant tones, either ascending or descending, than for any other condition regardless of the sound frequency (Fig. 3b). Accordingly, neurons in the mPFC showed poorly tuned frequency response areas. We tested a broad range of sound frequencies and sequences for each multiunit and found that mPFC cells exhibit a strong and dynamic sensitivity for stimulation context rather than for spectral processing.

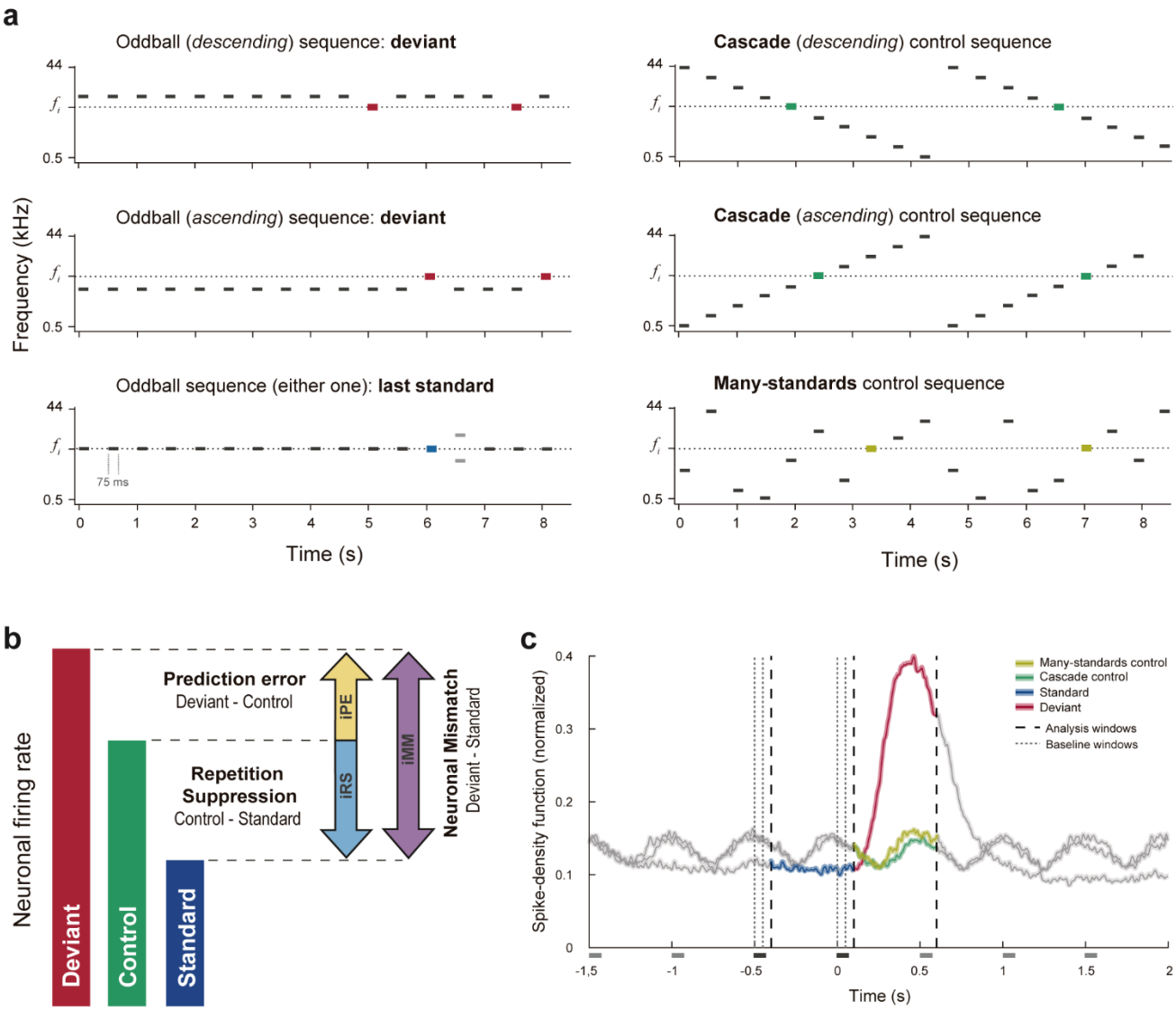


Fig. 1. Experimental design**. a** Stimulation paradigms were generated by selecting 10 pure tones, spaced by 0.5 octaves, for each multiunit. Among the 10 tones, a target tone f_i (i=1–10; highlighted in color) was part of an oddball ascending or descending sequence as a deviant or standard stimulus (left column). Deviant events were presented in a pseudo-random fashion with a minimum separation of three standard events. Two conditions controlled for the repetition suppression effect exerted over the target tone: cascade and many-standards sequences (right column). The many-standards sequence comprises the random presentation of the 10 tones, and thus, the target tone is unpredictable. In the cascade sequences, the 10 tones are presented in a regular succession of ascending or descending frequency, making the target tone completely predictable. Responses to deviant events are compared to frequency-matching tones presented as the last standard event (i.e., the standard preceding a deviant tone), the many-standards control, and the corresponding cascade ascending or descending sequence. **b** Decomposition of the index of neuronal mismatch (iMM), under the framework of the predictive coding theory. The iMM is calculated as the difference between the responses to deviant and last standard

conditions. The index of repetition suppression (iRS) is calculated as the subtraction of the standard response to one of the controls. The remaining part of the mismatch signal not assigned to iRS is the index of prediction error (iPE). **C** Population firing rate as normalized spike-density functions (mean ± SEM), n = 83 multiunits in the mPFC) for 8 consecutively presented tones within a cascade sequence (green trace), many-standards sequence (yellow trace) and oddball sequence with the last standard tone (blue trace) before and after a deviant event (red trace). Tones are illustrated as gray horizontal lines and lasted 75 ms with an inter-stimulus interval of 500 ms. Baseline spontaneous firing rates were computed in the time window from 0-50 ms after stimulus onset (gray dotted lines). Analysis windows to compute spike counts (black dashed lines) were selected 100-600 ms after stimulus onset. Adapted from Parras *et al.*, (2017); Carbajal & Malmierca, (2018)[13,14].

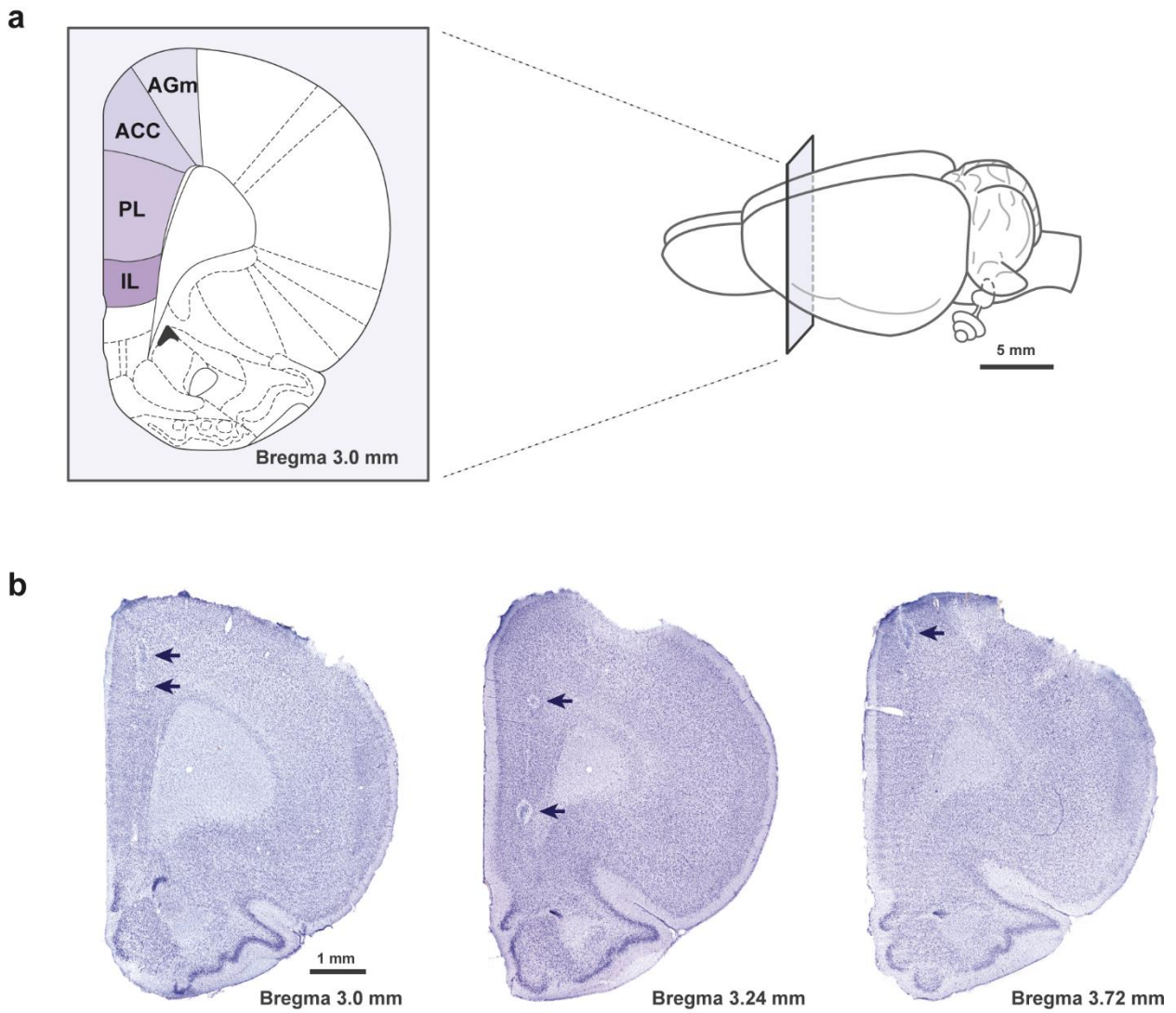


Fig. 2. Anatomical location of neurophysiological recordings. **a** Location of fields in the rat mPFC on a brain atlas with respect to a schematic lateral view of the brain. Adapted from Parras *et al.*, (2017); Carbajal & Malmierca, (2018)[13,14]. **b** Coronal Nissl-stained sections of electrolytic lesions performed at three rostrocaudal levels and penetration depths in the left mPFC of three different animals. Arrows point to the electrolytic lesions. Left section, two lesions in the ACC at 3 mm anterior to bregma, 0.5 mm mediolateral (ML) and 0.7 and 1.5 mm dorsoventral (DV). Middle section, lesion in the PL at 3.24 mm anterior to bregma, 0.9 mm ML and 1.9 mm DV and lesion in the IL at 3.24 mm anterior to bregma, 0.8 mm ML and 4.1 mm DV. Right section, lesion in the AGm at 3.72 mm anterior to bregma, 0.4 mm ML and 0.4 mm DV.

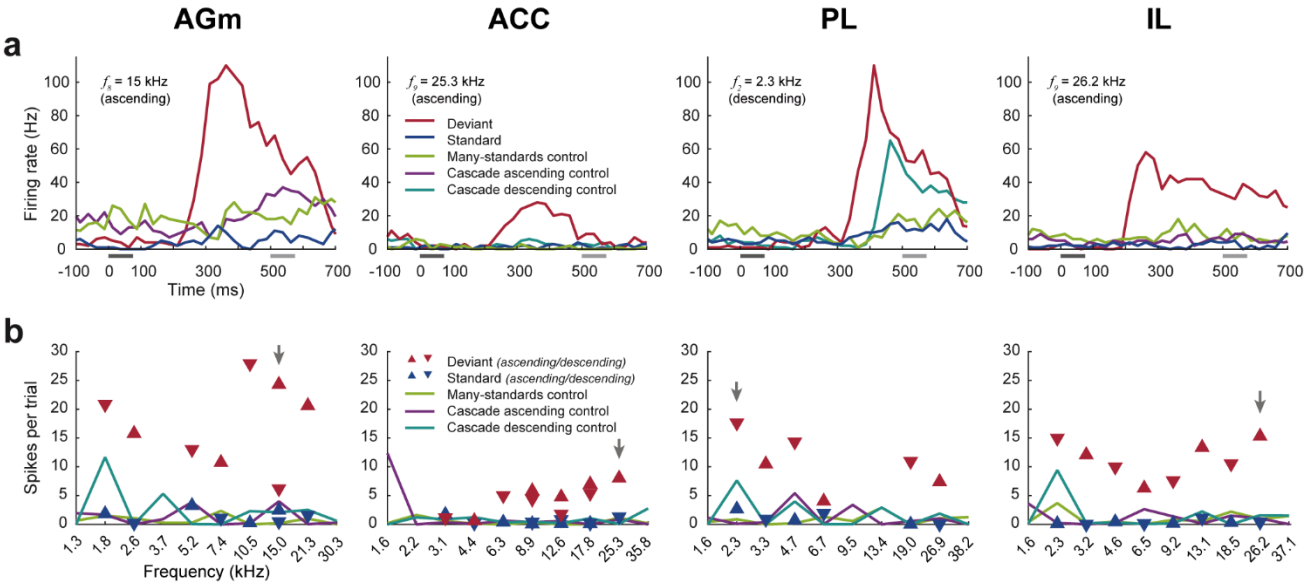


Fig. 3. Context-dependent responses in representative multiunit examples of each field of the mPFC. **a** Peristimulus time histograms comparing the neuronal responses to a pure tone (target tone indicated in each panel) in each condition tested. *Ascending* and *descending* denote whether the target tone's frequency was lower or higher than the previous tone, respectively, within the oddball and cascade sequences. Horizontal gray bars indicate the epoch of sound stimulation; conditions are assigned according to the tone in dark gray. Note that deviant tones, either ascending or descending, evoked a higher firing rate compared to standard, cascade and many-standards tones. Bin size: 25 ms. **b** Neuronal responses of the same examples in *a* (frequency indicated with an arrow), to all tested frequencies and conditions. Each neuronal example was recorded at a fixed sound level, from left to right, 40, 30, 40 and 40 dB SPL. Note that overall, regardless of the sound frequency, responses are larger to deviant tones than to any other condition.

Neurons in mPFC show maximal levels of prediction error signals.

To interpret neuronal responses within the predictive coding framework, we presented the oddball paradigm and control sequences to estimate the contribution of RS and PE to mismatch signals in all mPFC fields. For each pure tone tested, we calculated baseline-corrected spike counts in the time window between 100 to 600 ms after stimulus onset. To allow comparisons across the neuronal population, we normalized the neuronal responses using Euclidean vector normalization to the three conditions, deviant, standard, and controls (either cascade or many-standards). At the population level, deviant stimuli evoked the most robust neuronal discharges in the four fields of the mPFC, being nearly maximum as compared to other conditions (Table 1, Fig. 4a). The median evoked firing rate in response to the deviant stimuli was significantly larger than that to the standard, many-standards and cascade conditions; and all but the deviant condition were not significantly different from each other (within-field multiple comparisons Friedman test; Table 1). Hence, the distribution of spike counts in response to deviant frequencies did not overlap with the remaining conditions, while the distributions of responses to cascade, many-standards, and standard frequencies largely overlapped (Fig. 4a).

Neurophysiological responses to many-standards and cascade sequences were not statistically different from each other within each field (Wilcoxon signed-rank test, Table 1) or considering the whole population sample (Wilcoxon signed-rank test, $n = 384$ tested frequencies, $z = -0.512$, $p = 0.609$). Differences between the responses to deviant and standard stimuli with both controls also yielded similar results (Table 1). Therefore, in the following, we will report only analyses obtained under the cascade control because this sequence not only controls for the presentation rate, as the many-standards condition does, but also for the refractoriness of the response and RS effects due to the predictability of cascade tones (see Methods)[30].

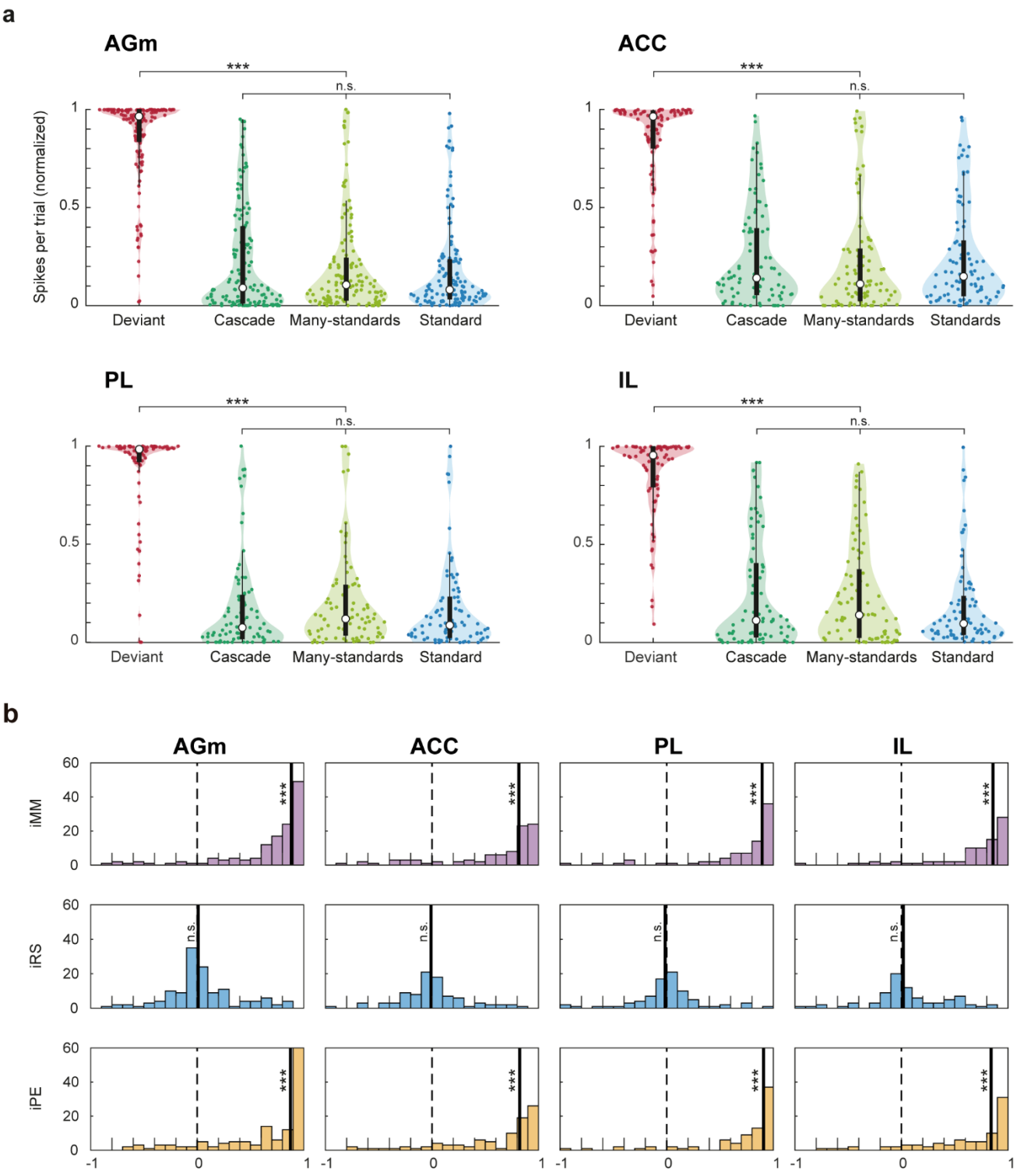


Fig. 4. Maximal prediction error signals at the population level on each field of the mPFC. **a** Density distribution of normalized responses of all frequencies tested as each condition (AGm, n = 132; ACC, n = 90; PL, n = 81; and IL, n = 81 frequencies). Colored dots represent sample points of each measured firing rate. Box plots are shown in black, with white dots denoting the population median. In all fields, deviant responses are significantly larger than both controls and standard responses, which were equivalent. **b** Histograms represent distributions of the three indexes that were computed as response differences between pairs of stimulus conditions shown in *a* (deviant, standard and cascade control). Solid

black lines indicate medians. The iRS is close to zero and virtually negligible in all fields; this indicates that the large iMM values are almost exclusively due to the contribution of the high iPEs in the mPFC. **a-b** Statistical significance denoted as *n.s.* (non-significant) and *** ($p < 0.001$), within-field multiple comparisons Friedman test (see Table 1).

We computed the index of neuronal mismatch (iMM = deviant – standard), which is the classical difference between the standard and deviant tone in the oddball paradigm and equivalent to the classic stimulus-specific adaptation index[14]. Importantly, this design allows decomposing the neuronal mismatch into two fundamental signals under a predictive coding framework, namely, RS and PE. The index of repetition suppression (iRS = control – standard) measures the response decrease from the control condition to the standard condition, which is due to repetition effects[31]. The index of prediction error (iPE = deviant – control) denotes the relative response increase from the control condition to the deviant condition, which represents the error signal generated when a prediction is inaccurate[13,30]. The iMM revealed values close to the index maximum and was significantly larger than zero in all the mPFC fields (within-field multiple comparisons Friedman test, Table 1, Fig. 4b). More remarkably, all these robust mismatch signals were due to strong and highly significant PE signals, because RS signals were absent and not significant in any field; iRS of most tested frequencies were centered around zero with very few extreme positive or negative values (Fig. 4b). By contrast, the distribution of the iPE and iMM per frequency tone showed that the majority of the values were skewed towards the largest values (Fig. 4b). These effects were largely independent of recording location in the mPFC. The median iPE and iMM were slightly larger in the PL than the ACC, whereas all the other paired comparisons between fields did not significantly differ from each other (Wilcoxon rank-sum test; iPE: PL-ACC p=0.024, AGm-ACC p=0.158, AGm-PL p=0.391, AGm-IL p=0.560, ACC-IL p=0.468, PL-IL p=0.146; iMM: PL-ACC p=0.011, AGm-ACC p= 0.069, AGm-PL p=0.324, AGm-IL p=0.513, ACC-IL p=0.259, PL-IL p=0.155). As expected, the

iRS was not significantly different between any pair of fields (Wilcoxon rank-sum test, AGm-ACC p=0.282, AGm-PL p=0.737, AGm-IL p=0.827, ACC-PL p=0.581, ACC-IL p=0.281, PL-IL p=0.562). Additionally, we compared these indices between two groups of cortical layers resulting from a restrictive histological assessment, discarding inconclusive lesions on laminar borders: layers 2/3 (n = 19 multiunits & 92 frequencies) and layers 5/6 (n = 22 multiunits & 113 frequencies). Median iPE, iRS and iMM were comparable between both laminar subdivisions (Median values, layers 2/3: iPE=0.862, iRS=0.011, iMM=0.873, layers 5/6: iPE=0.898, iRS=0.005, iMM=0.904; Wilcoxon rank-sum test iPE: p=0.129, iRS: p=0.585, iMM: p=0.062). These findings strongly support the conclusion that the classically measured signals of mismatch in the mPFC are mostly due to PE signals at the neuronal level.

The temporal dynamics of mismatch and prediction error signals coincide in time.

To identify the overall response pattern of each mPFC field, we computed the population temporal dynamics of the average firing rate as normalized spike-density functions (SDF; Fig. 5a). The neuronal firing rate was very similar between cascade and many-standards controls in all the fields at any time point. The responses to cascade and many-standards tones showed a regular increase and decrease in amplitude during the sequence that suggests an encoding of temporal predictability of the consecutive presented tones (Fig. 5a, see also Fig. 1). Consistently among fields, ~200 to 700 ms after stimulus onset, the population presented a stronger firing for the deviant condition than the cascade, many-standards, and standard conditions. Neurons in mPFC showed a robust firing to deviant events, spanning up to the next tone (Fig. 5a). To analyze the emergence of predictive signals around stimulus presentation within each field, we computed the average iPE and iMM for each tested tone recorded in 35 time-windows of 20 ms width in the epoch of -50 to 650 ms around tone onset (Fig. 5b). We tested the indices for significance against zero (Wilcoxon signed-rank test, FDR-

corrected for 35 comparisons, $p < 0.05$). iPE signals started to be significantly different from 0 at 120 ms in the PL, followed by the IL at 140 ms and later in the AGm and ACC at 180 ms from stimulus onset (Fig. 5b). In all mPFC areas, iPE signals exceeded half of the index maximum for a sustained length, ~180 to 640 ms. Only one iPE was slightly higher than the iMM at 260–280 ms time window in the IL (Fig. 5b, Wilcoxon signed-rank test, FDR-corrected for 35 comparisons, $p < 0.05$). The remaining iMMs were roughly equal to iPEs, and both functions fluctuated homogenously in the entire analysis epoch in all fields. The temporal emergence of the iPE was comparable to the iMM, underpinning the conclusion that the neuronal temporal dynamics of mismatch signals are primarily due to PE signals.

In order to analyze the variability of latencies within the population response, we plotted the SDF per tested frequency tone with heat maps for each sound condition and field (Fig. 5c). Deviant events were the condition with the largest number of sound-evoked responses (86.2% of the tested tones). Therefore, latencies were assessed as the timing of SDF maxima in an analysis window between 100 and 700 ms relative to the onset of the deviant response and then, progressively sorted as increasing time of maximum firing (Fig. 5c). SDF to identical frequency tones in the cascade and standard conditions preserved the same ordering in the heat maps as deviant tones. However, SDFs to standard and cascade conditions were generally weak or even absent, and thus, distinct peaks of maximum firing are unclear in their corresponding heat maps (Fig. 5c). In the mPFC population, latencies of maximum firing rates to deviant events varied notably among frequencies and were uniformly distributed within each field (Fig. 5c). Median latencies were significantly later in the PL (467 ms post-stimulus onset) than in the other fields (AGm 400, ACC 397, IL 389 ms; Wilcoxon rank-sum test, PL-AGm p=4.43E-04, PL-ACC p=4.48E-04, PL-IL p=1.50E-04). Response latencies in the AGm, ACC, and IL were statistically equivalent (Wilcoxon rank-sum test, AGm-ACC p= 0.729, AGm-IL p=0.490, ACC-IL p=0.756). At the population level, this broad range of maximum

neuronal discharges comprises a sustained response window or the maintenance of internal representations over time. This period of activity could be used to update the inference model to unpredicted deviant events via recurrent PE signals[32,33].

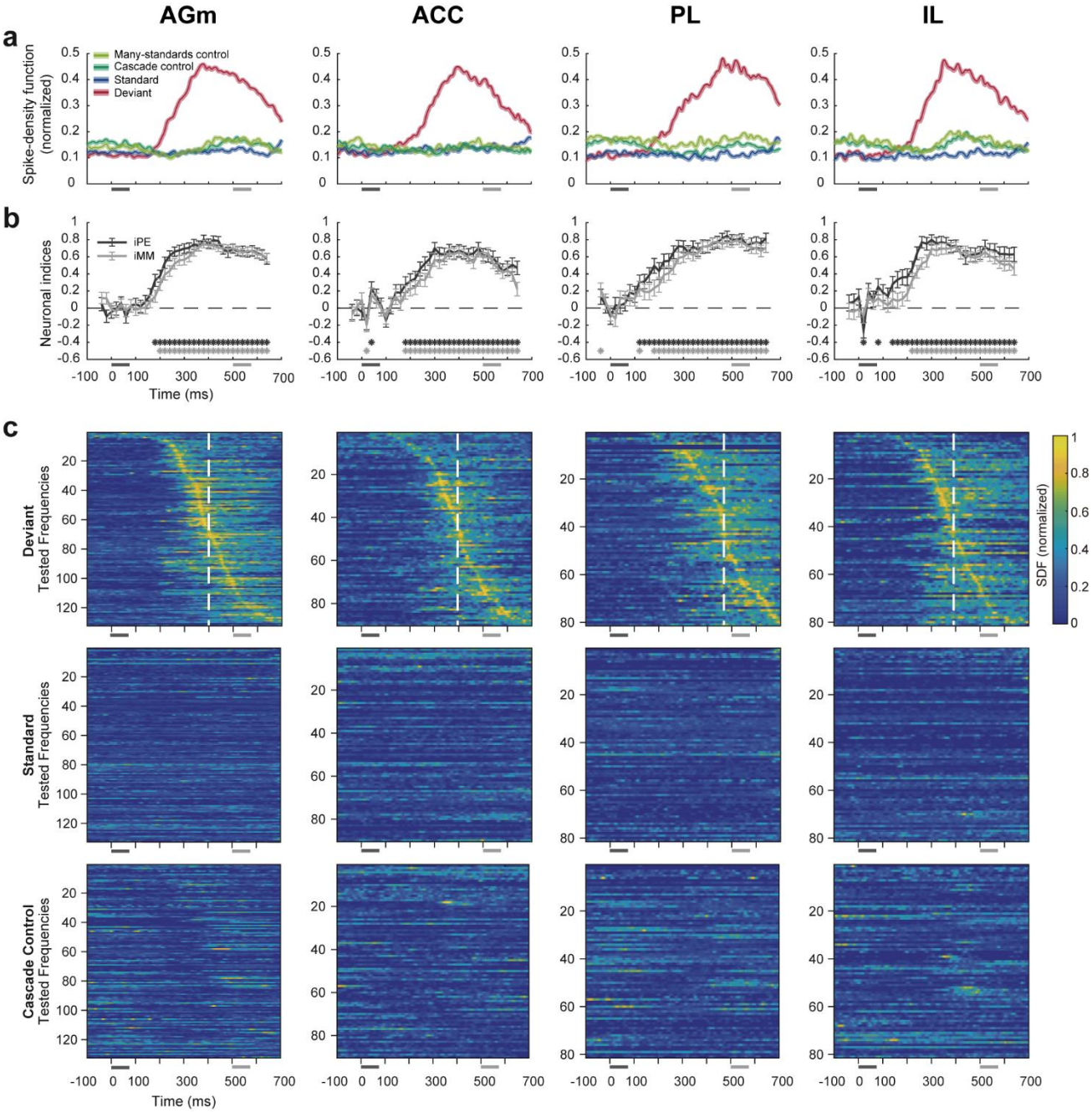


Fig. 5. Temporal population dynamics of mismatch and prediction error signals coincide in time on each field of the mPFC. **a** Population average firing rate as normalized spike-density functions (SDF, mean ± SEM) for the tested conditions. mPFC neurons show a long and robust response to deviant events decaying with the next standard tone. **b**

Time course of the average iMM and iPE computed in 20 ms time-windows (mean ± SEM). Asterisks denote indexes different from 0 (Wilcoxon signed-rank test for 35 comparisons, corrected for false discovery rate = 0.1, $p < 0.05$). Note that iMM and iPE signals were significant for an extended time interval in which both index values fluctuated homogenously within each field. **C** Heat maps showing on each horizontal line the SDF of each tested frequency as deviant, standard or cascade condition. SDFs were sorted from shorter to longer peak latencies in the deviant condition; this sorting was preserved for the standard and cascade conditions. Dashed white lines illustrate the median peak latency for deviant events. Note the absence of clear peaks of maximum firing to the standard and cascade conditions, whereas response latencies to deviant events are variable and uniformly distributed across frequencies. **a-c** Horizontal gray bars indicate the epoch of sound stimulation; conditions are assigned according to the tone in dark gray. Tested frequencies: AGm, $n = 132$; ACC, $n = 90$; PL, $n = 81$; and IL, $n = 81$.

Fast time course of the repetition suppression effect to predictable auditory input.

In order to explore the dynamics of adaptation or RS effect to repetitive stimuli over time, we averaged the responses to deviant, standard, and cascade stimuli across recordings for every trial number within the sequence. Thus, we averaged firing rates at their absolute position within the sequence and generated the time course of responses from the beginning of the sequence within each field (Fig. 6a). The variability of deviant, many-standards and cascade responses in all fields was minimally explained with the tested models (linear, exponential, double exponential, inverse polynomial, and power-law models; adjusted $r^2 < 0.15$). Within these conditions, two tones were never repeated and did not undergo the same influence of the repetitive effect as standard events. By contrast, a power-law model of three parameters adjusted to the variance of standard responses per field ($y(t) = at^b + c$; adjusted r^2: AGm, 0.328; ACC, 0.189 ; PL,0.277 ; IL, 0.392). The IL exhibited the smallest plateau steady-state of responses reached at the end of the sequence (Fig. 6b, dashed line; *c* parameter [with 95% confidence intervals]: AGm, 3.238 [3.172–3.305]; ACC, 2.597 [2.547–2.647]; PL, 2.857 [2.791–2.924]; IL, 2.257 [2.205–2.31]). The ACC and PL cortices showed medium values of the steady-state plateau. We found the highest values of the constant parameter in the AGm

(Fig. 6b). All these values were significantly different from each other across areas. The steady-state values highlight a possible hierarchy for the suppressive effect exerted to the response of repeated standard tone. The deeper the neuronal location was, the stronger the suppressive force they experienced.

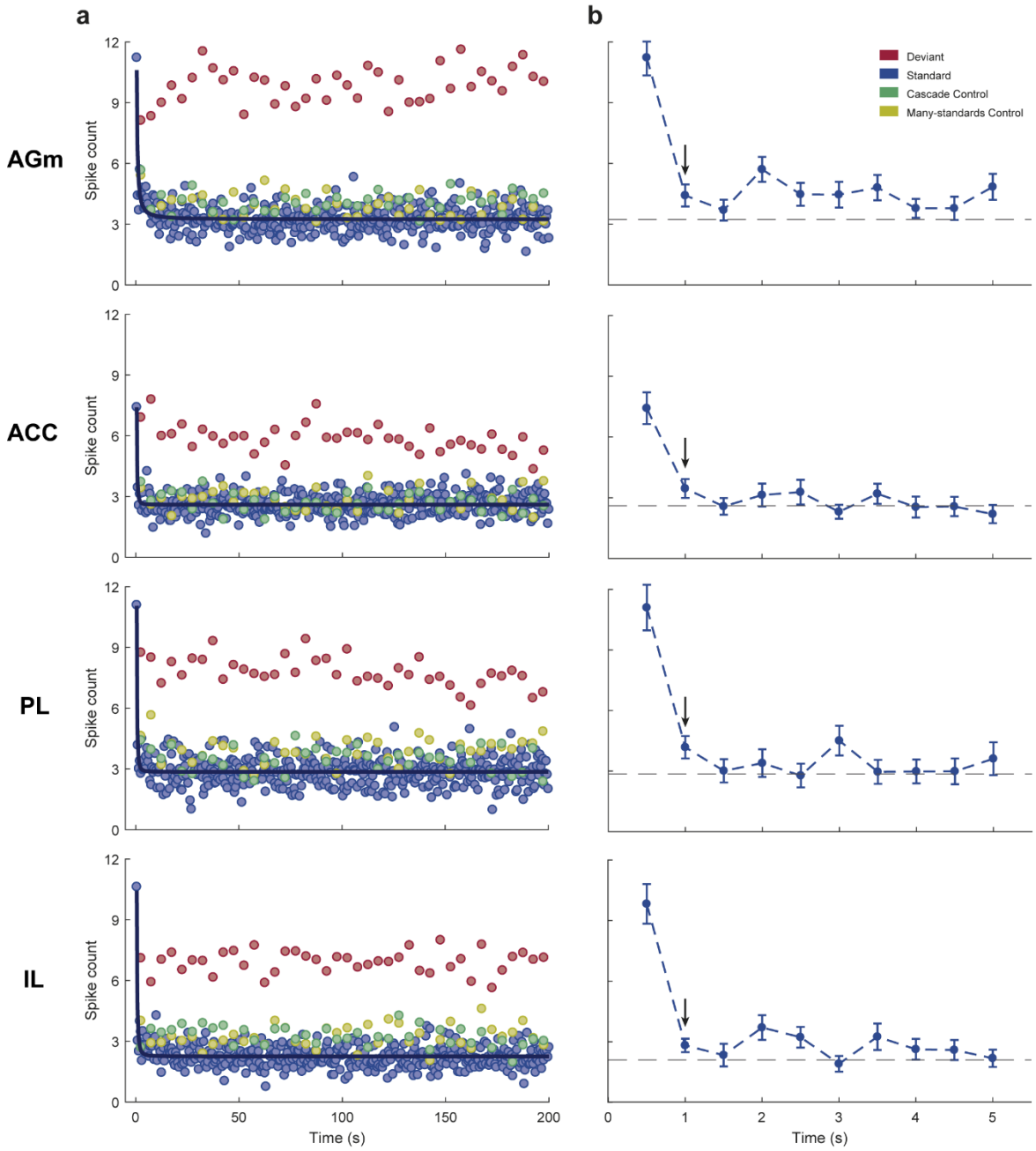


Fig. 6. Fast time course of the repetition suppression effect exerted over standard events. **a** Average baseline-corrected spike counts per trial number of presentation order within the sequence and field of the mPFC. Each dot represents the measured spikes per trial in relation to a given stimulus time from the beginning of the sequence. Tested frequencies:

AGm, *n* = 132; ACC, *n* = 90; PL, *n* = 81; and IL, *n* = 81. The time course of standard responses followed a power-law model of three parameters (see Results). **b** Detail of the average responses for the first 10 standard stimuli (mean ± SEM). Arrows point to the trial in which the response is lower than half of the initial firing rate to the first standard event. Responses are suppressed to half of the initial response with two stimulus repetitions. Dashed lines mark the steady-state parameter of the power-law fit. This level of response suppression was reached after the repetition of three standard tones and maintained to the end of the sequence.

The models demonstrated a comparable rate of RS among fields (*b* parameter [with 95% confidence intervals]: AGm, -1.258 [-1.521 to -0.9939]; ACC, -2.349 [-3.542 to -1.156]; PL, -2.393 [-3.359 to -1.426]; IL, -2.053 [-2.601 to -1.504]). In the mPFC, the repetition effect is very robust and fast. Thus, it only needed a second stimulus repetition to generate a response decay less than 50% of the initial response (AGm 39.4%; ACC 46.6%, PL 37.6%, IL 28.6%; Fig. 6b, arrows). A third repetition further attenuates the response, to levels comparable to the steady-state, where the firing rate remains constant until the end of the sequence (Fig. 6b, note the overlap between mean ± standard error of the mean [SEM] of the third stimulus with the black dashed line). Thus, RS effects played a role for a very short time scale in the mPFC. Only two repetitions of the same tone are required to maximally suppress spiking on the third stimulus repetition and subsequent standard presentations. In other words, two repetitions generate a regularity rule that suppresses neuronal firing or PE signals when acoustic input matches the current prediction.

Strong responses to unpredictable sounds under a regular context of silence.

We studied the effect of yet another type of regular stimulation context on the PE signals generated by unpredictable events. We presented the deviant-alone paradigm as a sequence of deviant tones separated by silent periods of a minimum of 1.925 s (equivalent to 3 silenced standards in the oddball condition with an interstimulus interval of 500 ms and 75 ms deviant tones). In a subset of 9 multiunits (6 rats) from the previously reported data, we tested 39 frequency tones in the deviant-

alone condition and compared them against regular deviant and standard tones of the same frequency across the mPFC. Median spike counts for deviant and deviant-alone conditions were statistically comparable and their response pattern around stimulus presentation was alike at any time point (multiple comparisons Friedman test; Supplementary Table 1, Supplementary Fig. 1). Thus, the unpredictable effect of the sounds is what matters most to generate a strong PE signal when the predictions are incorrect. Hence, silence can be understood as the inferred cause of spontaneous activity in the auditory system and the generative model needs to be updated to any sound to generate the perception of something over nothing[34]. The deviant-alone tone elicits a sufficiently robust change in evidence to update the model with high PE signals comparable to those required to update the prediction of standard tones when a deviant event is presented.

Discussion

In this study, we recorded neuronal activity while presenting an oddball stimulus paradigm across all mPFC fields in the anesthetized rat. Neurons in mPFC showed robust neuronal mismatch signals to unpredictable events. We performed a quantitative separation of neuronal mismatch signals into PE and RS components with the cascade and many-standard control sequences as we did previously at lower levels of the auditory hierarchy[14]. We found that maximal iMMs are almost exclusively due to iPEs. We also verified that this PE mechanism generalizes to multiple frequency sounds within the same neurons. At the population level, we observed a delay period of sustained activity beyond the presentation of deviant events, which may represent a time window that eliminates PEs progressively. Importantly, we found that RS over repeated stimuli was extremely fast, such that only two repetitions of the standard tone suffice to encode a regularity representation. The fact that this generative system of PE signals is present in urethane-anesthetized rats, in the absence of

behavioral relevance or wakefulness, suggests the mechanism of hierarchical inference as a fundamental process of the rat mPFC.

In the rat mPFC, we found maximal iMMs, which were disentangled into maximal iPEs and non-significant iRSs using two controls, the many-standards and cascade sequences [30,35]. These indices were computed with robust responses to oddball deviant events and minimal responses to standard and control events. The many-standards and cascade sequences were analogously encoded in the mPFC as it was also shown in human FCs[36]. Both controls are probably processed as sequences where alternation is more likely than repetition[36–38]. Previous studies using suitable and similar controls to separate frontal mismatch signals into PE and RS were limited to the mesoscopic network level. LFP recordings in the awake rat showed stronger responses to deviant than to many-standards stimuli[27,28]. A recent study demonstrated specific responses to deviant stimuli and not to the many-standards control in some ECoG electrodes in the PFC of alert macaques[39]. Despite the methodological and species differences, the previous results are similar to ours. ECoG is a population network measurement that records high-gamma activity (80–150 Hz), reflecting multiunit- and synaptic activity[24]. Only one study has reported neuronal activity under the oddball paradigm in the PFC and the experiments yielded inconclusive results on how local are the generators of mismatch signals because few neurons demonstrated a mismatch component[29]. These experiments were confined to the dorsolateral PFC and the selected stimuli may explain the differences from our result in rat and even from previous primate evidence. Thus, robust mismatch signals have been recorded in alert macaques using ECoG even in the same dorsolateral PFC[25,39] and humans (Phillips et al, 2016; Dürschmid et al, 2016; Nourski et al, 2018a). Importantly, human data also support our findings at the neuronal level showing strong iPEs in the mPFC. Human ECoG recordings in PFC demonstrated PE signals by comparing predictable deviants stimuli versus unpredictable deviant stimuli immersed in a train of repeated tones[19]. Although our stimulation

protocols are quite different from those employed by Dürschmid and colleagues (2016), their predictable deviant and standard tones yield comparable responses, which is analogous to our finding of similar responses between predictable cascade and standard tones[19]. Hence, our results are consistent with a generative mechanism of PE signals that are largely abstracted from low-level physical sound features and highly modulated by context in the mPFC[18,32].

MMN signals are currently best interpreted as evidence of predictive activity in the form of PEs[2,8]. The MMN has interacting generators in the temporal and prefrontal networks[16,17], but the relative contribution of the PFC is especially unclear and has not been studied thoroughly at neuron-level resolution. The high iMMs systematically found across the rat mPFC may underlie the large-scale mismatch responses of FCs. Accordingly, LFP recordings in the mPFC of the awake rat also demonstrated mismatch signals[26]. Deviant evoked LFPs were significantly larger than standards tones at ~40 ms (peak N1) after stimulus onset with the strongest response in the time window of 100–500 ms[26]. Our responses match the period of ~100–500 ms and the N1 peak may be reduced or delayed under anesthesia as seen in human data[40]. Likewise, the spiking activity recorded in our project may be less sensitive to N1 refractoriness effects or earlier sources located on temporal cortices compared to the large-scale LFP activity in the awake rat[19,26,41]. Therefore, the neurons recorded in the current study may represent the neuronal correlate of mismatch signals classically measured at the frontal network scale in rats[26–28], humans[19,22,23] and across sensory modalities[42–45].

Our findings under urethane-anesthesia, in the absence of attention or reward-related behavior, suggest that the mechanism of hierarchical inference is a fundamental process of the rat mPFC. Using the same methodological approach, we recently showed slightly higher iPEs in the awake state compared to urethane anesthesia in the rodent inferior colliculus and auditory cortex (AC)[14]. This finding suggests a modulatory role of alertness and attentiveness in the generative system of PE

signals that probably extrapolates to the PFC [46–49]. Furthermore, we have previously demonstrated that PE hierarchically increased from lemniscal to non-lemniscal subdivisions and from the midbrain to the thalamus to ACs [14]. Our results demonstrate that iPEs are profoundly amplified in mPFC as compared to ACs (~86% vs. ~19% of the index maximum, respectively), while iRSs vanish from ACs (~36%) to the mPFC (Adapted from Parras et al., 2017; Fig. 7)[14]. In other words, responses to our predictable events will be "filtered out" at lower levels of the hierarchy such as ACs to prevent their arrival at a higher cortical processing level, i.e., mPFC. By contrast, deviations from expectations that cannot be explained in lower levels will be forwarded to the mPFC and registered as robust iPEs (Fig. 7)[10,32]. Hence, the rat mPFC acts as a higher level in the hierarchy and PE are propagated in a bottom-up fashion from the AC to PFC as also suggested by human data[19,22,23,50] and that from macaques[25], and in support of the predictive coding framework[10,11,51]. Our results provide a link of empirical auditory evidence at the cellular level for the predictive coding theory between animal models and humans.

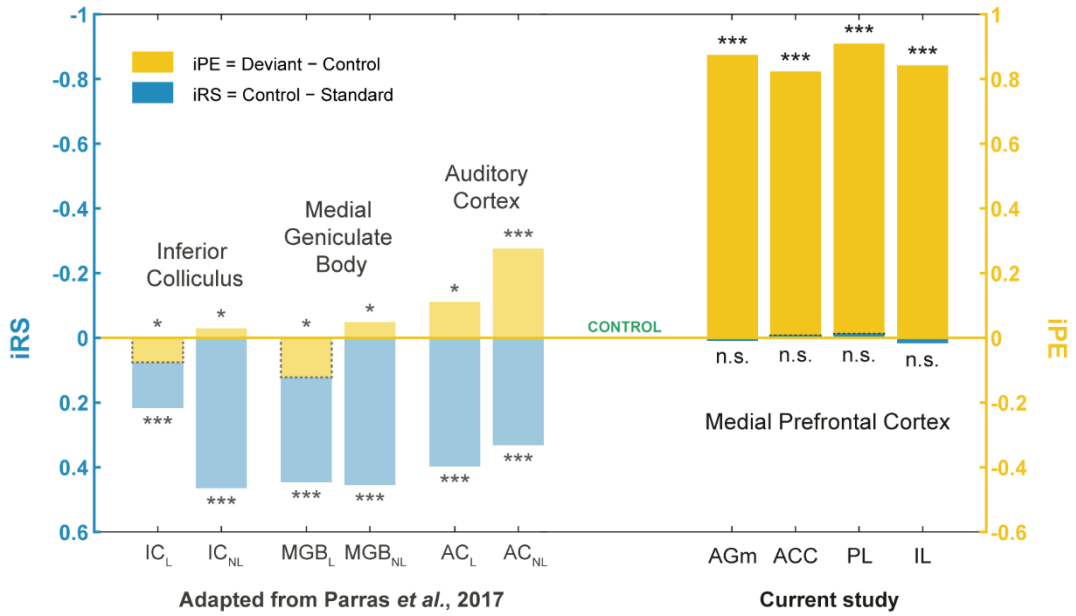


Fig. 7. Summary: Progression of iPEs and iRSs along the auditory hierarchy from auditory midbrain to thalamus to the auditory cortex as shown in Parras *et al.*, (2017) compared to the current dataset from mPFC. Experimental conditions were alike, with stimulus presentation rates adapted to each station: inferior colliculus and medial geniculate body at 4 Hz, auditory cortex at 3 Hz and mPFC at 2 Hz. Asterisks denote statistical significance of indices against zero median

(within-field multiple comparisons Friedman test; *n.s.* [non-significant], $*p < 0.05$, $**p < 0.01$, $***p < 0.001$). The generative system of prediction error signals increases progressively from subcortical structures towards cortical stations and from lemniscal towards non-lemniscal divisions. Higher in the hierarchy, at the mPFC, iPEs are profoundly amplified to reach maximum levels as compared to lower stations. Conversely, iRSs are progressively dampened from inferior colliculus to thalamus and vanish from the auditory cortex to the mPFC.

We also explored the dynamics of the RS effect exerted to the reiteration of standard tones in the rat mPFC. After two repetitions of the standard tone, a regularity representation was encoded and PE signals were efficiently attenuated to subsequent matching stimulation inputs[32]. Frontal human event-related potentials also demonstrated that responses are substantially reduced after a second tone repetition and the third and subsequent tones differed only slightly in their evoked response[52]. Interestingly, this is precisely what we have found in our anesthetized rat model. Our prior work in ACs showed that the non-lemniscal AC in the rat needed 5-6 stimulus repetitions to suppress PE signals maximally, being, therefore, slower than the mPFC[53]. Hence, the present data is consistent with and further support the hierarchical predictive model where higher areas should encode a regular contextual expectation faster and then send a prediction signal to lower areas to cease responsiveness to subsequent matching inputs[10,25,32,54].

In the mPFC, the neuronal population demonstrated a diverse range of response latencies as previously reported in FCs[55]. This range of responses generates a delay-period of persistent and sustained activity beyond the deviant event, which probably holds an internal presentation of the stimuli spanning to the subsequent standard tone[56,57]. Thus, these results suggest that the mPFC generates and maintains an internal memory representation of the stimulation sequence by means of perceptual learning, which suffices to create a PE to those stimuli not matching that representation. Hence, the prolonged neuronal firing may represent an intracortical recurrent message passing among or within hierarchically disposed processing levels to progressively eliminate PE and adapt

the brain's internal model[32,33]. The unique mPFC microcircuitry made of dense recurrent intrinsic connectivity and extrinsic connections with a range of brain areas is an ideal anatomical substrate to support complex computations such as the PE extinction[57–59]. Four main divisions comprise the mPFC with a shift of afferent projections along the dorsoventral axis. A majority of sensorimotor inputs arrive in the AGm and ACC, and mostly limbic inputs connect with the PL and IL[60]. Our results indicated minor differences among fields with the PL possessing larger iPEs and iMMs and longer latencies than the other fields. We could confidently assign about 50% of the recorded multiunits to either layers 2/3 or layers 5/6, and we found no differences in iPE, iRS or iMM among them. Still, the functional relationship across fields, brain areas and cortical layers awaits future studies. For example, a more detailed functional correlation of predictive activity across areas could be tested with dynamic causal modeling and neurons could be allocated to feedforward and feedback layers with current-source density analyses[23,54,61,62]. Also, the specific contribution of different neuronal types such as somatostatin- and parvalbumin-positive interneurons, which are known to be involved in the generation of mismatch responses in primary sensory cortices awaits future studies[63,64].

In summary, our findings demonstrate that mPFC neurons signal the detection of unpredictable stimuli in the auditory scene and support a hierarchical predictive model of perceptual processing, where PE propagates bottom-up and the mPFC acts as one of the highest level in the hierarchy (Fig. 7). We show that neuronal predictive activity in the form of PE underlies mismatch signals. These neurons most likely form the neuronal basis of error signaling in the predictive coding framework. To the best of our knowledge, this is the first study that demonstrates robust mismatch responses and predictive activity independent from attention at the neuronal level in FCs. Our findings unify previous evidence for predictive coding based on cognitive neuroscience and neuronal physiology

and provide an incentive for future studies to discover how these neurons interact in microcircuits to generate sensory predictions.

Methods

Surgical procedures.

All procedures were approved and performed in accordance with the Bioethics Committee of the University of Salamanca, the European Communities Directive (86/609/EEC, 2003/65/EC and 2010/63/EU) and the RD 53/2013 Spanish legislation for the use of animals in research. Experiments were designed following the study of Parras *et al.*, (2017)[14]. We conducted experiments on 33 female Long-Evans rats aged 9–17 weeks with body weights between 200–300 gr. Rats were anesthetized with urethane (1.9 g/kg, intraperitoneal, ip). To ensure a stable deep anesthetic level, we administered supplementary doses of urethane (~0.5 g/kg, ip) when the corneal or pedal withdrawal reflexes were present. Urethane preserves balanced neural activity better than other anesthetic agents having a modest balanced effect on inhibitory and excitatory synapses [65]. Normal hearing was verified with auditory brainstem responses (ABR) recorded with subcutaneous needle electrodes, using a RZ6 Multi I/O Processor (Tucker-Davis Technologies, TDT) and processed with BioSig software (TDT). ABR stimuli consisted of 0.1 ms clicks at a 21/s rate, delivered monaurally to the right ear in 10 dB steps, from 10 to 90 decibels of sound pressure level (dB SPL), using a closed-field speaker. Every 10 hours, we administered 0.1 mg/kg of atropine sulfate (subcutaneous), 0.25 mg/kg of dexamethasone (intramuscular) and 5–10 ml of glucosaline solution (subcutaneous) to ameliorate the presence of bronchial secretions, brain edema and prevent dehydration, respectively. Animals were artificially ventilated through a tracheal cannula with monitored expiratory [CO_2] and accommodated in a stereotaxic frame with hollow specula to facilitate direct sound delivery to the ears. Rectal temperature was maintained at ~37 ºC with a homeothermic blanket system (Cibertec).

We surgically exposed bregma by making an incision in the scalp at the midline and retracting the periosteum. A craniotomy of ~3 mm in diameter was performed above the left mPFC and the dura was removed.

Neurophysiological recordings.

We recorded neuronal activity to look for evidence of predictive coding signals under acoustic oddball stimulation across fields of the mPFC of the urethane-anesthetized rat: AGm, ACC, PL and IL. The rodent mPFC combines anatomo-electrophysiological elements of the primate dorsolateral PFC and ACC at a rudimentary level [66]. Experiments were conducted in an electrically shielded and sound-attenuating chamber. Recording tracts were orthogonal to the brain surface of the left mPFC: ~2.5–4.68 mm rostral to bregma, ~0.2–1.8 mm lateral to the midline and ~0.2–4.5 mm dorsoventrally. Therefore, we covered the four fields of the mPFC and various cortical layers (II-VI). We performed extracellular neurophysiological recordings with glass-coated tungsten microelectrodes (1.4–3.5 MΩ impedance at 1 kHz). We used a piezoelectric micromanipulator (Sensapex) to advance a single electrode and measure the penetration depth. We visualized electrophysiological recordings online with custom software generated with MATLAB (Mathworks) and OpenEx Suite (TDT). Multiunit activity was extracted automatically by manually setting a unilateral action potential threshold above the background noise as an accurate estimation of neuronal population dynamics [67]. Analog signals were digitalized with a RZ6 Multi I/O Processor, a RA16PA Medusa Preamplifier and a ZC16 headstage (TDT) at 12 kHz sampling rate and amplified 251x. Neurophysiological signals for multiunit activity were band-pass filtered between 0.5 and 4.5 kHz.

Histological procedures and verification of recording sites.

The neuroanatomical location of the recording tracts was marked with electrolytic lesions. Post-mortem brains were fixed with 4% paraformaldehyde in phosphate-buffered saline and cryoprotected in 30% sucrose. 40 µm sections were cut in the coronal plane with a freezing microtome and Nissl-stained with 0.1% cresyl violet to visualize cytoarchitectural landmarks. Histological assessment of the electrolytic lesions to any of the fields of the mPFC was processed blindly to each animal history. Multiunits locations were assigned to AGm, AC, PL or IL within a rat brain atlas, accordingly with the histological verification and the stereotaxic coordinates in the three axes of recording tracts[68].

Experimental design and stimulation paradigms.

The sound stimuli were generated using the RZ6 Multi I/O Processor (TDT) and custom software programmed with OpenEx Suite (TDT) and MATLAB. Sounds were presented monaurally in a close-field condition to the ear contralateral to the left mPFC, through a speaker. We calibrated the speaker to ensure a flat spectrum up to ~73 dB SPL between 0.5 and 44 kHz, and the second and third signal harmonics were at least 40 dB lower than the fundamental at the loudest output level.

Trains of white noise bursts of 75 ms duration with 5 ms onset and offset ramps were presented to search for neurophysiological responses to acoustic stimuli. While searching, sound presentation rate and intensity were modified online to prevent strong response adaptation. All experimental paradigms consisted of pure tones of 75 ms duration with 5 ms onset and offset ramps at a presentation rate of 2 Hz. To identify neurons suitable for recording, we computed their frequency-response area or receptive field, which consisted of tones of various frequency and intensity combinations that ranged from 1 to 44 kHz (in 4–6 frequency steps/octave) and 0 to 70 dBs (10 dB steps) and were presented randomly with 1–3 repetitions per tone. We found sound-driven multiunits, but neurons in the mPFC did not show clear auditory receptive fields for pure tones. Thus, for each multiunit, we selected tones within the audible range to generate control sequences

(Supplementary Fig. 1). Control sequences consisted of 10 tones evenly spaced by 0.5 octaves delivered at a fixed sound intensity for the same multiunit and varied among multiunits. Among those 10 tones, we selected pairs of consecutive frequencies to generate oddball sequences. Both control and oddball sequences lasted 400 tones with a 500 ms interstimulus interval (2 Hz). Each target or studied tone of a specific frequency was presented as a deviant or standard condition within an oddball paradigm, as well as part of the many-standards and cascade control sequences. This approach allowed comparing the same physical stimuli within various stimulations contexts. We presented the sequences in a random fashion, with periods of ~12 mins of silence between sequences to minimize long-term habituation effects[69].

We used the oddball paradigm to find neuronal evidence of predictive coding signals to the violation of regularity rules. Oddball sequences consisted of frequently repeating stimuli (standard tones) presented with a 90% probability, which were pseudo-randomly interleaved with rare events (deviant tones) occurring with a 10% probability (Fig. 1a). Sequences started with a minimum of 10 repetitions of standard stimuli, and a minimum of 3 standard events separated deviant tones. We classified deviant sounds as ascending or descending depending on whether the frequency of the standard tone was lower or higher, respectively (Fig. 1a).

We generated the many-standards and cascade controls as sequences with 10 tones in a different presentation order to account for the contribution of RS effect of mismatch signals exerted to the target tone. Thus, we can identify the remaining part of the mismatch signal that is not explained by the RS effect as the PE signal (Fig. 1b). Both sequences control for the presentation rate because they were delivered at the same rate than the oddball paradigm. The many-standards control is the consecutive presentation of blocks of 10 tones randomly ordered within the block (Fig. 1a) [35]. The target tone was among those ten tones and presented at the same ratio as the deviant condition. The

frequency separation (in octaves) between the tones in the many-standards sequence was equal to the separation between deviant and standard in the oddball sequence. The target tone is unpredictable in both oddball and many-standards sequences. The many-standards control cannot establish a prediction or internal rule as the oddball paradigm because an unpredictable sequence of tones replaces the repetition of standard tones. Therefore, the many-standards control cannot account precisely for the undergone RS effect during the oddball sequence.

We also used two cascade control sequences, which consist of the regular presentation of 10 tones (the same as in the many-standards sequence) in ascending or descending frequency succession (Fig. 1a) [30]. PE signals to the target tone are minimized in the cascade sequence because it is embedded in a predictable context, which conforms the internal representation rule and undergoes a RS effect. Ascending and descending deviant targets were compared with the corresponding ascending and descending cascade paradigms. Thereby, we controlled for the state of refractoriness and pitch gliding effects that the preceding tone could exert over the target stimulus. In sum, the cascade sequence makes a more rigorous control (for review [13]).

After controlling for RS effects on target tones, we studied the effect of yet another type of regular stimulation context on the PE signals generated by deviant events. Accordingly, in a subset of the recorded multiunits, we presented the deviant-alone paradigm [70]. This paradigm consisted of the same pseudo-random sequence of deviants as oddball paradigms in which all the standard tones were completely silenced. Thus, deviant events were played consecutively with variable time spans separated by a minimum of 1.925 s (equivalent to 3 silenced standards in the oddball condition with an interstimulus interval of 500 ms and 75 ms deviant tones). Although this deviant-alone control does not account for the level of refractoriness or RS effects, it elicits a sensory-memory violation within a regular context to generate a PE. In consequence, we can compare the effect of disrupting a

prediction by a deviant event in two stimulation contexts: under a regular presentation rate of a different sound frequency that causes RS or continuous silence that does not generate RS.

Data analyses and statistical testing.

Offline data analyses were performed with MATLAB functions, the Statistics and Machine Learning toolbox and custom-made MATLAB scripts. Computing PSTHs with the 40 trial repetitions, we measured multiunit responses to each tested tone and condition (deviant, standard, many-standards and cascade stimuli). In the case of standard tones, we analyzed the last standard before a deviant event to have a comparable number of stimulus repetitions. PSTHs were smoothed with a 6 ms Gaussian kernel in 1 ms steps to calculate the spike-density function (SDF) over time ("ksdensity" function). Thereby, we obtained the mean and standard error of the mean (SEM) of spiking rates from -100 to 700 ms around tone onset. The SDF of the deviant responses of the mPFC population showed a response latency of ~150 ms with a sustained firing spanning up to the next tone (Fig. 1c). We calculated the baseline spontaneous firing rate as the mean firing rate from 0–50 ms during the tone presentation. We measured baseline-corrected spike counts as the spiking activity that exceeded the firing rate of the baseline. This baseline-corrected spike count is the area above the baseline and below the SDF in the period of 100–600 ms after stimulus onset. To avoid overlap of consecutive tone responses, the response analysis window preserved the interstimulus interval of 500 ms and was delayed 100 ms (Fig. 1c).

Tested sound frequencies recorded within each multiunit that did not respond with a significant enhancement from baseline to at least one of the conditions (deviant, standard, cascade or many-standards) were excluded from the analyses. To test for the statistical significance of the baseline-corrected spike count, we performed a Monte Carlo simulation, which is a probability simulation that withdraws numerical values from several random samplings. We simulated 10000 PSTHs with a

Poisson model of a constant firing rate equivalent to the baseline spontaneous spiking activity and thus, a null distribution of baseline-corrected spike counts was generated from the PSTHs. We computed a *p*-value for the original baseline-corrected spike count as $p = (g + 1)/(N + 1)$, where g is the count of null measures ≥ baseline-corrected spike count and $N = 10000$ is the size of the null sample. The significance level was set at $\alpha = 0.05$.

Although the many-standards and cascade sequences controlled for RS effects to target tones, we can separate the neuronal mismatch signal into RS and PE with just one control (Fig. 1b). Both conditions outcomes were comparable and we limited our analyses to the cascade condition, which is a more rigorous control (see previous section and Results). To compare across neuronal populations, we normalized responses of each multiunit to the same frequency tone in the three conditions as follows:

$$Normalized\ Deviant = Deviant/N$$

$$Normalized\ Standard = Standard/N$$

$$Normalized\ Control = Control/N$$

where

$$N = \sqrt{Deviant^2 + Standard^2 + Control^2}$$

is the Euclidean norm of the vector defined by the deviant, standard and cascade responses. Thereby, normalized responses are the coordinates of a 3D unit vector defined by normalized deviant, normalized standard and normalized cascade responses that ranged between 0 and 1. This normalized vector has an identical direction to the original vector defined by the non-normalized data and equal proportions among the three response measurements.

With these normalized baseline-corrected spike counts, we next computed the indices of neuronal mismatch (iMM), repetition suppression (iRS), and prediction error (iPE) as:

$$iMM = Normalized\ Deviant - Normalized\ Standard$$

$$iRS = Normalized\ Control - Normalized\ Standard$$

$$iPE = Normalized\ Deviant - Normalized\ Control$$

Index values ranged between -1 and 1 and facilitated the quantitative decomposition of neuronal mismatch into RS and PE since

$$iMM = iRS + iPE$$

Parras and colleagues (2017) proved that the iMM is extensively equivalent to the classic $Stimulus - specific\ adaptation\ index = (Deviant - Standard)/(Deviant + Standard)$, and its basis is not simply adaptation but predictive activity in higher-order areas. Lastly, to analyze the emergence of predictive signals around stimulus presentation, we also calculated the average iPE and iMM in 35 time-windows of 20 ms width from -50 to 650 ms relative to event onset.

In order to discern the response latencies among fields of the mPFC, we measured latencies as the timing of SDF maxima in the analysis window from 100–700 ms relative to deviant stimulus onset and then, sorted progressively as increasing time of maximum firing. SDF to the same frequency tones in the cascade and standard conditions preserved the same ordering as deviant events.

Since the present data were not normally distributed, we assessed statistical significance with non-parametric or distribution-free tests. We specifically used the Wilcoxon signed-rank, the Wilcoxon

rank-sum and the Friedman test for comparisons of baseline-corrected spike counts, normalized responses, iMM, iRS, iPE and response latencies. Statistical significance was assessed at *p*-values < 0.05. We corrected *p*-values for false discovery rate (FDR = 0.1) of multiple comparison tests with the Benjamini-Hockberg method.

To verify that the current sample size was sufficient to identify significant PE signals, we estimated the sample size required per field. To obtain a statistical power of 0.8 for the iPE, we used the 'sampsizepwr' function in MATLAB as follows:

$$MinSampleSize = sampsizepwr(\text{'t'}, [0\ std(iPE)], \max\left(0.05, abs\left(mean(iPE)\right), 0.8\right)$$

where iPE is the distribution of iPE values for our sample of frequency tested tones per field (Table 1). Note that our sample sizes were ~14 times larger than the required sample due to the homogeneity of neuronal responses of the mPFC. As mentioned, these recordings were also diverse in anatomical localization and therefore, we consider the current sample size as a representative sample of each field.

To analyze the time course of adaptation or RS effect, we measured the deviant, standard and control responses of each target tone as average baseline-corrected spike counts for every trial number within the sequence, for each field separately [53]. We included all the standards tones, not just the last standard before a deviant event as previously. Thereby, we ordered average spike counts at their absolute position within the sequence and generated the time course of responses from the beginning of the sequence. Then, we fitted these time series to various models, namely, linear, exponential, double exponential, inverse polynomial and power-law with two or three coefficients. We used the

"fit" function in MATLAB that computes the confidence intervals of the fitted parameters and the adjusted-r^2, the coefficient of determination of the function fit.

For the additional data set with the deviant-alone paradigm, tests of sound-driven enhanced responses, SDF, baseline-corrected spike counts and normalized responses followed the same previously described analyses. This time, the three compared conditions were the deviant-alone, deviant and last-standard tones. Since this was an additional experiment to compare the influence to deviant responses of different stimulation contexts, the whole sample was merged along the mPFC (Table S1).

Code availability. The scripts and functions written in MATLAB to generate the results and analysis during the current study are available from the corresponding author on reasonable request.

Data availability. The data sets generated and analyzed during the current study are available from the corresponding author on reasonable request.

References

1. Whitmire, C. J. & Stanley, G. B. Rapid Sensory Adaptation Redux: A Circuit Perspective. *Neuron* **92**, 298–315 (2016).
2. Stefanics, G., Kremlácek, J. & Czigler, I. Visual mismatch negativity: a predictive coding view. *Front. Hum. Neurosci.* **8**, 19 (2014).
3. Musall, S., Haiss, F., Weber, B. & von der Behrens, W. Deviant Processing in the Primary Somatosensory Cortex. *Cereb. Cortex* **27**, 863–876 (2017).
4. Escera, C. & Malmierca, M. S. The auditory novelty system: an attempt to integrate human and animal research. *Psychophysiology* **51**, 111–23 (2014).
5. Zelano, C., Mohanty, A. & Gottfried, J. A. Olfactory Predictive Codes and Stimulus Templates in Piriform Cortex. *Neuron* **72**, 178–187 (2011).
6. Sams, M., Paavilainen, P., Alho, K. & Näätänen, R. Auditory frequency discrimination and event-related potentials. *Electroencephalogr. Clin. Neurophysiol.* **62**, 437–48 (1985).
7. 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).
8. 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).
9. Morlet, D. & Fischer, C. MMN and novelty P3 in coma and other altered states of consciousness: A review. *Brain Topogr.* **27**, 467–479 (2014).

10. Friston, K. A theory of cortical responses. *Philos. Trans. R. Soc. B Biol. Sci.* **360**, 815–836 (2005).
11. Friston, K. The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* **11**, 127–138 (2010).
12. Bastos, A. M. *et al.* Canonical Microcircuits for Predictive Coding. *Neuron* **76**, 695–711 (2012).
13. Carbajal, G. V. & Malmierca, M. S. The Neuronal Basis of Predictive Coding Along the Auditory Pathway: From the Subcortical Roots to Cortical Deviance Detection. *Trends Hear.* **22**, 233121651878482 (2018).
14. Parras, G. G. *et al.* Neurons along the auditory pathway exhibit a hierarchical organization of prediction error. *Nat. Commun.* **8**, 2148 (2017).
15. Giard, M. H., Perrin, F., Pernier, J. & Bouchet, P. Brain generators implicated in the processing of auditory stimulus deviance: a topographic event-related potential study. *Psychophysiology* **27**, 627–40 (1990).
16. Alho, K. Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes. *Ear Hear.* **16**, 38–51 (1995).
17. Deouell, L. Y. The Frontal Generator of the Mismatch Negativity Revisited. *J. Psychophysiol.* **21**, 188–203 (2007).
18. Fogelson, N., Shah, M., Scabini, D. & Knight, R. T. Prefrontal cortex is critical for contextual processing: evidence from brain lesions. *Brain* **132**, 3002–10 (2009).
19. Dürschmid, S. *et al.* Hierarchy of prediction errors for auditory events in human temporal and frontal cortex. *Proc. Natl. Acad. Sci.* **113**, 6755–6760 (2016).
20. Dürschmid, S. *et al.* Direct Evidence for Prediction Signals in Frontal Cortex Independent of Prediction Error. *Cereb. Cortex* (2018). doi:10.1093/cercor/bhy331
21. Li, Q. *et al.* Single-Trial EEG-fMRI Reveals the Generation Process of the Mismatch Negativity. *Front. Hum. Neurosci.* **13**, 168 (2019).
22. Nourski, K. V. *et al.* Processing of auditory novelty across the cortical hierarchy: An intracranial electrophysiology study. *Neuroimage* **183**, 412–424 (2018).
23. Phillips, H. N. *et al.* Convergent evidence for hierarchical prediction networks from human electrocorticography and magnetoencephalography. *Cortex.* **82**, 192–205 (2016).
24. Lachaux, J.-P., Axmacher, N., Mormann, F., Halgren, E. & Crone, N. E. High-frequency neural activity and human cognition: past, present and possible future of intracranial EEG research. *Prog. Neurobiol.* **98**, 279–301 (2012).
25. Chao, Z. C., Takaura, K., Wang, L., Fujii, N. & Dehaene, S. Large-Scale Cortical Networks for Hierarchical Prediction and Prediction Error in the Primate Brain. *Neuron* **100**, 1252-1266.e3 (2018).
26. Imada, A., Morris, A. & Wiest, M. C. Deviance detection by a P3-like response in rat posterior parietal cortex. *Front. Integr. Neurosci.* **6**, 127 (2013).
27. Harms, L. *et al.* Mismatch negativity (MMN) in freely-moving rats with several experimental controls. *PLoS One* **9**, e110892 (2014).
28. Jodo, E. *et al.* Neonatal exposure to an inflammatory cytokine, epidermal growth factor, results in the deficits of mismatch negativity in rats. *Sci. Rep.* **9**, 7503 (2019).
29. Camalier, C. R., Scarim, K., Mishkin, M. & Averbeck, B. B. A Comparison of Auditory Oddball Responses in Dorsolateral Prefrontal Cortex, Basolateral Amygdala, and Auditory Cortex of Macaque. *J. Cogn. Neurosci.* **31**, 1054–1064 (2019).
30. Ruhnau, P., Herrmann, B. & Schröger, E. Finding the right control: the mismatch negativity under investigation. *Clin. Neurophysiol.* **123**, 507–412 (2012).
31. Baldeweg, T. Repetition effects to sounds: evidence for predictive coding in the auditory system. *Trends Cogn. Sci.* **10**, 93–4 (2006).
32. Auksztulewicz, R. & Friston, K. Repetition suppression and its contextual determinants in predictive coding. *Cortex* **80**, 125–140 (2016).
33. Heilbron, M. & Chait, M. Great Expectations: Is there Evidence for Predictive Coding in Auditory Cortex? *Neuroscience* **389**, 54–73 (2018).
34. Hullfish, J., Sedley, W. & Vanneste, S. Prediction and perception: Insights for (and from) tinnitus. *Neurosci. Biobehav. Rev.* **102**, 1–12 (2019).
35. Schröger, E. & Wolff, C. Mismatch response of the human brain to changes in sound location. *Neuroreport* **7**, 3005–8 (1996).
36. Wiens, S., Szychowska, M., Eklund, R. & van Berlekom, E. Cascade and no-repetition rules are comparable controls for the auditory frequency mismatch negativity in oddball tasks. *Psychophysiology* **56**, e13280 (2019).

37. Horváth, J. & Winkler, I. How the human auditory system treats repetition amongst change. *Neurosci. Lett.* **368**, 157–161 (2004).
38. Mittag, M., Takegata, R. & Winkler, I. Transitional Probabilities Are Prioritized over Stimulus/Pattern Probabilities in Auditory Deviance Detection: Memory Basis for Predictive Sound Processing. *J. Neurosci.* **36**, 9572–9 (2016).
39. Takaura, K. & Fujii, N. Facilitative effect of repetitive presentation of one stimulus on cortical responses to other stimuli in macaque monkeys - a possible neural mechanism for mismatch negativity. *Eur. J. Neurosci.* **43**, 516–528 (2016).
40. Nourski, K. V *et al.* Auditory Predictive Coding across Awareness States under Anesthesia: An Intracranial Electrophysiology Study. *J. Neurosci.* **38**, 8441–8452 (2018).
41. O'Reilly, J. A. Double-epoch subtraction reveals long-latency mismatch response in urethane-anaesthetized mice. *J. Neurosci. Methods* 108375 (2019). doi:10.1016/j.jneumeth.2019.108375
42. Stefanics, G. & Czigler, I. Automatic prediction error responses to hands with unexpected laterality: an electrophysiological study. *Neuroimage* **63**, 253–61 (2012).
43. Cléry, H., Andersson, F., Fonlupt, P. & Gomot, M. Brain correlates of automatic visual change detection. *Neuroimage* **75**, 117–122 (2013).
44. Kimura, M. & Takeda, Y. Automatic prediction regarding the next state of a visual object: Electrophysiological indicators of prediction match and mismatch. *Brain Res.* **1626**, 31–44 (2015).
45. Zhang, Z. *et al.* Do theta oscillations explain the somatosensory change detection mechanism? *Biol. Psychol.* **143**, 103–112 (2019).
46. Fiser, A. *et al.* Experience-dependent spatial expectations in mouse visual cortex. *Nat. Neurosci.* **19**, 1658–1664 (2016).
47. Leinweber, M., Ward, D. R., Sobczak, J. M., Attinger, A. & Keller, G. B. A Sensorimotor Circuit in Mouse Cortex for Visual Flow Predictions. *Neuron* **95**, 1420-1432.e5 (2017).
48. Oemisch, M. *et al.* Feature-specific prediction errors and surprise across macaque fronto-striatal circuits. *Nat. Commun.* **10**, 176 (2019).
49. Sikkens, T., Bosman, C. A. & Olcese, U. The Role of Top-Down Modulation in Shaping Sensory Processing Across Brain States: Implications for Consciousness. *Front. Syst. Neurosci.* **13**, 31 (2019).
50. Phillips, H. N., Blenkmann, A., Hughes, L. E., Bekinschtein, T. A. & Rowe, J. B. Hierarchical Organization of Frontotemporal Networks for the Prediction of Stimuli across Multiple Dimensions. *J. Neurosci.* **35**, 9255–64 (2015).
51. Friston, K. Hierarchical Models in the Brain. *PLoS Comput. Biol.* **4**, e1000211 (2008).
52. Garrido, M. I. *et al.* Repetition suppression and plasticity in the human brain. *Neuroimage* **48**, 269–279 (2009).
53. 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).
54. Keller, G. B. & Mrsic-Flogel, T. D. Predictive Processing: A Canonical Cortical Computation. *Neuron* **100**, 424–435 (2018).
55. Fritz, J. B., David, S. V, Radtke-Schuller, S., Yin, P. & Shamma, S. A. Adaptive, behaviorally gated, persistent encoding of task-relevant auditory information in ferret frontal cortex. *Nat. Neurosci.* **13**, 1011–1019 (2010).
56. Funahashi, S. Functions of delay-period activity in the prefrontal cortex and mnemonic scotomas revisited. *Front. Syst. Neurosci.* **9**, 2 (2015).
57. Parr, T., Rikhye, R. V., Halassa, M. M. & Friston, K. J. Prefrontal Computation as Active Inference. *Cereb. Cortex* 1–14 (2019). doi:10.1093/cercor/bhz166
58. Wang, Y. *et al.* Heterogeneity in the pyramidal network of the medial prefrontal cortex. *Nat. Neurosci.* **9**, 534–42 (2006).
59. Zaborszky, L. *et al.* Neurons in the basal forebrain project to the cortex in a complex topographic organization that reflects corticocortical connectivity patterns: an experimental study based on retrograde tracing and 3D reconstruction. *Cereb. Cortex* **25**, 118–37 (2015).
60. Hoover, W. B. & Vertes, R. P. Anatomical analysis of afferent projections to the medial prefrontal cortex in the rat. *Brain Struct. Funct.* **212**, 149–179 (2007).
61. Garrido, M. I., Kilner, J. M., Kiebel, S. J. & Friston, K. J. Dynamic causal modeling of the response to frequency deviants. *J. Neurophysiol.* **101**, 2620–31 (2009).
62. Bastos, A. M. *et al.* A DCM study of spectral asymmetries in feedforward and feedback connections between

visual areas V1 and V4 in the monkey. *Neuroimage* **108**, 460–75 (2015).
63. Chen, I.-W., Helmchen, F. & Lutcke, H. Specific Early and Late Oddball-Evoked Responses in Excitatory and Inhibitory Neurons of Mouse Auditory Cortex. *J. Neurosci.* **35**, 12560–12573 (2015).
64. Hamm, J. P. & Yuste, R. Somatostatin Interneurons Control a Key Component of Mismatch Negativity in Mouse Visual Cortex. *CellReports* **16**, 597–604 (2016).
65. Hara, K. & Harris, R. A. The Anesthetic Mechanism of Urethane: The Effects on Neurotransmitter-Gated Ion Channels. *Anesth. Analg.* **94**, 313–318 (2002).
66. Seamans, J. K., Lapish, C. C. & Durstewitz, D. Comparing the prefrontal cortex of rats and primates: Insights from electrophysiology. *Neurotox. Res.* **14**, 249–262 (2008).
67. Trautmann, E. M. *et al.* Accurate Estimation of Neural Population Dynamics without Spike Sorting. *Neuron* **103**, 292-308.e4 (2019).
68. Paxinos, G. & Watson, C. *The rat brain in stereotaxic coordinates*. (Elsevier, 2007).
69. Lu, K. *et al.* Implicit memory for complex sounds in higher auditory cortex of the ferret. *J. Neurosci.* (2018). doi:10.1523/JNEUROSCI.2118-18.2018
70. Ruusuvirta, T., Penttonen, M. & Korhonen, T. Auditory cortical event-related potentials to pitch deviances in rats. *Neurosci. Lett.* **248**, 45–48 (1998).

Acknowledgments

We thank Drs. Conrado A. Bosman, Edward L. Bartlett, Yves Boubenec, Nell B. Cant, Lucia Melloni and Guillermo V. Carbajal for their constructive criticisms and fruitful discussions on previous versions of the manuscript. We also thank Drs. Gloria G. Parras and Javier Nieto-Diego for their assistance with neurophysiological recordings and data analyses, and Antonio Rivas Cornejo for the histological processing. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 722098 (LISTEN) and the Spanish MINECO (Grant # SAF2016-75803-P) to MSM.

Author Contributions

Conceptualization: L.C.R. and M.S.M.; Investigation: L.C.R., Formal analysis: L.C.R. and D.PG.; Software: L.C.R., D.P.G.; Visualization: L.C.R.; Writing – original draft: L.C.R.; Writing – review & editing: L.C.R., D.P.G., and M.S.M. Supervision: D.P.G. and M.S.M.; Funding Acquisition: M.S.M.

Competing interests: The authors declare no competing financial interests.

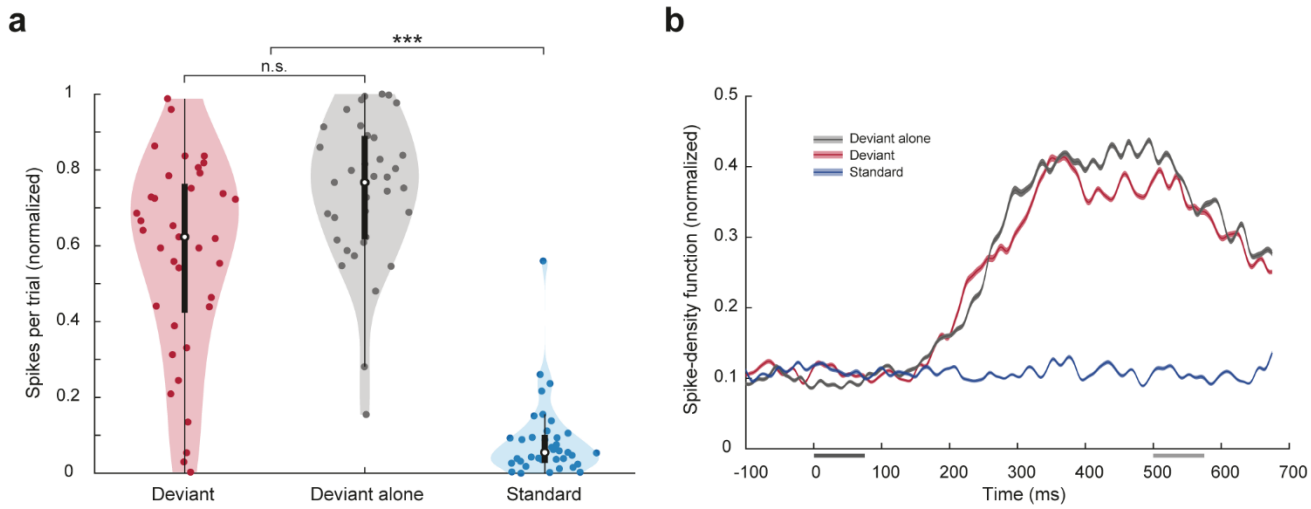
	AGm	ACC	PL	IL
# Multiunits	25	20	20	18
# tested frequencies/ estimated minimum sample size of frequencies	132/7	90/8	81/6	81/7
Raw median spike counts				
Deviant	5.707	2.816	4.301	4.244
Standard	0.525	0.561	0.412	0.318
Cascade	0.424	0.522	0.296	0.315
Many-standards	0.501	0.38	0.544	0.499
Normalized median spike counts				
Deviant	0.966	0.964	0.984	0.955
Interquartile range	0.840 – 0.997	0.805 – 0.990	0.923 – 0.996	0.796 – 0.996
Standard	0.083	0.149	0.088	0.096
Interquartile range	0.037 – 0.232	0.053 – 0.327	0.028 – 0.228	0.043 – 0.233
Cascade	0.092	0.141	0.075	0.113
Interquartile range	0.015 – 0.400	0.059 – 0.399	0.020 – 0.237	0.031 – 0.400
Many-standards	0.106	0.112	0.121	0.142
Interquartile range	0.030 – 0.239	0.028 – 0.286	0.040 – 0.289	0.029 – 0.368
Wilcoxon signed-rank test for medians of Cascade and Many-standards controls				
p-value	0.409	0.419	0.401	0.878
Raw spike count differences, Friedman test				
Deviant − Standard	5.182	2.255	3.889	3.926
p-value	**1.30E-22**	**4.15E-12**	**4.76E-16**	**2.52E-14**
Deviant − Cascade	5.283	2.294	4.005	3.928
p-value	**7.91E-22**	**2.90E-14**	**2.84E-14**	**8.40E-14**
Cascade − Standard	-0.101	-0.039	-0.116	-0.003
p-value	0.8535	0.5023	0.608	0.8751
Deviant − Many-standards	5.206	2.436	3.757	3.745
p-value	**7.04E-23**	**5.03E-15**	**4.61E-14**	**2.74E-13**
Many-standards − Standard	-0.024	-0.18	0.132	0.181
p-value	0.538	0.371	0.346	0.4795
Normalized spike count differences using Cascade controls, Friedman test				
iMM = Deviant − Standard	0.884	0.815	0.896	0.858
p-value	**1.30E-22**	**4.15E-12**	**4.76E-16**	**2.52E-14**
iPE = Deviant − Cascade	0.875	0.823	0.909	0.842
p-value	**7.91E-22**	**2.90E-14**	**2.84E-14**	**8.40E-14**
iRS = Cascade − Standard	0.009	-0.008	-0.013	0.017
p-value	0.853	0.502	0.608	0.8751
Normalized spike count differences using Many-standards controls, Friedman test				
iMM = Deviant − Standard	0.888	0.812	0.896	0.867
p-value	**1.28E-25**	**4.15E-12**	**2.15E-17**	**1.11E-15**
iPE = Deviant − Many-standards	0.869	0.854	0.861	0.824
p-value	**7.04E-23**	**5.03E-15**	**4.61E-14**	**2.74E-13**

iRS = Many-standards − Standard	0.018	-0.042	0.035	0.043
p-value	0.538	0.371	0.346	0.479

Table 1. Summary of the main neuronal population dataset for each separated field in the mPFC. Number of recorded multiunits, total number of tested sound frequencies in all multiunits together and estimated minimum sample size of tested frequencies required to find a significant index of prediction error (see Methods). Median indices of neuronal mismatch (iMM = deviant – standard), repetition suppression (iRS = control – standard) and prediction error (iPE = deviant – control) can be computed with cascade or many-standards stimuli as control. Significance is set at $p < 0.05$ and significant *p*-values are highlighted.

	mPFC	
# Multiunits	9	
# tested frequencies	37	
Raw median spike counts		
Deviant	5.404	
Standard	0.469	
Deviant-alone	6.966	
Normalized median spike counts		
Deviant	0.623	
Interquartile range	0.426 - 0.760	
Standard	0.055	
Interquartile range	0.030 - 0.097	
Deviant-alone	0.767	
Interquartile range	0.621 - 0.886	
Spike count differences; Friedman test	Raw	Normalized
Deviant − Standard	4.934	0.568
p-value	**1.68E-07**	**1.68E-07**
Deviant – Deviant-alone	-1.562	0.144
p-value	0.163	0.163
Deviant-alone − Standard	6.496	0.712
p-value	**3.45E-11**	**3.45E-11**

Supplementary Table 1. Summary of the neuronal population data subset with the deviant-alone paradigm in the mPFC. Number of recorded multiunits and the total number of tested sound frequencies in all multiunits together. Significance is set at $p < 0.05$ and significant *p*-values are highlighted.



Supplementary Figure 1. Strong responses to the deviant-alone paradigm across the entire mPFC. **a** Density distribution of normalized responses per stimulus repetition of each frequency tested in each condition. Colored dots represent sample points of each measured firing rate (n = 9 multiunits & 37 frequencies). Box plots are shown in black, with white dots denoting the population median. Statistical significance denoted as *n.s.* (non-significant) and *** ($p < 0.001$) with a Friedman test (see Table 2). **b** Population average firing rate as normalized spike-density functions for the different conditions (mean ± SEM). Horizontal gray bars indicate the epoch of sound stimulation; conditions are assigned according to the tone in dark gray. Note that deviant-alone responses are largely equivalent to regular deviant responses.