



Research

Cite this article: Rogalla MM, Rauser I, Schulze K, Osterhagen L, Hildebrandt KJ. 2020 Mice tune out not in: violation of prediction drives auditory saliency. *Proc. R. Soc. B* **287**: 20192001.
<http://dx.doi.org/10.1098/rspb.2019.2001>

Received: 28 August 2019
 Accepted: 7 January 2020

Subject Category:

Behaviour

Subject Areas:

behaviour, neuroscience, cognition

Keywords:

animal behaviour, expectation, auditory, statistical learning

Authors for correspondence:

Meike M. Rogalla
 e-mail: meike.rogalla@uol.de
 K. Jannis Hildebrandt
 e-mail: jannis.hildebrandt@uol.de

[†]Submitting author.

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.4819476>.

Mice tune out not in: violation of prediction drives auditory saliency

Meike M. Rogalla^{1,2,†}, Inga Rauser¹, Karsten Schulze¹, Lasse Osterhagen^{1,2} and K. Jannis Hildebrandt^{1,2}

¹Department of Neuroscience, Division of Auditory Neuroscience, and ²Cluster of Excellence, Hearing4all, Carl von Ossietzky University, Oldenburg, Lower Saxony 26129, Germany

MMR, 0000-0003-3103-0053; KJH, 0000-0003-4455-7495

Successful navigation in complex acoustic scenes requires focusing on relevant sounds while ignoring irrelevant distractors. It has been argued that the ability to track stimulus statistics and generate predictions supports the choice of what to attend and what to ignore. However, the role of these predictions about future auditory events in drafting decisions remains elusive. While most psychophysical studies in humans indicate that expected stimuli are more easily detected, most work studying physiological auditory processing in animals highlights the detection of unexpected, surprising stimuli. Here, we tested whether in the mouse, high target probability results in enhanced detectability or whether detection is biased towards low-probability deviants using an auditory detection task. We implemented a probabilistic choice model to investigate whether a possible dependence on stimulus statistics arises from short-term serial correlations or from integration over longer periods. Our results demonstrate that target detectability in mice decreases with increasing probability, contrary to humans. We suggest that mice indeed track probability over a timescale of at least several minutes but do not use this information in the same way as humans do: instead of maximizing reward by focusing on high-probability targets, the saliency of a target is determined by surprise.

1. Introduction

Animals using acoustical information for navigation are nearly continuously confronted with numerous sounds from different sources. Relevant stimuli need to be detected while irrelevant ones should be ignored. For this process of differentiation, the ability of tracking stimulus statistics (if a stimulus occurs with high or low probability) is essential, sets expectations and creates predictions about future auditory events [1,2]. While there is general agreement on the principal importance of expectation in auditory perception, there are different ways in which these predictions may be used to guide decisions.

On the one hand, high-probability, expected and relevant signals may selectively improve perception and form an important part of the analysis of complex auditory scenes [3–5]. Adult humans are better at detecting expected auditory stimuli in reward-based auditory listening tasks [6]. Humans internally monitor the probability of a stimulus and adapt their behaviour according to the stimulus statistics [7–9]. This behaviour does not require awareness of the subject and is driven by unconscious expectations [10]. Within this framework, the improvement of detectability is based on the expectation as an implicit cue and biases detection performance towards expected targets [11].

While most psychophysical studies indicate that expected stimuli serve as implicit cues for selective listening, most work studying the physiology of auditory processing highlights the detection of unexpected, surprising stimuli. Stimuli are more salient when presented rarely to the auditory system due to pre-attentive mechanisms [1,12]. Within this framework, the evaluation of stimulus statistics serves to detect novelty, emphasizing changes in the auditory scene rather than enabling tracking of task relevant information.

Novelty coding in the auditory system has successfully been interpreted within the framework of predictive coding [13]. Stimuli are carried on to higher sensory areas as prediction error signals, updating predictions for the interpretation of sensory input. Whether and how prediction error signals shape the *detectability* of a signal is less clear: on the one hand, larger prediction error signals could result in higher detectability of surprising signals [14]. On the other hand, top-down processes are thought to assign higher weights to expected signals and thereby render them more salient perceptually [15,16].

Thus, tracking of stimulus probability may influence auditory processing in two contrary ways: on the physiological level, low-probability sounds elicit maximal responses, but during listening tasks, *relevant* high-probability sounds shape expectations, and thereby improve their detectability. While physiological evidence for deviant detection spans all the way from animal models to humans [13,17], behavioural assessment of the effects of target probability is largely restricted to humans.

Although rodents serve as widely used animal models to study auditory phenomena, little is known about their ability to monitor stimulus probability and how it shapes detectability. Here, we asked how target probability influences auditory perception in mice. More specifically, we tested whether target probability is used as an implicit cue for expectation-biased detection or whether detection is biased towards low-probability deviants. To this end, we devised three different psychophysical tasks and tested three separate sets of mice. First, we used faint tones in noise and varied the probability of a given tone frequency between different sessions. This paradigm resembles the ‘listening band phenomenon’, the most prominent example of probability-guided detection bias in the human literature [6]. Subsequently, we tested whether the probability-dependence generalizes to other detection tasks, namely streaming paradigms, in which a target has to be detected in one out of multiple streams. Here, we separately tested for effects on the detection of both spectral and temporal stimulus dimensions. Finally, we present a probabilistic choice model to investigate whether the dependence on stimulus statistics arises from short-term serial correlations or from integration over longer periods. A previous version of this work has been placed on a preprint server [18].

2. Results

(a) Experiment 1: tone-in-noise detection

When humans are asked to detect faint tones in a noise background, performance for high-probability targets is better than for those played with low probability, even if listeners are not consciously aware of the probabilities [9]. This is usually explained by a focusing on specific auditory filters, thereby listening selectively to a certain frequency range [7]. In our first experiment, we aimed to test whether mice are able to track target probabilities from session to session and display a preference for either high- or low-probability targets. We devised a behavioural paradigm (figure 1), in which mice were trained to indicate the detection of faint tones embedded in a noise background by leaving a small pedestal after the presentation of a target (figure 1*a*). A typical single session contained 60 targets and lasted approximately 30 min. In order to test the animals near their individual thresholds, we first tested a single-frequency in each session, varying the level of the tones to determine the

threshold (figure 1*c*, upper panel). In the next step, we presented tones with varying probability as targets in mixed sessions (lower panel in figure 1*c*). We hypothesized that if mice displayed selective listening to high-probability tones they should (1) be better at tone detection in the single-frequency sessions compared to the mixed sessions and (2) show better performance for the high-probability compared with the low-probability stimulus within the mixed sessions.

Contrary to our hypothesis, all animals tested showed higher sensitivity in the mixed than in the single-frequency session tested before (example data in figure 2*a*; repeated measures ANOVA, $F_{1,20} = 32.2$, $p < 0.001$). Within the mixed sessions, the impact of stimulus probability on the preference of the mice for low-probability tones was confirmed. Sensitivity was positively influenced by surprise, quantified as the prediction error (inverse conditional probability of the stimulus, [19]). This relation was highly significant, both when taking the single-frequency sessions into account and for mixed sessions only, and independent of the frequency that was played. We concluded that tone detection performance of mice is clearly biased by stimulus probability, and that low-probability sounds were detected more reliably.

(b) Experiment 2: frequency change detection in streams

Contrary to the behaviour displayed in experiment 1, a strategy to focus on high-probability sounds would have maximized rewards. A possible explanation for mice not taking advantage of tracking probabilities is that they are not able to focus on a single-frequency band in a continuous noise background with very sparse tones appearing at random times. We reasoned that a more natural situation could be the presence of multiple streams of tones that allow to selectively track one of these streams [20,21]. We therefore designed an experiment in which animals had to detect a frequency change in either one of two continuous streams of tone pips (figure 3*a*). The repetition rate was rapid (5 Hz for either stream) and the tone streams were more than an octave apart in frequency, a parameter range that results in a clear two-stream percept in most animals, including rodents [22,23]. To this end, we trained a batch of naive mice ($n = 6$). Again, we varied the probability that a target could appear in either of the two streams. In one set of sessions, frequency changes would be inserted in either one of the two streams only. In a second set, targets appeared in both streams with equal probability. Sessions were randomized in order to avoid sequence effects.

When we compared the mean sensitivity for the two different probability levels, we observed a higher mean sensitivity for the mixed sessions for all tested frequency changes (figure 3*c*). As in Experiment 1 (figure 2), targets were more salient to the mice if they were distributed between the two streams than if they were played in one of the two streams only. This was confirmed when we compared all animals for both streams (figure 3*e*; *rmANOVA*, $F = 6.0$, $p = 0.0171$). Experiment 2 confirmed that detection of targets in mice is strongly biased by target probabilities from session to session, but detectability seems to be determined by surprise rather than be biased towards expected targets, despite the latter being the better strategy to maximize rewards in a given session.

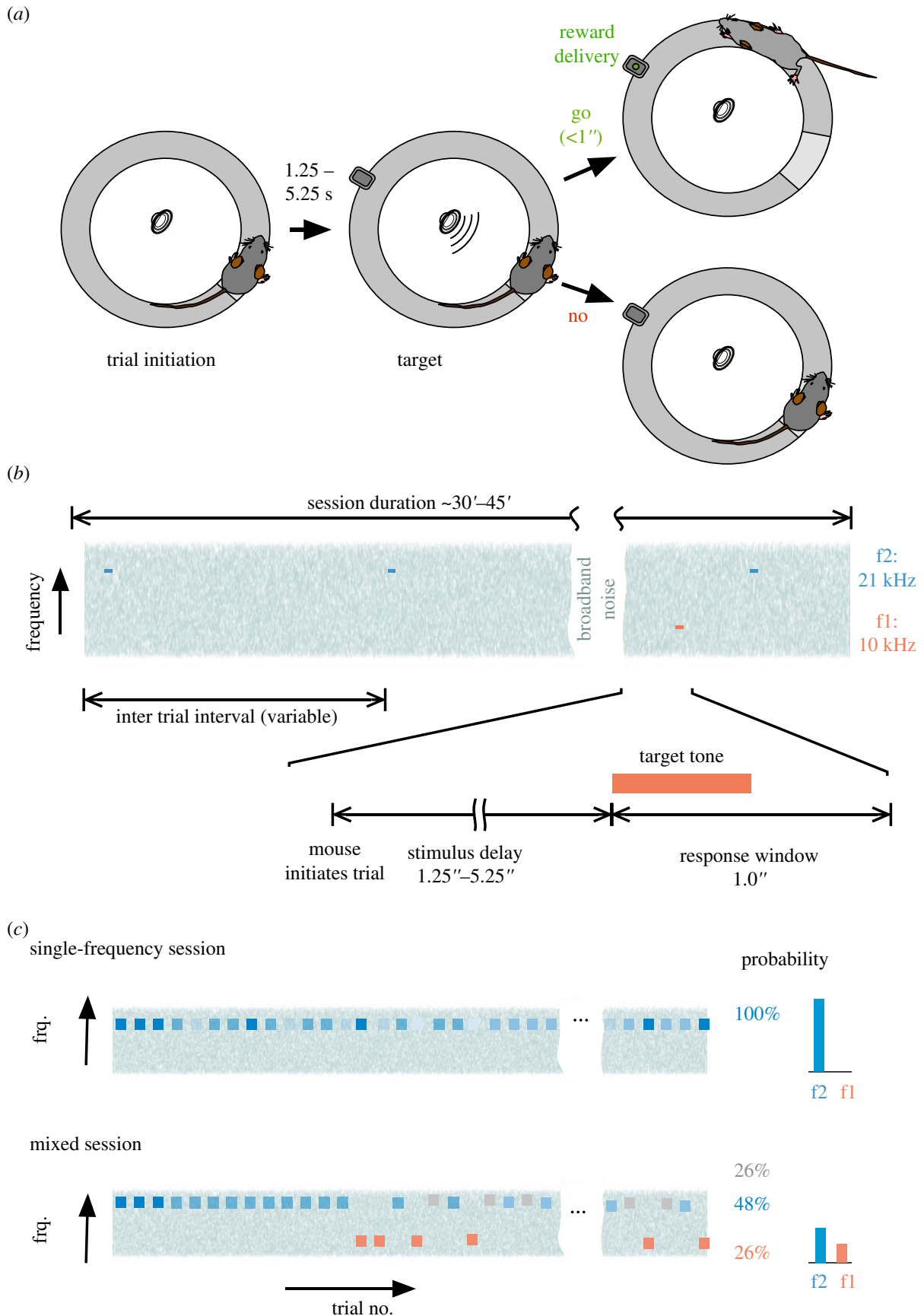


Figure 1. Behavioural paradigm and stimulus protocol used in Experiment 1. (a) Go/No-Go paradigm used in this study. Mice initiated a trial by climbing on a small pedestal on the circular platform. After a variable waiting interval, a target was presented. Animals received a reward if they left the platform within 1 s after target presentation. The next trial could be initiated immediately. (b) Timeline of one experimental session. Throughout the entire session, a broadband noise stimulus was presented. Once a trial was initiated, a 500 ms pure tone was presented after a random stimulus delay. In a single session, an animal had to complete 73 or 78 trials, which typically lasted 30–45 min. (c) Different probabilities of single-frequency pure tone targets in different sessions. In single-frequency sessions, the level of the tones was varied, but only pure tones of either frequency f_1 (10 kHz) or f_2 (21 kHz) were presented. In mixed session, level was held constant near the behavioural threshold, but three different frequencies were presented. In any one session, either f_1 or f_2 was presented with 48% probability and the respective other with only 26%. In addition, a tone of the frequency close to the high-probability targets was presented in 26% of the trials. (Online version in colour.)

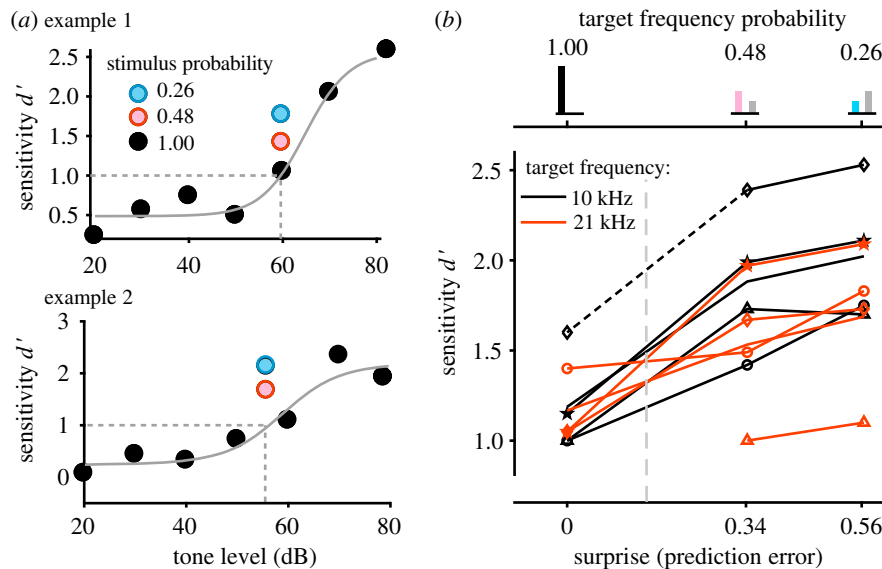


Figure 2. Results for Experiment 1—tone-in-noise detection. (a) Example performance of two different animals for the tone-in-noise stimuli at a single tone frequency, presented with different probabilities. Before the mixed-frequency experiments, animals were tested individually for their thresholds at each tone frequency by presenting tones of a single-frequency (probability 100%) at different levels to construct psychometric functions (black circles, grey line). In the mixed experiments, tones with a level corresponding to a d' of 1 (dashed line) were presented with probabilities of 48% (red circle) or 26% (blue circle). (b) Population data for all four animals at the two different frequencies used (red, 21 kHz; black, 10 kHz). The values for a probability of 100% were taken from the psychometric function obtained after the mixed experiments. Histograms above the graph visualize the probability of the tone in the respective sessions, the x-axis shows the surprise quantified as the prediction error. Note that larger numbers indicate more surprising stimuli. Total number of sessions included: 208 (152 for the mixed session, 56 for the psychometric functions). (Online version in colour.)

(c) Experiment 3: gap detection in streams

Since the two streams were separated by frequency and target changes were along the same dimension, we aimed to test whether our results would generalize to other stimulus dimensions. Therefore, we trained a naive set of animals ($n=7$) to detect temporal irregularities in the form of short gaps, introduced into one of the two streams (figure 3*b*). Here, we used three probabilities for each condition: targets in only one of the two streams (100%), or 66.7% and 33.3% probability in sessions with targets in both streams. As already observed for the frequency changes, sensitivity for detection of gaps strongly depended on target probability, with the best detectability for low-probability targets in the mixed sessions, and lowest detection performance for targets in only one out of two streams (figure 3*d*). We observed this effect for both possible target streams in all animals (figure 3*f*, rmANOVA, $F_{1,206} = 30.0$, $p < 0.001$). Experiment 3 confirmed our results from the previous experiments and generalizes the saliency of surprising targets to temporal features as well.

(d) Probabilistic choice model

We observed higher detection performance for low-probability stimuli in three different behavioural experiments. However, this does not necessarily mean that the animals were tracking long-term probability. When manipulating probability, the structure of the randomized trial sequences is changed as well: in sessions in which one type of target is presented with low probability, stimuli are more often preceded by a different target than if presented in high-probability sessions. A simple switch of the stream being monitored after each trial could explain our results just as well as tracking probability over a time course of up to tens of minutes. In order to test whether the animals were tracking probability over longer timescales or simply displaying short-term trial-history effects, we devised a probabilistic choice model (figure 4*a*). The

model included the factors stimulus intensity, stimulus probability within the session, and recent history of stimuli presented in the immediately preceding trials. The model was fit separately for each mouse and experiment, in versions including or excluding probability and history terms. If the probability-dependence was due to recent history effects, a model including only the respective term should perform equally well as one including both probability and history, and better than one that takes only probability into account. Inclusion of the probability term significantly improved model performance (figure 4*b*). By contrast, inclusion of the recent history term (up to four preceding trials) improved the model only marginally (figure 4*b*).

The average interval between two trials was 30.2 ± 10.3 s (mean \pm s.d., $n=528$ sessions from all three experiments). Since there was little effect of recent trial history up to at least four trials, perception in the mice was apparently shaped by long-term probability on the timescale of several minutes at least. In line with this, we could not find a difference between hit rates after a switch of the stimulus class between two trials or a repetition of stimuli from the same class (figure 4*d*). We also did not find a change in overall strategy between mixed and pure sessions—false alarm rates did not differ between those session (figure 4*c*).

3. Discussion

Stimulus statistics in auditory scenes have been suggested to shape auditory perception in two contrary ways: (1) a focus on novelty detection, favouring low-probability sounds [17] and (2) improving detectability of expected, high-probability sounds, maximizing overall detection rate [6,10]. Here, we tested whether detectability in mice is biased rather towards low- or high-probability target sounds. To this end, we conducted three different experiments varying the probability

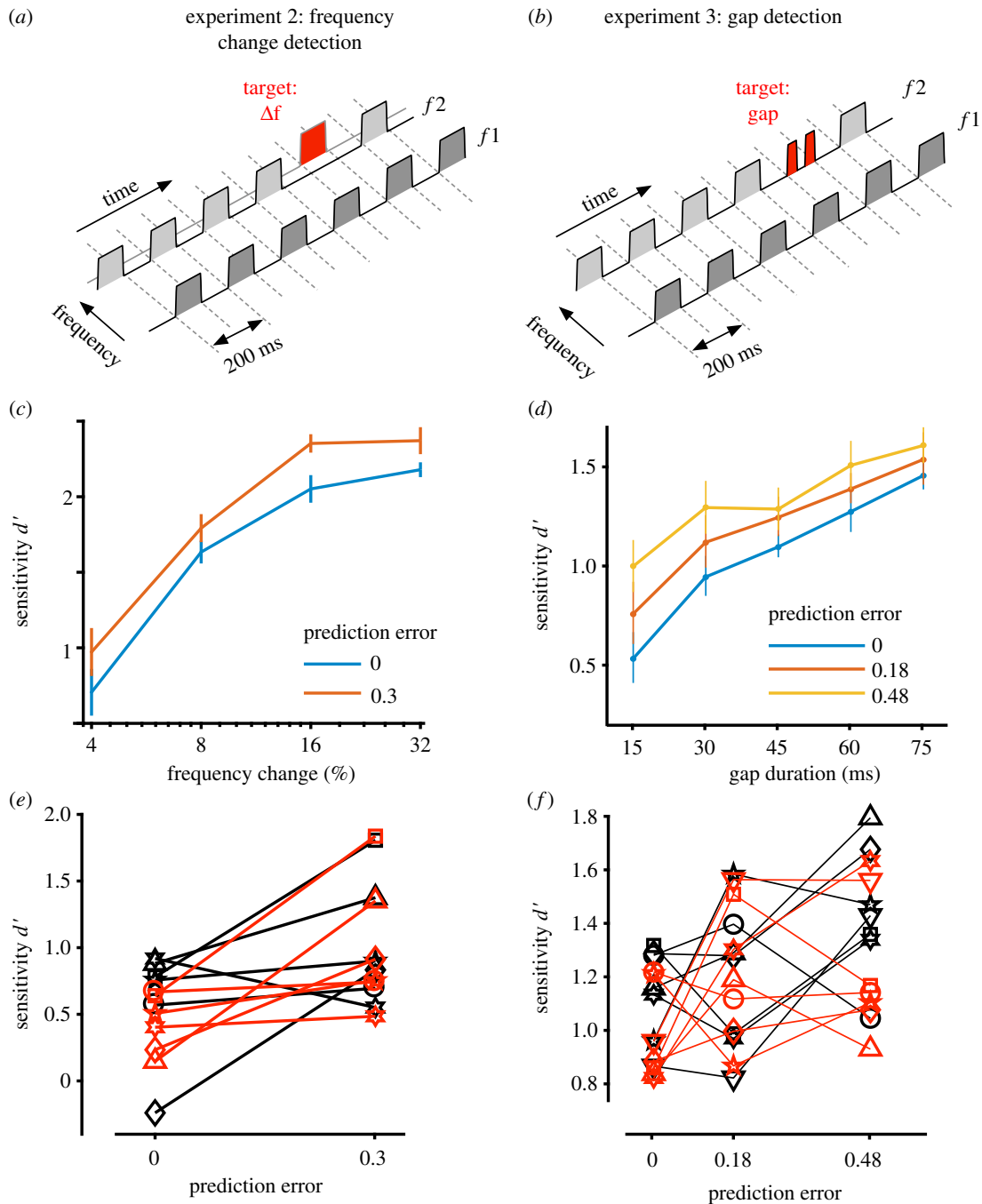


Figure 3. Stimulus paradigms and results for Experiments 2 and 3. (a) Paradigm for Experiment 2: two continuous, interleaved streams of tone pips with different frequencies were presented. $f_1 = 10$ kHz, $f_2 = 21$ kHz. Animals had to detect a change in the frequency in either of the two streams. (b) Paradigm for Experiment 3: a short gap was inserted into one of the two narrowband noise streams (centre frequency same as f_1 and f_2 in (a)) as a target for detection. (c) Mean performance of all animals ($n = 6$) in Experiment 2 tested at different values of frequency change at either 50% (red) or 100% (blue) probability. Error bars show standard error of the mean (s.e.m.). (d) Mean performance of all animals ($n = 7$) in Experiment 3 tested at different gap durations at 33.3% (yellow), 66.7% (red) or 100% (blue) probability. Error bars depict s.e.m. (e) Sensitivity as a function of prediction error for Experiment 2—tone change detection—for all animals tested in both frequency streams (black, 10 kHz; red, 21 kHz). Each line joins data from an individual mouse for targets with a frequency change of 16%. (f) Sensitivity as a function of prediction error for Experiment 3—gap detection—for all animals, mean across all gap durations in either of the two frequency streams (black, 10 kHz; red, 21 kHz). (Online version in colour.)

of targets. While humans direct their attention to the most probable target out of several acoustic channels or streams, target detectability in mice decreased with increasing probability. Thus, the more surprising a stimulus was, the more reliably it was detected. This was confirmed in three independent experiments using three separate sets of animals, one with changing probability of target frequency in noise (figure 2) and two using a streaming paradigm (figure 3) with either a spectral or temporal variation to

be detected. Finally, our probabilistic choice model best predicted animal behaviour for all three tasks if it took overall probability into account, but not if we considered recent trial history (figure 4b). These results suggest that mice indeed track probability over a timescale of at least several minutes, but do not use this information in the same way as humans: instead of maximizing reward by focusing on high-probability targets, the saliency of a target is determined by surprise.

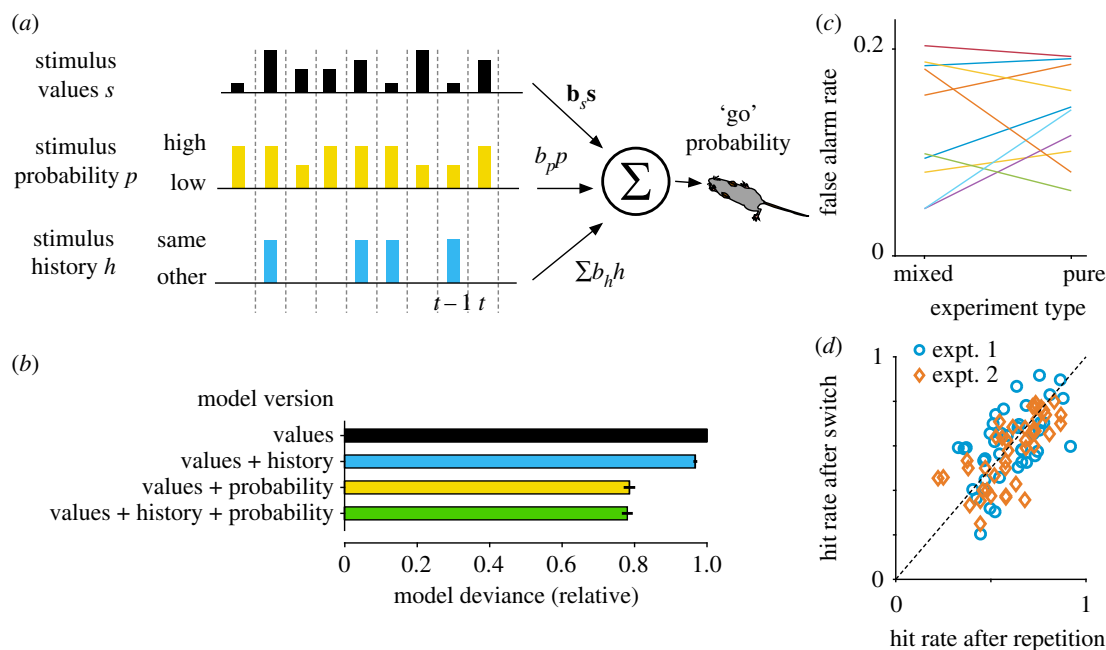


Figure 4. Probabilistic choice model. (a) Schematic illustration of the probabilistic choice model. The full model includes stimulus intensity, stimulus probability within the session for each stimulus, and recent history of stimuli presented in the immediately preceding trial steps $t - i$. Stimulus values were different in each experiment: SNR ratio in Experiment 1, frequency shift in Experiment 2 and gap duration in Experiment 3. The model was fit for each mouse and experiment, in four different versions, including either all three factors, stimulus values only, values + history or values + probability. (b) Performance of the four model versions, plotted as deviance of model output to the data, relative to the model including stimulus values only. Note that smaller numbers mean better model performance. Bars represent mean deviations from all animals in the three experiments \pm s.e.m. (c) False alarm rate depending on whether stimuli of one class were presented as the only stimuli in the session (pure) or whether they were combined with other stimuli (mixed). Each line represents mean false alarm rates from a single animal. (d) Influence of immediate trial history on hit rate: x -axis: hit rate when the stimulus in the trial before was drawn from the same class as the current stimulus (repetition), y -axis: hit rate to stimuli that were preceded by a stimulus from another class (switch). (Online version in colour.)

(a) Different strategy or mouse-specific auditory processing?

It seems that mice are very good at something that humans find hard and vice versa. Are our results in mice really caused by a different strategy with respect to target probability or can it be explained by more basic differences in their auditory system? Mice have much wider auditory filters [24], so our stimuli could have been merged into one perceptual category, such that no separate streams would have built up. However, in all three experiments, we used targets that were more than an octave apart, far above the frequency discrimination threshold of mice [25]. It was not assessed whether the two sequences used in experiments 2 and 3 resulted in a streaming percept, as this was not the focus of our study and our results do not critically depend on the sequences being perceived as streams. However, all animal species that have been tested so far showed evidence for streaming for stimuli separated by one octave and upwards [22], including rats [23], which have similar auditory filters to mice. Furthermore, our positive results on the effect of target probability and its generalization across paradigms provide evidence for perceptual separation rather than merging.

(b) Deviance detection in the auditory system

We found clear evidence for mice to favour unexpected, surprising stimuli. An explanation for this behaviour is that in mice, the deviance signal (prediction error) is not weighted by probability-cued top-down processes as it is in humans. There is a large body of work on enhanced neural representation

of deviant stimuli in the auditory system in both animal models and humans. At the single-cell level, stimulus-specific adaptation (SSA) describes the enhancement of the neural representation of low-probability sounds [1,17]. A typical paradigm is the presentation of a sequence of tones of two different frequencies, with varying relative probability [26]. Experiment 1 of this study is such a paradigm, but the ratio of stimulus duration (0.5 s) and the very long inter-stimulus interval (mean of all sessions: 30.2 s) has not been reported before. However, timescales up to several minutes are reasonable based on the measurement of adaptation time constants in the cat auditory cortex [27]. The stimuli in our Experiments 2 and 3 extended this to two synchronously presented sequences of standard and deviants—Experiment 2 using frequency shifts and Experiment 3 using temporal deviants. SSA is likely to shape responses to the deviant targets in both of the streams in either experiment. Deviants in both streams are very rare: approximately 30 s inter-trial intervals with a background pulse repetition rate of 5 Hz result in a deviant probability of approximately 1% in session with the target in one stream and approximately 0.5% in sessions with equal distribution between the two streams. Neural sensitivity for such small changes has not been reported yet, but there is no principle reason why they cannot exist. SSA has been shown to extend beyond simple pure tone patterns [28] and to more complex statistical structure of the sensory context [29]. Similar to the findings presented here (figure 4), SSA is sensitive to average statistics rather than recent history [30].

It is very difficult to directly observe a correlate of neural deviance representation at the behavioural level. This is probably due to its pre-attentive nature—deviance detection is

observable in passively listening subjects [31] as well as in anaesthetized animals [12]. In active listening tasks, implicit cueing could then use the predictive signal to improve the representation of high-probability sounds through top-down control of sensory processing [10] by inverting the sign [16]. However, in mice we may be able to directly observe the strength of a prediction error signal at the perceptual level.

(c) Sensory ecology

Deviance detection is of foremost importance for the detection of sudden, potentially dangerous changes in the environment. In mice, as potential targets of predation, deviance detection may have priority over probability-cued selective attention. On the contrary, both humans and carnivores [21] may be cued implicitly using target probability—providing evidence that the effect of stimulus probability on perception may be due to sensory ecology rather than taxonomy. Both primates and carnivores might use their auditory senses to tune in and follow potential prey or conspecific communication signals. Interestingly and consistent with the strategy reported here for mice, threatening stimuli in humans are best perceived if they occur with relatively low-probability [32]. By contrast, implicit cueing in reward-based tasks usually enhances perception of high-probability signals [6,10,11].

An alternative explanation for our results would be that mice are not able to provide top-down influence and scale the deviance signals accordingly. However, recent work suggests that mice are able to selectively attend at least *explicitly* cued visual patterns [33] or auditory streams [34]. This could indicate that mice do not lack a mechanism for top-down control of input-signal scaling, but it is not activated by probability-cueing. Instead, if mice use contextual auditory information mainly for the detection of threats, this rule may be hard-wired and not under the control of top-down signals.

Our results suggest that in mice, predictions based on complex statistic regularities are computed along the sensory pathway, but mostly used to suppress ongoing input, similar to sensory adaptation on shorter timescales [1,12]. The development of top-down modulation of error signals in carnivores and primates may have been added later on to this first step of probabilistic analysis of complex sensory scenes.

(d) Perspective

In summary, our study provides evidence for animal detection behaviour being shaped directly by prediction error. This finding could be very helpful for future work on prediction-guided behaviour, since we may be able to study the neural mechanisms underlying extraction of complex contextual sensory information without the confounding of the interplay with top-down modulation shaped by the task. The mouse model offers unrivalled possibilities to record and manipulate neural activity in the behaving animal. In future studies, this may not only enable measurements of neural deviance detection during relevant behaviour. It also offers the perspective of direct manipulation of potential mechanisms, with the observed behaviour as readout to infer causal relationships.

4. Material and methods

(a) Experimental model and subject details

In total, 17 adult male mice bred at the University of Oldenburg animal facilities were used in the experiments (Experiment 1, $n = 4$; Experiment 2, $n = 6$; Experiment 3, $n = 7$). All mice had a C57BL/6.CAST-*Cdh23*^{Ahl+} background (the Jackson laboratory, #002756) and were between three and nine months old. We used this line because it does not display the age-dependent hearing loss which is present in other C57BL/6 lines [35,36]. Animals were kept at a reversed 12 L : 12 D cycle, all experiments were performed during the dark period. Animals had unlimited access to water but were food-deprived to a moderate extent (85–90% of their ad libitum weight) and single-housed in standardized enriched cages but with visual and olfactory contact to neighbouring animals.

(i) Behavioural paradigm

All three experiments were performed using the following reward-based go/no-go paradigm. Animals were placed on an annular platform made from wire mesh (figure 1a). The raised platform was placed in a custom sound-proof chamber that was lined with pyramid foam. On one side of the platform, a small pedestal was installed. Once the animals ascended the pedestal, a random, variable waiting time started, drawn from a distribution between 1.25 and 5.25 s (in steps of 1 s). After this pseudo-random interval, a target was presented. The onset of the target triggered a 1 s response window. If the animals descended within the window ('go'), a food pellet (0.02 g, Dustless precision pellets rodent, grain based, Bio-Serv, #F0163) was delivered at the opposite side of the annular platform. If the animals stayed on the pedestal, a new trial was presented after a newly drawn waiting time. A typical session contained 60 targeted trials and 25 sham trials and lasted 30–40 min. For all different target stimuli, the distribution of waiting times was the same, both within and across sessions. Animals were tested once or twice per day with at least 2 h between subsequent sessions. All experiments were controlled by custom software (GitHub repository: <https://github.com/Spunc/PsychDetect>) written in MATLAB (The MathWorks, RRID: SCR_001622). Pellet dispenser and light barriers were custom build (University of Oldenburg workshop) and controlled by a microcontroller (Arduino UNO, Arduino AG, Italy) connected to a Windows PC.

(ii) Estimation of false alarm rates and sensitivity

For the experiments, each trial was drawn from five, non-overlapping response windows. Thus, randomly choosing one of the five windows to descend from the platform in a given trial would theoretically have resulted in 20% chance level. However, since the paradigm is not strictly forced choice, we needed to estimate the real chance level for each session in order to quantify detectability in terms of the sensitivity d' (see section Data analysis and statistics). Here d' is independent of the individual decision boundary and thus more comparable across sessions and animals than uncorrected hit rates [37]. In order to estimate false alarm rates for each session, we introduced sham trials with a designated 'stimulus' delay, but without presentation of a target at that delay. If the animals descended the platform within the response window of a sham trial, a false alarm was scored—if not, the response was counted as a correct rejection (electronic supplementary material, figure s1D). Descending the platform at any other time without a target was not counted. Depending on the paradigm 25–33% of trials in a given session were sham trials, with the same distribution of waiting times as in the target trials.

(iii) Stimuli

For sound presentation, a speaker (Vifa XT 300/K4, Denmark) was mounted in the sound-proof chamber approximately 0.5 m above the pedestal. Sound was generated using a high-fidelity sound card (Fireface UC, RME, Germany) connected to the PC. Sound was played back at either 192 kHz (experiment 1) or 96 kHz (Experiments 2 and 3) sampling rate. The speaker was calibrated at the approximate position of the head of the animals using a measurement microphone (model 40BF, G.R.A.S., Denmark).

(iv) Paradigms

Experiment 1: tone-in-noise detection. Tones in noise served as a target in Experiment 1. Once a session started, broadband noise (4–64 kHz, 60 dB) was constantly played until the end of the session. Pure tone of either 10 or 21 kHz served as targets (2 ms cosine ramps, 500 ms duration). In the sessions containing only one target frequency, the level for that frequency was varied between 20 and 80 dB in steps of 10 dB in order to obtain a psychometric function. Psychometric functions were fit with a logistic function and an individual signal-to-noise ratio (SNR) threshold was estimated. In the mixed sessions, we used the level corresponding to the individual SNR thresholds, estimated as the point on the psychometric curve with a d' value of 1. During the mixed session, targets were played back at three different frequencies: (1) high-probability (0.48), 'priming' frequencies, which were present in the first 10 trials and also throughout the rest of the session, (2) the low-probability (0.26) target frequency, only played back from trial no. 11 onwards, and (3) target tones of a third frequency close to the priming frequency, played with low probability (0.26). These last stimuli were not used for further analysis, since they had not been used in corresponding sessions with high-probability. Only trials 11 onward (after the priming) were used for analysis. Each animal performed at least 10 sessions for both priming frequencies. Measurement of psychometric function was repeated after the mixed sessions in order to rule out effects of perceptual learning when comparing single-frequency with mixed sessions. Prior to the described experiments, animals have been trained in a tone-in-noise detection task with either one of the two target frequencies (10 or 21 kHz) which were randomly chosen for each session until the performance reached a stable level in several consecutive sessions (electronic supplementary material, figure s1E).

Experiment 2: frequency change detection in streams. For Experiment 2, two alternating tones with frequencies of 10 and 21 kHz (1.07 octaves) were played at rate of 5 tones/s throughout the experimental session. Tone duration was 100 ms including 2 ms cosine ramps. The level of each individual tone was roved between 60 and 66 dB SPL (randomly) in order to avoid the detection of a differences in loudness when the shift in frequency occurred. The frequency of a tone from either tone sequence was shifted upwards by 4%, 8%, 16% or 32%. Mice had to report the appearance of the frequency shift within 700 ms after onset of the shifted tone. Within a session, targets appeared either in only one of the two tone sequences ('single', probability 1 and 0, respectively) or in either of the two sequences ('mixed', probability 0.5 for each sequence). Each animal completed at least eight sessions for each of the mixed session types. Prior to experiments, animals have been trained in a frequency change detection task in a single stream, frequencies (10 or 21 kHz) were chosen randomly for each training session. After receiving a stable and similar performance in both streams in several consecutive sessions, experiments with the two alternating streams were introduced. Mice responded towards the simultaneous presentation immediately without a decline in performance.

Experiment 3: gap detection in streams. The temporal structure of the sequences in Experiment 3 was the same as in Experiment 2, but instead of pure tones, narrowband noise with a bandwidth

of 0.25 octaves around 10 or 21 kHz was used. The level of narrowband pulses was fixed at 60 dB SPL. In the target pulses, gaps with duration of 15, 30, 45, 60 and 75 ms were introduced (including 2 ms cosine pulses). The response window was 1 s. For Experiment 3, we used three different probabilities: 1 (target only in one sequence), 0.66 or 0.33. Each animal completed at least eight sessions for each session type. For this specific paradigm, the auditory training was performed in the same way as Experiment 2 but was started with a broadband instead of a narrowband noise, centre frequencies of 10 and 21 kHz were randomly chosen. After animals showed a stable performance for both individual streams, the bandwidth of the noise was slowly reduced in each session until the final bandwidth of 0.25 octaves was reached. Subsequently, final experiments with both streams were conducted.

(b) Quantification and statistical analysis

(i) Data analysis and statistics

In all three experiments, for each session i and stimulus class s , the sensitivity d' was calculated as

$$d'_{i,s} = z(H_{i,s}) - z(FA_i),$$

where $z()$ is the inverse of normal cumulative function, $H_{i,s}$ is the hit rate for the stimuli with parameters s in the i th session $P(\text{response} | \text{stimulus } s)$ and FA_i is the false alarm rate $P(\text{response} | \text{sham})$.

In order to check for significant effects of stimulus probability on the sensitivity, we fit a generalized mixed effects model (MATLAB *fitglm*), with the d' values as response variable and probability and stimulus parameters as factors. In Experiment 1, the stimulus parameter factor was target tone frequency. For Experiment 2, relative frequency shifts were entered as factor. For Experiment 3, the stimulus factor was gap duration. For each experiment, we performed repeated measures ANOVA (rmANOVA, MATLAB) and report both F -values and exact p -values up to the fourth decimal.

(ii) Probabilistic choice model

To account for different factors affecting animal choice behaviour we devised a probabilistic choice model, similar to what has been used before in order to include history in psychophysics [38].

The probability p_{go} to jump at a given trial t in a behavioural session is given by

$$p_{go}(t) = \frac{1}{1 + e^{L(t)}},$$

with the response variable $L(t)$, that is a weighted sum of three main terms: (1) the stimulus parameters $\mathbf{s}(t)$, (2) the overall probability of the stimulus to appear in the given channel $p(t)$, and the stimulus history $\mathbf{h}(t)$:

$$L(t) = \mathbf{b}_s \cdot \mathbf{s}(t) + b_p p(t) + \mathbf{b}_h \cdot \mathbf{h}(t).$$

The stimulus parameters depend on the paradigm. For Experiment 1, this was absolute stimulus frequency and the SNR ratio. For Experiment 2, this was the absolute frequency of the stream the target appeared in and the frequency shift of the target. For Experiment 3, we entered absolute frequency of the target stream and the gap duration in the target pulse.

The probability term is constant across a given session and only depends on the target channel. The history term is described by

$$\mathbf{b}_h \cdot \mathbf{h}(t) = \sum_{i=1}^n b_h^i h(t-i),$$

where $h(t - i)$ is 1 if the target in the $(t - i)$ th trial before the current was in the same channel and 0 if it was presented in the respective other.

The weights were fit using the MATLAB function *glmfit* with a logit link and no constant term. For each animal, sessions were combined into sets that each contained all probability distributions (four single sessions in Experiments 1 and 3, three sessions in Experiment 2). For each experiment and animal, at least five such sets were combined randomly and corresponding models were fitted, resulting in a total of 86 sets. For each such set four versions of the model were fitted, the full model above and the following reduced versions.

Stimulus parameters only:

$$L(t) = \mathbf{b}_s \cdot \mathbf{s}(t),$$

Stimulus parameters + probability:

$$L(t) = \mathbf{b}_s \cdot \mathbf{s}(t) + b_p p(t),$$

Stimulus parameters + history:

$$L(t) = \mathbf{b}_s \cdot \mathbf{s}(t) + \sum_{i=1}^n b_h^i h(t - i).$$

For each set and model version, the deviance between the animal's response and the probability p_{go} was collected and normalized to the model deviance for the model version including stimulus parameters only.

Ethics. All experiments were approved by the responsible authorities (Lower Saxony State Office for Consumer Protection and Food Safety, license no. 33.9-42502-04-13/1271).

Data accessibility. All data presented in this manuscript are available from the Dryad Digital Repository at: <https://dx.doi.org/10.5061/dryad.t76hdr7x9> [39]. The code for the experiment control and data registration can be found on GitHub: <https://github.com/Spunc/PsychDetect>.

Authors' contributions. Conception and design of the work: K.J.H., M.M.R. and L.O. Data collection: M.M.R., I.R. and K.S. Data analysis and interpretation: M.M.R., L.O. and K.J.H. Drafting the article: M.M.R. and K.J.H. Critical revision of the article: L.O., M.M.R. and K.J.H.

Competing interests. We declare we have no competing interests.

Funding. This work was supported by the DFG Cluster of Excellence 'Hearing4all' (grant no. EXC 1077/1).

Acknowledgements. The authors thank C. Koeppl, M. S. Malmierca and G.V. Carbajal for their helpful comments and productive scientific discussions and D. Gonschorek for the excellent support. Additionally, the authors would like to thank M. Maleki and A. A. Korte for the support with animal handling and behavioural training.

References

- Carbajal GV, Malmierca MS. 2018 The neuronal basis of predictive coding along the auditory pathway: from the subcortical roots to cortical deviance detection. *Trends Hear.* **22**, 1–33. (doi:10.1177/2331216518784822)
- Skerritt-Davis B, Elhilali M. 2018 Detecting change in stochastic sound sequences. *PLoS Comput. Biol.* **14**, e1006162. (doi:10.1371/journal.pcbi.1006162)
- Bregman AS. 1994 *Auditory scene analysis: the perceptual organization of sound*. New York, NY: MIT Press.
- Sussman ES. 2007 A new view on the MMN and attention debate: the role of context in processing auditory events. *J. Psychophysiol.* **21**, 164–175. (doi:10.1027/0269-8803.21.34.164)
- Woods KJP, McDermott JH. 2015 Attentive tracking of sound sources. *Curr. Biol.* **25**, 2238–2246. (doi:10.1016/j.cub.2015.07.043)
- Scharf B, Quigley S, Aoki C, Peachey N, Reeves A. 1987 Focused auditory attention and frequency selectivity. *Percept. Psychophys.* **42**, 215–223. (doi:10.3758/BF03203073)
- Bargones JY, Werner LA. 1994 Adults listen selectively; infants do not. *Psychol. Sci.* **5**, 170–174. (doi:10.1111/j.1467-9280.1994.tb00655.x)
- Gordon N, Koenig-Robert R, Tsuchiya N, van Boxtel JJ, Hohwy J. 2017 Neural markers of predictive coding under perceptual uncertainty revealed with Hierarchical Frequency Tagging. *eLife* **17**, e3000233. (doi:10.7554/eLife.22749)
- Greenberg GZ, Larkin WD. 1968 Frequency-response characteristic of auditory observers detecting signals of a single frequency in noise: the probe-signal method. *J. Acoust. Soc. Am.* **44**, 1513–1523. (doi:10.1121/1.1911290)
- Wolmetz M, Elhilali M. 2016 Attentional and contextual priors in sound perception. *PLoS ONE* **11**, e0149635. (doi:10.1371/journal.pone.0149635)
- Girshick AR, Landy MS, Simoncelli EP. 2011 Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nat. Neurosci.* **14**, 926–932. (doi:10.1038/nn.2831)
- Parras GG, Nieto-Diego J, Carbajal GV, Valdés-Baizabal C, Escera C, Malmierca MS. 2017 Neurons along the auditory pathway exhibit a hierarchical organization of prediction error. *Nat. Commun.* **8**, 2148. (doi:10.1038/s41467-017-02038-6)
- Heilbron M, Chait M. 2018 Great expectations: is there evidence for predictive coding in auditory cortex? *Neuroscience* **389**, 54–73. (doi:10.1016/j.neuroscience.2017.07.061)
- den Ouden HEM, Kok P, de Lange FP. 2012 How prediction errors shape perception, attention, and motivation. *Front. Psychol.* **3**, 548. (doi:10.3389/fpsyg.2012.00548)
- Kanai R, Komura Y, Shipp S, Friston K. 2015 Cerebral hierarchies: predictive processing, precision and the pulvinar. *Phil. Trans. R. Soc. B* **370**, 20140169. (doi:10.1098/rstb.2014.0169)
- Kok P, Rahnev D, Jehee JFM, Lau HC, de Lange FP. 2012 Attention reverses the effect of prediction in silencing sensory signals. *Cereb. Cortex* **22**, 2197–2206. (doi:10.1093/cercor/bhr310)
- Khouri L, Nelken I. 2015 Detecting the unexpected. *Curr. Opin. Neurobiol.* **35**, 142–147. (doi:10.1016/j.conb.2015.08.003)
- Rogalla MM, Rauser I, Schulze K, Osterhagen L, Hildebrandt KJ. 2019 Mice tune out not in: violation of prediction drives auditory saliency. *bioRxiv*, 633388. (doi:10.1101/633388)
- Gill P, Woolley SMN, Fremouw T, Theunissen FE. 2008 What's that sound? Auditory area CLM encodes stimulus surprise, not intensity or intensity changes. *J. Neurophysiol.* **99**, 2809–2820. (doi:10.1152/jn.01270.2007)
- Lakatos P, Musacchia G, O'Connell MN, Falchier AY, Javitt DC, Schroeder CE. 2013 The spectrotemporal filter mechanism of auditory selective attention. *Neuron* **77**, 750–761. (doi:10.1016/j.neuron.2012.11.034)
- Schwartz ZP, David SV. 2018 Focal suppression of distractor sounds by selective attention in auditory cortex. *Cereb. Cortex* **28**, 323–339. (doi:10.1093/cercor/bhx288)
- Itatani N, Klump GM. 2017 Animal models for auditory streaming. *Phil. Trans. R. Soc. B* **372**, 20160112. (doi:10.1098/rstb.2016.0112)
- Noda T, Kanzaki R, Takahashi H. 2013 Stimulus phase locking of cortical oscillation for auditory stream segregation in rats. *PLoS ONE* **8**, e83544. (doi:10.1371/journal.pone.0083544)
- Lina IA, Lauer AM. 2013 Rapid measurement of auditory filter shape in mice using the auditory brainstem response and notched noise. *Hear. Res.* **298**, 73–79. (doi:10.1016/j.heares.2013.01.002)
- de Hoz L, Nelken I. 2014 Frequency tuning in the behaving mouse: different bandwidths for discrimination and generalization. *PLoS ONE* **9**, e91676. (doi:10.1371/journal.pone.0091676)
- Ulanovsky N, Las L, Nelken I. 2003 Processing of low-probability sounds by cortical neurons. *Nat. Neurosci.* **6**, 391–398. (doi:10.1038/nn1032)
- Ulanovsky N, Las L, Farkas D, Nelken I. 2004 Multiple time scales of adaptation in auditory cortex

- neurons. *J. Neurosci.* **24**, 10 440–10 453. (doi:10.1523/JNEUROSCI.1905-04.2004)
28. Nelken I, Yaron A, Polterovich A, Hershenhoren I. 2013 Stimulus-specific adaptation beyond pure tones. In *Basic aspects of hearing* (eds BCJ Moore, RD Patterson, IM Winter, RP Carlyon, HE Gockel), pp. 411–418. New York, NY: Springer.
 29. Yaron A, Hershenhoren I, Nelken I. 2012 Sensitivity to complex statistical regularities in rat auditory cortex. *Neuron* **76**, 603–615. (doi:10.1016/j.neuron.2012.08.025)
 30. Rubin J, Ulanovsky N, Nelken I, Tishby N. 2016 The representation of prediction error in auditory cortex. *PLoS Comput. Biol.* **12**, e1005058. (doi:10.1371/journal.pcbi.1005058)
 31. Tiitinen H, May P, Reinikainen K, Näätänen R. 1994 Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature* **372**, 90–92. (doi:10.1038/372090a0)
 32. McFadyen J, Smout C, Tsuchiya N, Mattingley JB, Garrido MI. 2019 Surprising threats accelerate evidence accumulation for conscious perception. *bioRxiv* 525519.
 33. Wang L, Krauzlis RJ. 2018 Visual selective attention in mice. *Curr. Biol. CB* **28**, 676–685. (doi:10.1016/j.cub.2018.01.038)
 34. Chapuis GA, Chadderton PT. 2018 Using temporal expectation to assess auditory streaming in mice. *Front. Behav. Neurosci.* **12**, 205. (doi:10.3389/fnbeh.2018.00205)
 35. Johnson KR, Erway LC, Cook SA, Willott JF, Zheng QY. 1997 A major gene affecting age-related hearing loss in C57BL/6 J mice. *Hear. Res.* **114**, 83–92. (doi:10.1016/S0378-5955(97)00155-X)
 36. Kane KL, Longo-Guess CM, Gagnon LH, Ding D, Salvi RJ, Johnson KR. 2012 Genetic background effects on age-related hearing loss associated with Cdh23 variants in mice. *Hear. Res.* **283**, 80–88. (doi:10.1016/j.heares.2011.11.007)
 37. Wickens TD. 2001 *Elementary signal detection theory*. New York, NY: Oxford University Press.
 38. Busse L, Ayaz A, Dhruv NT, Katzner S, Saleem AB, Schölvinck ML, Zaharia AD, Carandini M. 2011 The detection of visual contrast in the behaving mouse. *J. Neurosci.* **31**, 11 351–11 361. (doi:10.1523/JNEUROSCI.6689-10.2011)
 39. Rogalla MM, Rauser I, Schulze K, Osterhagen L, Hildebrandt KJ. 2020 Data from: Mice tune out not in: violation of prediction drives auditory saliency. Dryad Digital Repository. (<https://dx.doi.org/10.5061/dryad.t76hdr7x9>)