Natural sound characteristics explain perceptual categorization

Chi Chen[1,2,3,4], Freddy Trinh[5], Nicol Harper[5], Livia de Hoz[1,4,*]

[1]Department of Neurogenetics, Max Planck Institute of Experimental Medicine, Göttingen, Germany.
[2]International Max Planck Research School for Neurosciences, Göttingen, Germany.
[3]Göttingen Graduate School of Neurosciences and Molecular Biosciences, Germany.
[4]Charité Medical University, Neuroscience Research Center, Berlin, Germany.
[5]Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, UK.

*Corresponding author: livia.dehoz@charite.de

Keywords: auditory, behavior, discrimination, generalization, dimension, feature, prediction

Abstract

As we interact with our surroundings, we encounter the same or similar objects from different perspectives and are compelled to generalize. For example, we recognize dog barks as a distinct class of sound, despite the variety of individual barks. While we have some understanding of how generalization is done along a single stimulus dimension, such as frequency or color, natural stimuli are identifiable by a combination of dimensions. To understand perception, measuring the interaction across stimulus dimensions is essential. For example, when identifying a sound, does our brain focus on a specific dimension or a combination, such as its frequency and duration? Furthermore, does the relative relevance of each dimension reflect its contribution to the natural sensory environment? Using a 2-dimension discrimination task for mice we tested untrained generalization across several pairs of auditory dimensions. We uncovered a perceptual hierarchy over the tested dimensions that was dominated by the sound's spectral composition. A model tuned to the predictability inherent in natural sounds best explained the behavioral results, suggesting that the perceptual hierarchy parallels the predictive content of natural sounds.

Introduction

Animals, including humans, learn to discriminate between the diverse sensory stimuli in their environments and to generalize their behaviors to new instances of those stimulus types. For example, rain can sound dramatically different as it falls on different surfaces and yet we have no problem classifying it as rain and discriminating it from speech. What is the basis of this 'carving nature at its joints' (Plato, 370BC)[1], by which our brains separate the world into objects which fall into different classes e.g., footsteps, motorbikes, human voices, smell of rain? Do we discriminate two stimuli to be of different classes, or generalize them into the same class, because of isolated physical properties such as frequency or color, or because of their characteristics in some more abstract multi-dimensional space? Is this classification determined by the relative sensitivity of sensory neurons to different stimulus dimensions? And is this a reflection of the contribution of these dimensions to the information in natural stimuli?

Most of our knowledge of stimulus perception is derived from tests along a single stimulus dimension. For example, psychophysics and electrophysiology studies have shown that the visual system can discriminate or categorize along color, contrast, or orientation. Or that auditory system is able to discriminate or categorize along sound frequency[2–6], intensity[5–8], frequency modulation direction[9–12] or amplitude modulation frequency[4,13]. Yet stimulus discrimination along a single dimension can be influenced by stimulus features outside that dimension, as has been demonstrated in auditory processing[14–18]. While in the visual modality, dimension integration has been studied by training animals such as pigeons to use dimension combinations[19], or by asking human subjects to identify pop-up stimuli of different level of complexity[20], little is known about how multiple dimensions within any given modality are naturally integrated when classifying environmental stimuli. To understand the interactions between stimulus dimensions, as well as the hierarchy of importance of the dimensions, we focused on the auditory domain. We trained mice to perform 2-dimensional sound discriminations before testing untrained generalization. For example, we trained them to classify a high-frequency upward-sweep as safe and a low-frequency downward-sweep as unsafe. These two sounds differed, therefore, in both frequency and the direction of the modulation. We then tested generalization by examining how the mice classified novel stimuli that varied independently in both dimensions. For example, we determined if they judged a low-frequency upward-sweep to be safe, as the direction would suggest, or unsafe, as the

frequency would suggest. By measuring how mice perceived and acted upon these changes we aimed to understand which dimensions tend to dominate in the discrimination of sounds (e.g. frequency or sweep direction).

Furthermore, we also investigated whether there might be general principles of sound classification that could explain the observed behavioral patterns of discrimination and classification. We hypothesized that the degree to which a dimension influences discrimination is a consequence of the degree of representation of that dimension in the neural population activity, and that this in turn reflects the degree of representation in the diverse array of natural sounds experienced by an animal over its lifetime or evolutionary history. There are several possible principled frameworks that have been posited to govern this representation[21,22]. One hypothesis is that our brain seeks to efficiently compress the incoming input, removing redundant information[23–27]. Another hypothesis is that our brain seeks to predict future sensory input[21,22,28–30]. We hypothesized that the similarity or difference of novel sounds will be judged in this compressed or predictive space. To test these general hypotheses, we compared the capacity of a number of models to explain the behavioral responses of the mice.

RESULTS

We trained mice to discriminate between two frequency-modulated sounds, with one sound being ‘safe’, indicating water would be delivered at a spout upon a nose-poke, and the other being ‘unsafe, indicating an aversive air puff would be delivered instead of water. The two sounds differed in two out of four dimensions; frequency range, sweep direction, modulation rate, and duration (Fig 1a). For example, for sounds differing in frequency range and sweep direction, the safe sound could be a 9-18 kHz upward-sweep (Fig. 1b, blue frame), and the unsafe sound, a 6-3 kHz downward-sweep (Fig. 1b, red frame). Once the animals had learned to discriminate between the safe and the unsafe sounds, we sparsely introduced novel sounds (always safe) that varied in one or both of the two dimensions. Examining whether the mice responded to these sounds as ‘safe’ or ‘unsafe’ allowed us to explore how they generalize what they had learned. For example, a novel sound could have the frequency range of the safe sound but the direction of the unsafe sound, forcing the animal to make a choice whether to treat the new sound as similar to the safe or to the unsafe (Fig. 1b, other sounds in matrix). Measuring generalization gave us

access to the mouse's perception of the two sound dimensions at play and to establish whether decisions were based on just one of the dimensions or on both dimensions (Fig. 1c).

Each dimension pair was tested in a different group of mice. For all 4 dimensions, we chose a span of values known to elicit discriminative responses in neurons in the auditory system. The mice were trained in the Audiobox (Fig. 1d), an apparatus in which mice live in groups for the duration of the experiment (several weeks) while performing the task *ad libitum*. The Audiobox consists of 2 main compartments: a home-cage where food is always available and a 'drinking chamber' where water is delivered upon a nose-poke. As the mouse entered the chamber, a given sound began to play repeatedly for the duration of the visit. In those visits in which the 'safe sound' was presented, mice had access to water (Fig. 1e, left), whereas when the 'unsafe sound' was played, nose-poking was followed by an aversive air puff and no access to water (Fig. 1e, right). The task, therefore, resembled a Go/No-go paradigm. During the generalization phase, two novel sounds (appearing in 11% of the chamber visits each) were introduced, and changed every 2 to 4 days (Fig. 1f). These sounds behaved like 'safe sounds' in the sense that in the visits in which either of these sounds played, the door giving access to water opened upon a nose-poke (Fig. 1e, left. We measured a mouse's response to a given sound as its level of nose-poke avoidance, i.e. the percentage chance that it will not poke its nose in that visit (the 'response').

Frequency range but not direction of modulation controlled behavior

The first group of mice was trained to discriminate sweeps that differed in frequency range and sweep direction (9 to 18 kHz upward vs. 6 to 3 kHz downward sweeps, n = 9; Fig. 1b). Each day, discrimination was measured as the difference between the response to the safe sound (Fig. 1f, blue line) and that to the unsafe sound (Fig. 1f, red line). Mice that showed stable discrimination performance for at least three consecutive days (See Methods), were included in the analysis and then tested with novel sounds varying along both dimensions (Fig. 1b). Mice tended to either avoid or approach the novel sounds, reflecting whether they perceived them as similar to the 'unsafe' or the 'safe' sound respectively (Fig. 1f, black lines).

The mean response for this first set of sounds (safe, unsafe and novel, Fig. 2a, top), what we term the 'generalization pattern', is shown in both the form of a performance plot (middle) and color-coded matrix (bottom). We found that whether a novel sound was

treated more like the unsafe or safe sound, that is how it was generalized, depended mostly on its frequency (the average correlation coefficient between frequency and behavior, ρ(freq), was -0.89) and not its sweep direction (ρ(dir) = 0.17). Regardless of the direction, mice approached novel sounds of high frequency ranges (6-12 kHz and 9-18 kHz), and avoided novel sounds of low frequency ranges (3-6 kHz and 4-8 kHz). A two-way ANCOVA revealed a significant main effect of frequency range ($F(1,44) = 117.73$ $p = 0$), and a weak effect of sweep direction ($F(1,44) = 6.85$, $p = 0.01$), but no frequency × direction interaction ($F(1,44) = 2.52$, $p = 0.12$).

To rule out the possibility that the dominant role of frequency resulted from a more salient difference along the frequency dimension, we narrowed the frequency dimension and trained mice to discriminate sweeps of partially overlapping frequency ranges but opposite directions (6 to 12 kHz upward vs. 8 to 4 kHz downward sweeps; $n = 7$; Fig. 2b, top). The generalization pattern revealed a similar pattern as before. Mice discriminated the two sounds well and, when exposed to novel sounds, responded mainly to their frequency range ignoring direction (the average correlation coefficient between frequency and behavior, ρ(freq), was -0.85; between direction and behavior, ρ(dir) = -0.13; Fig. 2b, bottom). A two-way ANCOVA revealed a significant main effect of frequency range ($F(1,36) = 89.31$, $p = 0$), but no effect of sweep direction ($F(1,36) = 1.49$, $p = 0.23$)) or frequency x direction interaction ($F(1,36) = 1.62$, $p = 0.21$).

Physiological studies have found a systematic representation of preferred sweep direction along the tonotopic axis at both the cortical and subcortical level, such that low-frequency neurons (with a characteristic frequency of < 8 kHz) prefer upward sweeps and high-frequency neurons prefer downward sweeps[31–33]. To further confirm that the observed frequency dominance was not caused by the particular choice of frequency-direction combination, we trained mice with sweeps of two frequency ranges that were modulated in their preferred direction (9 to 18 kHz downward vs. 6 to 3 kHz upward sweeps; $n = 9$; Fig. 2c, top). Our results showed that the mice selectively attended to the frequency range of the novel sounds, and ignored sweep direction (the average correlation coefficient between frequency and behavior, ρ(freq), is -0.91; between direction and behavior, ρ(dir) = 0.08; Fig. 2c, bottom). A two-way ANCOVA for generalization patterns revealed a significant main effect of frequency range ($F(1,68) = 204.86$, $p = 0$), but no effect of sweep direction ($F(1,68) = 1.04$, $p = 0.31$) or frequency x direction interaction ($F(1,68) = 0.32$, $p = 0.57$) .

Up to this point we had used sweeps of 20 ms duration, based on discriminable durations in mice[34]. Since previous studies have shown that sweep direction discrimination

can be limited by the sweep duration[14,17,18], we now trained and tested mice with 40 ms long sweeps (9 to 18 kHz downward vs. 6 to 3 kHz upward sweeps; n = 7; Fig. 2d, top). We found that an increase in sweep duration did not change the generalization pattern (the average correlation coefficient between frequency and behavior, ρ(freq), is -0.89; between direction and behavior, ρ(dir) = 0.16; Fig. 2d, bottom). A two-way ANCOVA revealed a significant main effect of frequency range ($F(1,52) = 122.37$, $p = 0$), but no effect of sweep direction ($F(1,52) = 3.44$, $p = 0.07$) or frequency x direction interaction ($F(1,52) = 0.54$, $p = 0.47$) . Taken together, these results (Fig. 2a-d) show that sound frequency is perceived as more relevant than sweep direction.

The mouse behavior is consistent with the auditory system being optimized for the representation of the predictive features of natural sounds

We examined if some general principle could explain how the mice generalize over frequency and sweep direction, or any given set of dimensions. We hypothesized that the generalization patterns reflect the relevant characteristics of the auditory environments that the animal evolved in and was reared in. However, this begs the question of what defines the relevant characteristics. We considered two alternative hypotheses, that those characteristics of natural sound that are predictive of future input are important, or alternatively that those characteristics that allow a compressed representation are important.

To this end we constructed three models the cochleagram-only model, the prediction model and the compression model – and used them to predict the behavioural responses (Fig. 2e). The models consisted of several stages (Fig. 2f). The first stage of each model was a model of the cochlea[35], that decomposed the sound waveform into frequency bands. For each task, the sounds were put through this simple model to produce a 'cochleagram' for each sound, the activity in each frequency band over time resembling the activity in the auditory nerve or brainstem[35].

The second stage consisted of an encoding model. Although abstract, this stage of the model embodies the neural processing of the cochlea activity that provides a representation of sound used for decision. Given the temporal resolution of the model (5ms time bins) is approximately that of the primary auditory cortex[30,36], this is perhaps the most appropriate analogue, but auditory midbrain might also be appropriate. In the prediction model, we used a temporal prediction model[30] pre-trained to predict the future of natural sounds from

their past. This embodies that hypothesis that the neural representation is optimized for prediction of future natural input from its past. In the compression model, we used an autoencoder[37] pre-trained to compress natural sounds efficiently. This embodies the hypothesis that the neural representation is optimized for the efficient compression of the incoming natural sounds. Finally, the cochleagram-only model has no encoding model, instead applying the decision model directly to the cochleagram representation. This embodies the hypothesis that the neural representation was not optimized to predict or compress the statistics of natural stimuli beyond the frequency decomposition of the cochlea[38,39].

More specifically, for the encoding model for the prediction model we used a temporal prediction model based on that in Singer et al. (2018), which has been shown to produce tuning properties of auditory cortical neurons, except that we developed a new version that included recurrent connections. To do this, we trained a recurrent neural network to predict the immediate future (5ms) of cochleagrams of natural sounds from the recent past (20ms), using a large corpus of natural sounds (Fig. 1g). The network had a single recurrent hidden layer (1600 hidden units) whose activity over time provides the representational space and was the input to the decision model. The network connection weights were constrained to be sparse. The trained network was able to learn representations that resemble spectro-temporal receptive fields (STRFs) as found in the auditory cortical neurons (Fig. 1h, Supp. Fig. S2a). The encoding model for the compression model was similar, but instead of predicting the future, it estimated its own input - the cochleagram in the recent past (20ms). The network weights were constrained to be sparse, as was its activity (making it a sparse recurrent autoencoder). The trained network also learned representations that had some STRF-like characteristics (Supp. Fig S2b).

The final stage was the decision model, which was used to predict the mean response of the mice for each stimulus of each task (Fig. 2g). Here, for each task, the simple decision model was fit to map representations of the safe and unsafe stimuli from the encoding model activity to the corresponding behavioral responses. We then tested what response the full model would estimate for the novel sounds using the fitted decision model. The decision model was a logistic regression, which mapped from the encoding model activity to a number between 0 and 100%, which was the model's estimate of the animal response – chance of giving no-nose-pokes. The regression weights were constrained to be sparse - this drives any regression weights to zero that are not useful, finding a parsimonious

mapping. The strength of this constraint for a given task was set as the strength that on-average best predicted the responses to the novel sounds on all the other tasks.

The prediction model qualitatively captured the behavioral data relatively well, which was also the case for the compression model (Fig. 2e, top and middle). However, the slight dependence of response on sweep direction that is evident in the behavioral data is more apparent for the prediction model than in the compression model. The cochleagram-only model failed to respond differently for the different auditory stimuli (Fig. 2e, bottom), indicating that for this model it is better to give a fixed output than to let the output depend on the cochleagram.

The behavioral control of sweep modulation rate, direction, duration and frequency range reveals a hierarchy of behavioral relevance

To test whether and how these results extended to other dimension combinations, we performed the same experiment using sounds that differed in the combination of modulation rate and direction, sweep duration and direction, or sweep duration and frequency-range.

We trained mice to discriminate sounds that differed in modulation rate and direction but had the same center frequency of 8 kHz (100 octave/sec upward sweep as the safe sound vs. 50 octave/sec downward sweep as the unsafe sound; n = 8; Fig. 3a, top). We chose 50 and 100 octave/s because these rates are within the range (25-250 octaves/s) that elicit discriminable neural responses[40]. Generalization patterns showed that mice responded to novel sounds according mainly to their sweep direction, and largely ignoring modulation rate (the average correlation coefficient between direction and behavior, ρ(dir), was -0.83; between rate and behavior, ρ(rate) = -0.33; Fig. 3a, bottom). A two-way ANCOVA analysis revealed a significant main effect of sweep direction ($F(1,60 = 107.18$, $p = 0$), a significant main effect of modulation rate ($F(1,60) = 16.53$, $p = 0.0001$), but no interaction ($F(3,56)= 1.09$, $p = 0.30$). This result further supports the idea that sound dimensions interact with each other such that a given dimension (e.g. direction) can have a stronger effect on behavior than another (e.g. rate) but be ignored when presented with a third (e.g. frequency). This suggests a hierarchy of relevance of sound dimensions in sound discrimination.

We then tested a new set of mice with sounds that combined sweep duration and direction (20 ms 4-16 kHz upward sweeps as safe *vs* 80 ms 16-4 kHz downward sweeps as

unsafe; n = 10; Fig. 3b, top). For the sweep duration, we used 20 and 80 ms[34]. Responses changed monotonically as a function of changes in both sound dimensions (the average correlation coefficient between direction and behavior, ρ(dir), was -0.67; between duration and behavior, ρ(dur) = 0.63; Fig. 3b, bottom). It is worth noting that changing the duration of a sweep of fixed frequency range inevitably also alters the modulation rate. Here, the modulation rates for the safe and the unsafe sound were 100 and 25 Oct/s respectively, i.e. further apart than in the previous test, where they varied between 100 and 50 Oct/s. Since in the previous experiment the role of modulation rate in discrimination was weak, it is unlikely that mice based their choice on the velocity dimension instead of the duration dimension. A two-way ANCOVA revealed a significant main effect of sound duration ($F(1,76) = 119.23$, $p = 0$), a significant main effect of sweep direction ($F(1,76) = 132.03$, $p = 0$), but no duration x direction interaction ($F(1,76) = 0.37$, $p = 0.55$). This result indicates that both dimensions approximately equally contribute to sound discrimination.

Finally, we investigated how mice generalized along the combined dimensions of frequency range and sweep duration (9 to 18 kHz 80 ms sweeps as safe vs. 3 to 6 kHz 20 ms sweeps as unsafe; n = 8; Fig. 3c, top). We found that both dimensions exerted strong control over the animals' behavior at the lower frequency ranges (around the unsafe sound) while changes in either dimension were ignored near the safe sound (the average correlation coefficient between frequency and behavior, ρ(freq), was -0.66; between duration and behavior, ρ(dur) = -0.31; Fig. 3c, bottom). A two-way ANCOVA revealed a significant main effect of frequency range ($F(1,92) = 66.01$, $p = 0$), a significant main effect of duration ($F(1,92) = 14.13$, $p = 0.0003$) and a significant interaction between the two dimensions ($F(1,92) = 8.86$, $p = 0.004$). This suggests that the influence of duration on animals' behavior is frequency dependent.

Finally, we examined the models' predictions for these tasks (Fig. 3d). The prediction and compression models both qualitatively captured the behavioral data of direction/rate and direction/duration tasks fairly well (Fig. 3d, top and middle). For the frequency/duration task, the prediction model better captures the behavior than the compression model, as the compression model is not as sensitive when the novel sound became more different to the unsafe sound in frequency interval or duration. The cochleagram-only model performs poorly, as for the frequency/direction tasks, but differentiated response probabilities are observed for stimulus set frequency/duration, although it fails to produce response probabilities similar to the behavioral data (Fig. 3d, bottom).

Mice can be trained to discriminate along the non-preferred dimension, but their learning remains localized.

Up to this point we have tested the untrained response to sound stimuli varying in one or both dimensions with respect to either the safe or unsafe sounds. To further characterize the interaction dynamics between dimensions, we first forced mice to discriminate along a non-preferred dimension (e.g. sweep direction, when against frequency range) and then examined generalization again. In particular, we assessed whether the trained dimension (e.g. sweep direction, non-preferred when combined with frequency) can take over behavioral control in detriment of the other dimension (frequency range, preferred).

A new group of mice was trained to discriminate sweep pairs of opposite direction but same frequency range, so as to force direction discrimination (Fig. 4a, 5-10 kHz upward sweeps as safe vs. 10-5 kHz downward sweeps as unsafe, n = 8; Fig. 4a, top). Discrimination learning was slow (Supplementary Fig. 3j) but the safe and unsafe sound were well discriminated after 50 trials of training (Fig. 4a, colored symbols). As mentioned before, direction selectivity in the auditory system is tonotopically organized, such that the physiological preferred direction of neurons of low characteristic frequency is an upward sweep[32,33,41]. We expected, therefore, direction discrimination to be frequency-specific. Generalization to novel sounds was later measured with sweeps of successively different frequency range and direction (8-4 kHz downward sweeps and 6-12 kHz upward sweeps). Learning did not generalize across direction at different frequency ranges, as revealed by strong avoidance to both tested sounds despite opposite sweep direction (Fig. 4a, black symbols). To strengthen the conditioned-valence of the downward direction, we then conditioned the mice to the already tested downward sweep sound (8-4 kHz downward sweeps; Fig. 4b, colored symbols) before testing again generalization on the various novel sounds (Fig. 4b, black symbols). This conditioning did succeed in generating direction generalization across wider frequency ranges (Fig. 4b, black symbols). The average correlation coefficient between direction and behavior was -0.82, whereas the value between frequency and behavior was 0.12 (Fig. 4b). However, generalization across direction was still limited in frequency range, as reflected by the absence of categorical responses to the sweeps of frequency range beyond the trained ones (3-6 kHz upward and 18-9 kHz downward sweeps; Fig. 4b outer most black symbols). A two-way ANCOVA revealed a significant main effect of sweep direction ($F(1,60) = 74.2$, $p = 0$), but no effect

of frequency range ($F(1,60) = 1.89$, $p = 0.17$) or interaction between the two dimensions ($F(1,60) = 0.53$, $p = 0.47$). Therefore, generalization along the non-preferred dimension could be trained but remained specific in frequency, possibly because plasticity occurred only within a specific region of the tonotopic map.

We then examined the models' prediction for the task. We trained the decision model for the generalization test after the initial conditioning in the same manner as for the models previously described. For the second generalization test, we initialized the decision model parameters with the ones learnt from the first learning phase to emulate the chronological order of the learning process, and then trained the decision model to predict the behavioral responses for the two unsafe stimuli and one safe stimulus. The result shows that both the prediction and compression model were able reproduce the lack of generalizing to the test stimuli of flanking frequency ranges as observed in the behavioral experiment, whereas the cochleogram-only model could not (Fig. 4c).

Behavioral control of sound dimensions is plastic and amenable to training.

We then tested whether we could train mice that already shown a preferred dimension to discriminate along the other, non-preferred, dimension. Here we used the animals that had been trained with the combination of frequency and sweep-direction (Fig. 2a, repeated in Fig. 5a left), where they showed a preference for the frequency dimension, while ignoring direction. These same animals were now given a 2nd conditioning with the same safe sound but a high-frequency downward-sweep as the unsafe sound (Fig. 5a, middle), to forced them to discriminate along the direction dimension. The 2nd conditioning led to an increase in avoidance to the new unsafe sound (from 35% to 65%) but their discrimination did not reach levels comparable to those of the 1st conditioning even after 4 days (Fig. 5a right; paired t-test for d', $p = 0.002$). A similar result was observed in mice previously trained with the combination direction-rate (Fig. 5b, left). When the upward 50 octaves/s was explicitly conditioned to force discrimination along the velocity dimension, we observed increased avoidance to the new unsafe sound (from 56% to 69%) but weaker discrimination than in the original conditioning (Fig. 5b, middle and right; paired t-test for d', $p = 0.003$).

We next examined flexibility in a task in which both tested dimensions influenced behavior. Mice that had been trained with the combination of direction-duration (Fig. 3b), were divided in two equally sized groups of comparable performances. The first half were subsequently trained with an additional unsafe sound (the 2nd unsafe tone) with the same

direction as the initial unsafe sound but the shorter duration of the safe sound (Fig. 5c, n = 5). Thus, the safe and the unsafe sounds differed now only in sweep direction. After this 2nd conditioning, mice now avoided nose-poking in the presence of the added unsafe tone (Fig. 5c, right, colored symbols; 88% of avoidance, d' = 1.45). Importantly, learning not only affected the newly unsafe sound, but was extended to novel sounds in between the safe and unsafe ones (Fig. 5c, right, black symbols). The discrimination gradient was now mainly along the direction dimension. A two-way ANCOVA for performance of the retrained mice revealed a significant main effect of sweep direction ($F(1,36) = 264.73$, $p = 0$), but no effect of duration ($F(1,36) = 0.58$, $p = 0.45$) or interaction between those two dimensions ($F(1,36) = 0.62$, $p = 0.43$).

The other half of the mice (Fig. 5d) were exposed to an additional unsafe tone that had the same duration as the first unsafe tone but the direction of the safe sound. This resulted in safe and unsafe sounds differing only in sound duration (Fig. 5d, right, colored symbols). Interestingly, even though the newly unsafe sound was already treated like an unsafe sound by the animals after the initial conditioning (from 74% to 79%; $d'_{2nd} = 1.24$), responses to novel sounds changed dramatically. Mice now mainly discriminated along the duration dimension (Fig. 5d, right, black symbols). A two-way ANCOVA revealed a significant main effect of sweep duration ($F(1,28) = 51.28$, $p = 0$), a weaker but significant effect of direction ($F(1,28) = 6.27$, $p = 0.02$), but no interaction between those two dimensions ($F(3,24) = 0.65$, $p = 0.43$). Thus, changes in task conditions can shift the behavioral pattern from one where both dimensions are used in discrimination to one in which one dimension is ignored. In summary, mice could be trained to flexibly use a single dimension as the behavioral relevant one, however, discrimination along non-preferred dimension remained more difficult and did not generalize beyond narrow ranges.

As before, we then examined how the models' predictions for the task with 2 unsafe sounds. Here, both the recurrent prediction and compression model were able reproduce the lack of generalizing in the first task (Fig. 5e, f, the left column). In the second task, the second unsafe sound encouraged generalization across duration (Fig. 5c, right) or across sweep direction (Fig. 5d, right) in the behavioral data. This generalization is captured by both the prediction and compression models after second conditioning on duration, but for second conditioning on sweep direction the prediction model appears to capture the phenomenon far better than the compression model (Fig. 5e, right column). The cochleagram-only model produces the same uniform response probabilities and does not predict behavior accurately (Fig. 5e, f, bottom row).

Perception of 'envelope' features in mice

The 'envelope' of the sound stream is important for auditory perception, for example in speech processing[42,43]. The experiments described so far, focused on differences within the short frequency-modulation bouts presented always at 3 Hz (Fig. 1-5). Next, we investigated how the 'envelope', i.e. the contour, influenced generalization patterns. Since amplitude modulations and the periodic envelope are important characteristics for acoustic communication signals[44], we trained and tested mice with 100% sinusoidal amplitude-modulated (AM) stimuli (differing in dimensions of carrier frequency and modulation rate, Fig. 6a).

We chose well discriminable carrier frequencies that differed by 1 octave[45] and amplitude-modulations ranging between 5 and 20 Hz modulation rate[4] ($\Delta F > 1$ octave). Fig. 6a shows responses to trained and novel sounds. We found that, for all mice, performance was mainly controlled by the carrier frequency and influenced by the amplitude modulation (the average correlation coefficient between frequency and behavior, ρ(freq), was 0.83; between amplitude-modulation rate and behavior, ρ(AM) = -0.33). A two-way ANCOVA revealed a significant main effect of carrier frequency ($F(1,124) = 168.76$, $p = 0$), a significant main effect of modulation rate ($F(1,124) = 26.38$, $p = 0$), but no interaction between the two dimensions ($F(1,124) = 0$, $p = 0.98$).

We then tested sounds differing in dimensions of tone frequency and repetition rate (Fig. 6b), using rates ranging between 2 and 8 Hz (ΔF of 2 octaves), since neurons in auditory cortex selectively prefer stimulus rates ranging from 2 to 15 Hz[46]. Similar to amplitude-modulated stimulus, we found that the animals' performance was influenced by both dimensions (the average correlation coefficient between frequency and behavior, ρ(freq), was 0.85; between repetition rate and behavior, ρ(rep) = -0.17). A two-way ANCOVA revealed a significant main effect of tone frequency ($F(1,92) = 189.3$, $p = 0$), a significant effect of repetition rate ($F(1,92) = 6.72$, $p = 0.01$) and significant interaction between the two dimensions ($F(1,92) = 4.36$, $p = 0.04$). Thus, although tone frequency exerts a stronger control over behavior, repetition rate influences in a frequency-dependent manner.

Next, we examined the models' predictions for these tasks. The prediction and compression models are both able to qualitatively capture the responses fairly well (Fig. 6c, d). For stimulus set varying along carrier frequency and amplitude-modulation rate,

they both tend to overly disregard the modulation rate as a variable for the behavioral outcome, which might be due to limitations of the model architectures for capturing modulation frequencies at a low rate. For the task involving tone frequency and repetition rate, the compression model tends to somewhat overestimate the fall-off with lower carrier frequency of the response to the novel sounds, while the prediction model underestimates them. The cochleagram-only model predicts approximately the same response for all sounds in both tasks, and thus fails to mimic the structure seen in the behavioral data (Fig. 6c, d).

Summary and Quantitative comparison of models

For an overview of the effect of the different dimensions on performance, we collated across tasks the correlation coefficients between each dimension tested and performance (Fig. 7a). The variability across individual animals was small, suggesting that the dimensional relevance reflects an innate neural representation in the brain. While frequency consistently exerted considerable control over the behavior across all tests, direction and duration were more variable and their contribution to behavioral decisions depended on the second sound dimension they were tested with. Temporal features such as rate of frequency modulation, amplitude modulation, or repetition rate, had less control over behavior in this type of go/no-go discrimination. Thus, the different features tested here, and in the ranges used, differed in their influence over the behavior in a way that was dependent on the other dimensions of the tested sound, revealing a hierarchy of influence dominated by frequency (Fig. 7b).

This hypothesis is also supported by the result of the prediction and compression models when we calculated the correlation coefficient values for the model results in the same way as we do for the data (Fig. 7a). The prediction model was most similar to the data. We also quantified the results of each model more directly for each task by measuring the mean squared error (MSE) between the models' responses for the novel sounds and the true response of the mice. In Fig. 7c we plot the mean squared error for each task for the prediction model, the compression model and the cochleagram-only model. The cochleagram-only model performs worst in almost all cases indicating that the auditory-nerve level representation is not the space in which sound classes are discriminated. The compression and the prediction models both perform better than the

cochleagram-only model, with the prediction model appearing to perform the best overall, performing better than the compression model for four tasks (narrow-frequency/direction, frequency/direction 40ms, frequency/duration and duration/direction 2nd), worse in two tasks (direction 2nd and frequency/repetition rate) and very similarly in the other nine tasks.

Figure 7c also displays another variation on the prediction model, the scrambled-frequency-channel prediction model. Here, we trained the encoding model using data where, for each time bin, the cochleagram-activity was independently scrambled over frequency channels. This eliminated much of the natural statistical structure of the stimulus, embodying the hypothesis that natural statistics are unimportant for predicting the responses of the mice. The frequency scrambled model performed worse than the prediction and compression models, but better than cochleagram-only model. This provides further evidence that the generalization patterns of the mice, and the hierarchy of importance of sound dimensions in mouse perception, are a consequence of the encoding representation being optimized for representing natural stimuli.

In order to estimate the overall behavioral prediction capacity of each model type, we averaged the mean squared error of all predicted test sound responses for each model type (Fig. 7d). Overall, the prediction model does a better job at predicting the test responses (mean squared error: 0.015) compared to the compression model (mean squared error: 0.027), the frequency-scrambled model (mean squared error: 0.059) and the cochleagram-only model (mean squared error: 0.048). We also explored variations on the compression and prediction model types, by examining the models for a range of durations of the time windows being predicted or compressed by their encoding model part Fig. S5a-b). As in the main results in Fig. 7c-d, the prediction model that predicts the next time bin performed best overall, out of all the model types and variations.

Learning rate did not clearly influence generalization patterns

Next, we wanted to explore whether the generalization pattern was related to the learning rate during initial discrimination training between safe and unsafe sounds. With this aim, we plotted the response to the unsafe sound in blocks of 4 visits starting after the first conditioned visit (Supplementary Fig. 3 and 4, red). The average responses to the safe sound before (gray) and after conditioning (blue) are included. We did this for every dimension combination explored (Supplementary Fig. 3a-g). We also assessed learning rates for single dimension discriminations of frequency, direction and duration (Supplementary Fig. 3h-k).

Discrimination learning in all shown frequency-direction combinations was slow (Supplementary Fig. 3a-d) when compared to discrimination learning based on frequency information alone (Supplementary Fig. 3h-i) but faster when compared to learning based on direction information alone (Supplementary Fig. 3j). For the frequency-direction combination, discrimination learning was progressively faster across tasks in the order shown in Supplementary Fig. 3a-d, and reflected in the final discrimination performance (Supplementary Fig. 3l). Despite these progressive differences in learning rate, the subsequent generalization pattern was comparable across tasks: frequency dominated over direction, and direction had a weak control over behavior throughout (Fig. 2a). For example, even though using the physiologically-preferred direction[31–33] for the given frequency range sped up learning (compare Supplementary Fig. 3a and c), it had no effect on the role of direction in the subsequent generalization pattern.

A very different pattern of results was observed in the direction-duration combination task. When we looked at the learning rate in this task, we saw that sweep direction did not slow down learning (Supplementary Fig. 3f), despite it resulting in slow discrimination learning when alone (compared to duration; Supplementary Fig. 3k and m). One could interpret this as a lack of contribution of direction to direction-duration discrimination learning. And yet, direction contributed greatly to the generalization phase in this task.

Together, the data suggest that the learning rate does not always correlate with the influence of the learned dimensions on the perceived meaning of the sound. This pattern is also observed for the frequency-amplitude modulation and frequency-repetition rate combinations, where we observed different learning rates during initial training (Supplementary Fig. 4a-d) but similar pattern of interaction between frequency and either amplitude modulation or repetition rate in the generalization phase (Fig. 6).

Discussion

We set out to understand how sound dimensions are integrated in auditory perception and to what extent their differential control over behavior can be explained by their relative relevance in natural sounds. The results have consequences for our understanding of perception and the underlying circuits. Using an automatic behavioral paradigm in mice, with a design reminiscent of the implicit bias test used in humans[47], we tested the natural perception of combinations of sound dimensions. We trained mice to distinguish between safe and unsafe sounds that differed in two dimensions, for example frequency and sweep direction, and then tested whether they categorized the safety of novel sounds according to

one or both dimensions. A complex pattern of behavioral generalization over different tasks emerged. This revealed that the perception of acoustic dimensions is hierarchically organized and that this hierarchy determines how acoustic dimensions interact with each other. Furthermore, our modelling suggests that sounds are judged to be different or similar within a representational space optimized for the prediction of the immediate future of natural sounds from their past.

Hierarchical organization of the acoustic dimension processing

Mice categorized novel sounds as 'unsafe' or 'safe' often on the basis of one of the two tested dimensions and ignored the other dimension. This suggests a hierarchy of relevance of the dimensions examined, such that frequency dominates behavioral decisions, followed by frequency sweep direction and duration in equal measure, while sweep rate was largely ignored (Fig. 7b). Mice can however be actively trained to discriminate along their non-preferred dimension, for example sweep direction.

Our findings are consistent with other behavioral studies that did not address the dimension question specifically. For example, rats discriminate along the frequency dimension more easily than along amplitude modulation rate[4] or loudness[48]. In mice, maternal responsiveness to wiggling calls depends critically on both frequency and duration[49,50], which matches our inference that frequency range and duration were at high levels of the perceptual hierarchy.

The hierarchy of stimulus dimensions has a basis in the predictive features of natural sounds

We asked what determines the hierarchy of stimulus dimensions we observe - why do we observe the particular hierarchy that we do, and not, for example, sweep rate being more salient than sound frequency? We hypothesized that the hierarchy is a consequence of the statistics of the natural sounds experienced by an animal during its lifetime and evolution, which would be reflected in the neural representation in the brain (likely forebrain or midbrain). We hypothesized that it is in this representational space that sounds are categorized; for example, in this space, frequency differences could be more relevant than sweep direction differences, and these in turn could be more relevant than sweep rate differences.

Within this perspective, there are many possible frameworks which could determine the representational space learnt from natural sounds. One hypothesis is that the brain seeks to find a representation that compresses the incoming natural sounds, removing redundant information[23–27,51]. Another hypothesis is that the brain seeks to find a representation that enables the efficient prediction of immediate future natural sound input from recent past sound input[21,28,30,52]. To this end we built encoding models that learnt representations of natural sounds optimized for either hypothesis, compression or prediction. We then trained a decision model to estimate the responses of the mice to the safe and unsafe sounds, given the representations provided by these encoding models, and examined the responses predicted for the novel sounds, on which they had not been trained. The models' predictions of the responses to the novel sounds was remarkably consistent with that measured in the experiment. We did not see this consistency when the decision model was applied directly on the cochlea representation without an intervening encoding model, or when the encoding model was trained on non-natural sounds. This suggests that the representational space within which the mice categorize sounds is optimized for natural stimuli.

Although both the compression model and the prediction model typically perform well, overall the prediction model performs slightly better, with the compression model performing poorly for some tasks. This suggests that the representational space within which sound objects are judged to be similar or different may be a consequence of a neural representation optimized to represent the predictive features of natural sounds. For example, this suggests that sound frequency dominates over sweep direction in discrimination as past sound frequencies tend to be more informative about future sounds than the direction of frequency sweeps. Similarly, sweep direction is more relevant for discrimination than sweep modulation rate perhaps because the direction tends to be more informative for prediction than modulation rate. Hence, usefulness for prediction may at least partially determine the basis of sound categories.

Here have compared two unsupervised learning principles by which the brain may learn representations to guide behavior; compression and prediction. The idea that the brain seeks to efficiently compress incoming stimuli has a long pedigree, and includes ideas such efficient coding and redundancy reduction[23,24], sparse coding[26], and many versions of predictive coding[27]. The idea that brain seeks to predict future input, encoding only input which is temporally predictive rather than compressing all input, has been somewhat less explored, and includes ideas such as predictive information[21,22,28] and

temporal prediction[29,30] and also relates to slowness ideas[53,54]. Compression and prediction have typically been used to explain neural responses, rather than animal behavior as we have done. Optimizing models for compression has had some success in explaining certain spatial tuning properties in the visual cortex[25–27], as has prediction of future input[29,30]. However, while both compression[55–59] and prediction[30] have had some success in explaining tuning aspects of neurons in the auditory midbrain or auditory cortex, it has been argued that that prediction models perform somewhat better, particularly for temporal tuning properties[30]. Temporal prediction and compression are not mutually exclusive, and indeed could be combined (e.g. PredNet[60]) or perhaps differentially applied in different sub-systems. Furthermore, while it has been argued that unsupervised processes are important in guiding sensory representation, nevertheless reinforcement signals might also affect its sensory representations, rather than just affecting the decision processes as we assume in our models. Hence, while we have demonstrated a method to ascertain the models and principles governing the representations that guide sensory categorization and have found a prediction models to perform best, there are of course - as always - more prediction, compression and other models that could be explored.

Neurophysiological consideration of dimension perception

Is the hierarchical relevance of acoustic features reflected in the functional and anatomical organization of the auditory system? The lack of hierarchy in our encoding models suggests that an anatomical hierarchy is not required. The notion that sensory systems might be organized hierarchically, i.e. into a series of stages that progressively increase abstract representations, has been a popular hypothesis for decades[61]. Hierarchical processing of sensory information has been intensively studied in the visual modality[62–65]. The influential reverse hierarchy theory proposed by Ahissar and Hochstein[66] posits that high-level abstract inputs are prioritized for perception along the sensory hierarchy and that this hierarchy maps onto an anatomical one. The acoustic dimensions used in this study do not map well onto an anatomical hierarchy of sequential processing across different auditory stations. Neural selectivity to frequency is observed across all auditory stations, from the cochlea to the cortex, and is organized tonotopically[67]. Neural selectivity for sound duration, sweep direction, sweep rate, repetition rate, and modulation rate all appear as early as the auditory midbrain[40,41,68–70]. Thus, the observed hierarchical organization for the dimensions used in this study may not result from sequential feature

extraction along different auditory stations, but from the dynamic representation of each auditory feature in an early station, probably the auditory midbrain or cortex.

The distribution of response selectivity within stations in the auditory system might nonetheless influence the functional hierarchy we observed. For example, while almost all neurons show frequency selectivity, direction or modulation rate selectivity have been found in subpopulations of neurons in auditory cortex and midbrain[32,40]. Both these dimensions are in fact somewhat embedded in the tonotopic map[32,33,41]. It is therefore plausible that a given station in the auditory system processes the direction and rate dimensions of a specific sound, albeit by neurons that are selective for a specific range of frequencies. This is consistent with our finding that the learning of non-preferred dimension discrimination is frequency-dependent, i.e. direction discrimination does not generalize well outside the trained frequency range.

How salient or relevant of a given dimension is to the animal might influence the relative control this dimension has over behavior and the way sounds are generalized. On one hand, the intrinsic salience of sounds might be precisely determined by the natural statistics of the acoustic environment. It could be that frequency is always more salient than direction because it is found to be more informative in natural sounds. On the other hand, subtle differences in saliency might not affect the overall patterns of generalization and the hierarchy we observed. Although we did see differences in learning rates which might have been related with changes in saliency when we varied the characteristics of a given dimension, these did not in turn influence the generalization pattern. For example, changes in the specific direction that was paired with the low or high frequency sound influence the initial learning rate, and yet, these differences in learning rate had no real influence in the generalization pattern, which consistently showed that direction had little control over behavior compared to frequency. We were nonetheless careful to use a variable space covering a range that elicited discriminable responses in neurons in the auditory system.

To our knowledge, our experiment constitutes the first successful use of multidimensional generalization to infer untrained dimensional integration and the hierarchy of auditory information processing. These results yield important insights into the processes underlying auditory object processing and categorization. Tapping into the untrained behavioral response of mice to sounds combining different acoustic dimensions, we found that these dimensions are hierarchically organized in the control they exert over behavior. This emergent hierarchy is consistent with the predictive value of these

dimensions in natural sounds as demonstrated by models. Overall, and inferring beyond the auditory modality, the data support the more general idea that both the sensitivity of physiological responses and behavioral decisions to natural stimuli has evolved to optimally use the predictive value of the different stimulus dimensions in the natural environment.

MATERIALS AND METHODS

Animals

A total of 103 female C57BL/6JOlaHsd (Janvier, France) mice were used. All mice were 5-6 weeks old at the beginning of the experiment. Animals were housed in groups in a temperature-controlled environment (21 ± 1°C) on a 12 h light/dark schedule (7am/7pm) with access to food and water *ad libitum*.

All animal experiments were approved by the local Animal Care and Use Committee (LAVES, Niedersächsisches Landesamt für Verbraucherschutz und Lebensmittelsicherheit, Oldenburg, Germany) in accordance with the German Animal Protection Law. Project license number 33.14-42502-04-10/0288 and 33.19-42502-04-11/0658

Apparatus: the Audiobox

The Audiobox was a device developed for auditory research from the Intellicage (TSE, Germany). The Audiobox served both as living quarters for the mice and as their testing arena. Each animal was individually identifiable through an implanted transponder, allowing the automatic measure of specific behaviors of individual animals. The mice lived in the apparatus in groups of 7 to 10 for several days. Mice were supplied water and food *ad libitum* but are required to discriminate and respond to sounds when they go to the drinking ‘chamber’.

At least one day before experimentation, each mouse was lightly anaesthetized with Avertin i.p. (0.1ml/10g) or isoflurane and a sterile transponder (PeddyMark, 12 mm × 2 mm or 8 mm × 1.4 mm ISO microchips, 0.1 gr in weight, UK) was implanted subcutaneously in the upper back. Histoacryl (B. Braun) was used to close the small hole left on the skin by the transponder injection.

The Audiobox was placed in a temperature regulated room under a 12 h dark/light schedule. The apparatus consisted of three parts: a home cage, a drinking 'chamber', and a long connecting corridor (Supplementary Fig. 1a, b). The home cage served as the living quarter, where the mice have access to food *ad libitum*. Water was delivered in the drinking 'chamber', which was positioned inside a sound-attenuated box. Presence of the mouse in the 'chamber', a 'visit', was detected by an antenna located at the entrance of the chamber. The antenna read the unique transponder carried by each mouse as it enters the chamber. The mouse identification was then used to select the correct acoustic stimulus. A heat sensor within the chamber sensed the continued presence of the mouse. Once in the 'chamber', specific behaviors (nose-poking and licking) could be detected through other infrared sensors. All behavioral data was logged for each mouse individually. Access to water was controlled by opening or closing of the doors behind the nose-poking ports. Air puffs were delivered through an automated valve which is placed on the ceiling of the 'chamber'. A loudspeaker (22TAF/G, Seas Prestige) was positioned above the 'chamber', for the presentation of the sound stimuli. During experimentation, cages and apparatus were cleaned once a week by the experimenter.

Sounds

Sounds were generated using Matlab (Mathworks) at a sampling rate of 48 kHz and written into computer files. Intensities were calibrated for frequencies between 1 and 18 kHz with a Brüel & Kjær (4939 ¼" free field) microphone.

Mice were trained with pairs of either frequency-modulated (FM) sweeps, amplitude-modulation (AM) sounds, or pure tone pips. For 2-dimensional discrimination task, sound pairs used in training as safe and conditioned differed in two out of four chosen dimensions. For the FM sweep sounds these dimensions were frequency range, duration, sweep direction of modulation and velocity of the sweep. For example, when the two dimensions used were the frequency range and the sweep direction, the safe sound could be an upward sweep in the high frequency range while the unsafe sound would be a downward sweep in the low frequency range. Frequency was modulated logarithmically from low to high frequencies (upward sweep) or from high to low frequencies (downward sweep). Sweeps had a duration of 20 ms (default) or 40 ms, including 5 ms rise/decay, and one of four modulation velocities (50, 62.5, 87.5 or 100 octave/sec; with 50 octaves/s as default). The sweeps were presented at roving-intensities (70 dB ± 3 dB). For the AM stimuli, tested dimensions were carrier frequency and modulation rate. The AM sounds

had 100% sinusoidal modulation and had one of four carrier frequencies (6670, 8404, 10588 or 13340 Hz), as well as one of four modulation rates (5, 8, 12 or 20 Hz). For pure tone pips, the dimension of tone frequency and repetition rate were tested. Similar to the AM stimuli, the pure tones pips had one of four frequencies (6670, 8404, 10588 or 13340 Hz) and one of four repetition rates (2, 3, 5 or 8 Hz). Pure tones had a length of 20ms, with a 5ms rise/fall ramp and were presented with intensities that roved between 67 and 73 dB. For single dimensional discrimination, sounds pairs differed in sweep direction, duration or frequency were used for training.

Our choice of sounds was based on the existing literature. Rodents perceive differences in both sound frequency and sweep direction[9,12,33,40]. Mice show increased sensitivity for frequencies between 4 and 16 kHz[33], and easily discriminate frequency distances of at least 0.20 octave[45,71], therefore, we tested frequency ranges between 3 and 18 kHz such that the center frequency differs by more than 0.60 octave. The choice of a 50 octave/s velocity of modulation was based on both physiological[33] and behavioral[14] data.

The frequency-only discrimination data were from mice trained with upward frequency sweeps of the same range used in the task with 2 dimensions (3 to 6 kHz vs. 9 to 18 kHz). For direction discrimination, the mice described in Fig. 4a were used. These are mice that were trained to discriminate between sweeps of identical frequency range (5-10 kHz) but opposite direction of modulation. Data on duration-only discrimination were obtained from mice trained to discriminate FM upward sweeps (4 to 16kHz) that differed only in duration (20ms vs. 80ms).

Discrimination task

As described in our previous study[45], mice were trained to discriminate two sounds, the ‘safe’ and ‘unsafe sound, in the Audiobox. Throughout the duration of the experiment, one sound (i.e. 9 to 18 kHz FMs) was always ‘safe’, meaning that access to water during these visits was granted upon nose-poke. For the first 5 days, only the safe tone was played in each visit. The doors giving access to the water within the chamber were open on the first day of training and closed thereafter. A nose-poke from the mouse opened the door and allowed access to water. Another sound, a ‘unsafe one (i.e. 6 to 3 kHz FMs), was introduced in a small percentage of ‘unsafe visits and a nose-poke during these visits was associated with an air puff and no access to water (Supplementary Fig. 1c). Most of the visits, however remained safe (safe sound and access to water) as before. The probability of unsafe visits was 9.1% for the first 2 days, increased to 16.7% for the next 2 days, then

stayed at 28.6% until they showed steady discrimination performance for at least 3 consecutive days (Fig. 1f).

Mice that failed to learn the task, i.e. had no differential responses to the safe and the unsafe tone for more than 2 days, were excluded from the analysis. In total, 12 out of 105 mice were excluded.

Multidimensional generalization measurement

During generalization testing, visits consisted of 55.6% of safe visits, 22.2% of conditioned visits and 22.2% of novel visits in which a novel sound was presented. Nose poking during the presentation of the novel tone resulted in opening of the doors that gave access to water, meaning that novel visits were never accompanied by an air-puff, the were safe in nature.

Since the safe and unsafe sound differed in two dimensions, novel sounds represented all possible combinations of values used (see sounds) along each dimension. For example, when using the 9 kHz to 18 kHz upward FM sound as the safe sound and the 6 kHz to 3 kHz downward FM sound as the unsafe sound, tested stimuli (including the safe and unsafe sounds) resulted from factorially combining 4 different frequency ranges with 2 different sweep directions.

On each testing day, only two novel sounds were presented (in 11.1 % of total visits each) in addition to the safe and unsafe sounds. Each novel sound was tested for 4 days (~50 visits). For tasks using 2 dimensions with 2 (e.g. direction) and 4 (e.g. frequency range) values respectively, the order in which novel sounds varying in either or both dimensions were tested was as shown in Supplementary Fig. 1d-e. For tasks testing 2 dimensions with 4 values each, such as the combination of frequency and duration, novel sounds were tested in the order shown in Supplementary Fig. 1f-g.

Analysis of performance in the Audiobox

Data were analyzed using in-house scripts developed in Matlab (Mathwork). Performance traces for different stimuli were calculated by averaging the fraction of visits without nose-pokes over a 24-hour window. Discrimination performance was quantified by the standard measures from signal detection theory, the discriminability (d'). It was calculated with the assumption that the decision variables for the safe and unsafe tone have a Gaussian distribution around their corresponding means and have comparable variances.

The d’ value provides the standardized separation between the mean of the signal present distribution and the signal absent distribution. It is calculated as:

$$d' = Z(HR) - Z(FAR)$$

Where Z(p), p□[0 1], is the inverse of the cumulative Gaussian distribution, HR is the hit rate, where a hit is the correct avoidance of a nose-poke in a conditioned visit, and FAR is the false alarm rate, where a false alarm is the avoidance of a nose-poke in a safe visit. Since d’ cannot be calculated when either the false alarms reach levels of 100% or 0%, in the few cases where this happened we used 95% and 5% respectively for these calculations. This manipulation reduced d’ slightly, and therefore our d’ estimates are conservative.

Statistical analysis

Group comparisons were made using multiple-way ANCOVAs with type I sum of squares after testing for normality distribution using the Shapiro-Wilk test. Samples that failed the normality test were compared using Wilcoxon signed rank test. Multiple comparisons were adjusted by Bonferroni correction. For analysis of data consisting of two groups we used either paired t-tests for within-subject repeated measurements or unpaired t-tests otherwise. For data consisting of more than two groups or multiple parameters we used repeated-measures ANCOVA. All multiple comparisons used critical values from a t distribution, adjusted by Bonferroni correction with an alpha level set to 0.05. Means are expressed ± SEM. Statistical significance was considered if $p < 0.05$.

Models

Introduction

We developed models to predict the responses of the mice to the novel sounds in the different tasks. There were three stages (sub-models) composing the full models, each stage providing input for the subsequent stage.

1. A cochleagram – a simple model of the cochlea.

2. An encoding model – representing the processing of the cochlea input by the auditory nervous system, to the level of the midbrain or cortex. This provides a neural representation.

3. A decision model – representing the transformation from the neural representation to the animal behavioral response; the percentage change of not poking its nose.

There were four types of full model. These only differed in their encoding model (stage 2). The cochleagrams (stage 1) and the decision model (stage 3) had same form in all models. The models were:

1. Prediction model. This had a recurrent temporal prediction model as its encoding model, pre-trained to predict the immediate future of natural sounds.

2. Compression model. This had a sparse recurrent autoencoder as its encoding model, pre-trained to compress natural sounds.

3. Cochleagram-only model. This had no encoding model, with the decision model mapping straight from the cochleagram to the behavioral response.

4. Frequency-scrambled model. This was the prediction model, but pre-trained to predict the immediate future of natural sounds whose frequency information has been scrambled.

Preparing and using a full model consisted of the following process:

1. We trained the encoding model (if present) using a database of natural sounds. We then fixed the weights of the encoding model.

2. For each task, we trained a separate decision model to predict the behavioral responses for the safe and unsafe sounds, using the pre-trained encoding models activity to those sounds as input to the decision model. We then fixed the weights of the decision model for each task. This provides a separate full model for each task that only differ in the parameters of their decision models.

3. For each task, get the output (the ‘behavioral response’) of each full model for the novel sounds.

Below we detail each stage of the full models, including the different types of encoding model.

The cochleagram model

Sounds were converted into cochleagrams using a cochlear model, the spec-Hill model[35]. This very simple model has been demonstrated to provide good input to cortical models for predicting neural responses in primary auditory cortex[35]. This cochleagram model was identical to the previously published version[35] except for the sound frequency range and the time-bin size - we used 5 ms time bins. Briefly, a spectrogram was produced from the sound waveform by taking the power spectrum using 10-ms Hanning windows, overlapping by 5 ms. The power of adjacent frequency channels was summed using overlapping triangular windows (using code adapted from melbank.m, http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html), with 32 log-spaced center frequencies ranging from 500 Hz to 22,627 Hz. The resulting power in each time-frequency bin was then converted to log values, thresholded and then put through a Hill function to provide the cochleagram activity.

The encoding models

Training and validation data for all encoding models

We used an auditory dataset composed of natural sounds: birdsong and other animal calls, environmental sounds, and human speech by children and adults as well as infant vocalizations. Sound recordings of birdsong and mammal and insect calls were drawn from the Macauley Library at the Cornell Lab of Ornithology[i] (https://www.macaulaylibrary.org/), and the British Library Sounds Environment & Nature Collection (https://sounds.bl.uk/Environment). We also used a corpus of ferret vocalizations recorded in our laboratory, which was supplemented by a recording found on https://freesound.org/people/J.Zazvurek/sounds/155115/. Recordings of adult human speech were taken from http://databases.forensic-voice-comparison.net/[72]. We also used some human speech and environmental sounds such as snapping twigs which had been recorded in our laboratory in an anechoic chamber. Recordings of children speaking were from the CHILDES database[73,74] from TalkBank (talkbank.org). Baby vocalizations were from the OxVoc database[75].

All recordings were sampled at or resampled to 44,100 Hz, and converted into cochleagrams. The cochleagrams were divided into training and validation sets such that the proportion of different types of sound were approximately consistent in both datasets. Both sets were normalized to the mean and standard deviation of the training set. Gaussian white noise, with a signal to noise ratio of 6dB, was added to the cochleagram activity of both data sets. Adding such noise for temporal prediction models has been found to produce tuning properties that slightly better resemble the biology (Singer et al., 2020), and can be interpreted as internal neural noise. All cochleagrams were then split into one-second-long snippets. Then to remove uninformative snippets, those resulting snippets for which the root mean square of their cochleagram activity (over frequency channels and time bins) was below twice the standard deviation of the added noise (0.356) were removed. This gave a total of 68,480 snippets in the training set and 7,611 snippets in the test set. We denote the cochleagram activity for a given snippet by $s_f[t]$ where f is the frequency channel and t is the time bin. There are f = 1 to F = 32 frequency channels and t = 1 to T = 200 time bins in the snippet.

Temporal prediction model

The temporal prediction model was optimized to predict the immediate future activity of cochleagrams based on their recent past activity. It was based on the standard temporal prediction model[30] but with the addition of recurrent connections. This recurrent temporal prediction model was a recurrent neural network (RNN), and consisted of an input layer, a single recurrent hidden layer, and an output layer. The input layer activity was the activity in all 32 frequency channels of the immediate past 4 time-bins (20 ms) of the cochleagram. The output layer predicted the activity in the single upcoming future time-bin (5 ms) of the cochleagram. We also investigated predicting a span of 4 and 16 future time-bins (20 and 80 ms respectively).

The input layer of the recurrent temporal prediction model had i = 1 to $I = DF$ = 128 units. This represents the past d = 1 to D = 4 time-bins (20 ms) for all f = 1 to F = 32 frequency channels of the cochleagram. Thus, the activity of the input layer was the activity of the cochleagram in a window of the immediate past, but flattened out into a vector. Hence the activity of input unit $u_i[t] = s_f[t - d + 1]$, where $i = (d\text{-}1)f + f$.

Likewise, the output layer had k =1 to K = QF linear units. This represents all f' = 1 to F = 32 frequency channels for the q =1 to Q future time-bins to be predicted. As we noted

above, in the main figures Q = 1 time bin (5ms), however in supplementary Figure S5, we investigated Q = 4 (20 ms) and Q = 16 (80 ms). Thus, the target predicted by the output layer was the activity of the cochleagram in a window of the immediate future, but flattened out into a vector. Hence the activity of output unit $v_k[t]$ is trying to predict a target $y_k[t] = s_f[t+q]$ where $k = (q-1)f + f$.

The input layer was connected to a recurrent hidden layer with j = 1 to J = 1600 hidden units, which had the rectified linear function as their activation function.

$$\mathrm{g}(x) = \max(0, x)$$

For a given cochleagram snippet, the activity $a_j[t]$ of hidden unit j at time step t was given by the following equation:

$$a_j[t] = \mathrm{g}\left(b_j + \sum_{i=1}^{I} w_{ji} u_i[t] + \sum_{j'=1}^{J} w_{jj'} a_{j'}[t-1]\right)$$

Here, b_j is the bias value for hidden unit j, w_{ji} the weight from the input unit i to the hidden unit j, and $w_{jj'}$ is the recurrent weight from hidden unit j' to hidden unit j. Also, $a_{j'}[t-1]$ is the activity of hidden unit j' at time step t-1.

In turn, the activity $v_k[t]$ of output unit k at time t is given by:

$$v_k[t] = b_k + \sum_{j=1}^{J} w_{kj} a_j[t]$$

Here, b_k is the bias value for output unit k, w_{kj} is the weight from hidden unit j to output unit k.

The trainable parameters were the input weight matrix w_{ji}, the recurrent weight matrix $w_{jj'}$, the hidden unit bias vector b_j, the output weight matrix w_{kj} and the output unit bias vector b_k. These parameters were optimized to perform temporal prediction on the training

set by minimizing the mean squared error with L1-regularization on the parameters. The equation for the objective function that is minimized is:

$$E = \frac{1}{NCK}\sum_{n=1}^{N}\sum_{t=D}^{T-Q}\sum_{k=1}^{K}(v_{kn}[t] - y_{kn}[t])^2 + \lambda\left(\sum_{j=1}^{J}\sum_{i=1}^{I}|w_{ji}| + \sum_{j=1}^{J}\sum_{j'=1}^{J}|w_{jj'}| + \sum_{k=1}^{K}\sum_{j=1}^{J}|w_{kj}|\right)$$

Here we have introduced the additional subscript *n*, where *n* is the snippet number, and N = 9,050 is the number of snippets in the training set. Thus, $v_{kn}[t]$ is the prediction $v_k[t]$ for snippet *n* and $y_{kn}[t]$ is the target $y_k[t]$ for snippet *n*. Here, *C* is the number of valid time bins accounting for the spans of the input and target windows, $C = T\text{-}Q\text{-}D+1$. The L1-regularization strength is governed by the hyperparameter λ.

The training was performed using stochastic gradient descent on mini-batches of 100 training examples and the Adaptive Moment Estimation (ADAM) update function[76]. We also decreased the learning rate for ADAM in a step-wise manner throughout training. We trained the model for 1000 epochs, decreasing the initial set learning rate for ADAM (0.001) by a factor 10 every 150 or 200 epochs. Typically, the objective function reached a steady state within 500 epochs. All weights were initialized using the Glorot uniform initializer[77] except for the hidden-to-hidden weight matrix which was initialized using an identity matrix multiplied by 0.01. The bias vectors for both the hidden and the output layer were initialized to 0. To provide stability, the weights w_{ji}, $w_{jj'}$, w_{kj} were further regularized by using the method of constraining the incoming weight norms to a maximum of 1 throughout training[78].

We did a hyperparameter search for the model using different values of λ: 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7} and 0. We chose the model with the λ-value that resulted in the lowest mean squared prediction error on the held-out validation set. For the Q=1, 4 and 16 models this was $\lambda = 10^{-5}$.

Compression model

The compression model is optimized to encode a compressed representation of the input data, where the output layer reconstructs the original input from the compressed

encoding. The same snippet of the cochleagram is used for both the input and the output, which had the same input dimensions as for the recurrent temporal prediction model, 4 time-bins of the cochleagram with 32 frequency channels per time bin. We also investigated using 1 and 16 time-bins of the cochleagram as the input (5 ms and 80 ms respectively).

The network architecture was identical to the recurrent temporal prediction model apart from the output layer which had the same number of linear units ($I = K = 128$ input/output units = 32 frequency channels × 4 time-bin) as the input layer.

The output was described by

$$v_k[t] = b_k + \sum_{j=1}^{J} w_{kj} a_j[t]$$

Where $v_k[t]$ is the estimate of the input at time step t for output unit k.

The cost function is similar to the recurrent temporal prediction model apart from the target values being the current input rather than the future of the cochleagram. Hence, both the input and target is represented by $u_i[t]$. The cost was calculated as the mean squared error between the input and model reconstruction with L1 regularization on the parameters and the hidden unit activity to give a compressed representation.

$$E = \frac{1}{KCN} \sum_{k=1}^{K} \sum_{t=D}^{T} \sum_{n=1}^{N} (v_{kn}[t] - u_{in}[t])^2 + \lambda \left(\sum_{j=1}^{J} \sum_{i=1}^{I} |w_{ji}| + \sum_{j=1}^{J} \sum_{j'=1}^{J} |w_{jj'}| + \sum_{k=1}^{K} \sum_{j=1}^{J} |w_{kj}| \right) + \lambda_a \sum_{j=1}^{J} \sum_{t=D}^{T} \sum_{n=1}^{N} |a_{jn}[t]|$$

Here, $u_{in}[t]$ the activity of input unit i at time t for training example n, λ_a is the regularization parameter for the hidden unit activity, and $C = T\text{-}D+1$.

Training, initialization, and weight norm constraining was done exactly as with the temporal prediction model, and the objective function also typically reached steady state within about 500 epochs. Unlike the temporal prediction model there is no quantitative method for choosing λ and λ_a, so we relied on a qualitative approach that made sparse autoencoder comparable to the temporal prediction model. To do this we did a hyperparameter search for the sparse autoencoder model using different values of λ: 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7} and 0, while setting λ_a to 0. We chose the model with representations that qualitatively most closely resembled the receptive fields of auditory cortical neurons and had a similar number of active units to the temporal prediction model (Q=1). For the main sparse autoencoder model (D=4) this was a λ-value of 10^{-5}. Using this λ value, we then added regularization to the hidden unit activity and did a hyperparameter search over λ_a: 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7}, 10^{-8}, 10^{-9}, 10^{-10}, and 0. Again, we then chose the resulting model which formed representations qualitatively most similar the receptive fields of auditory cortical neurons and with a similar number of active units to the temporal prediction model (Q=1). For the main sparse autoencoder model (D=4) this was a λ_a-value of 10^{-7}. For the D=1 and D=16 models λ and λ_a were chosen in the same manner but with exploration over somewhat different λ and λ_a ranges, this resulted in $\lambda = 10^{-4}$ and $\lambda_a = 10^{-7}$ for the D=1 model and $\lambda = 7.5 \times 10^{-5}$ and $\lambda_a = 10^{-8}$ for the D=16 model.

Frequency scrambled encoding model

This was the temporal prediction model given above, but when training this encoding model, we scrambled the frequency channels of each natural sound cochleagram. More precisely, for each one second cochleagram snippet, the 32 frequency channels were randomly reordered. This random reordering was different and independent for each cochleagram snippet. The model was then trained in the exact same way as the temporal prediction model above, and it also typically settled within 500 epochs. The behavioral task stimuli (see below) were not frequency scrambled.

Measuring encoding model receptive fields

For the temporal prediction model (Q=1) and the sparse autoencoder model (D=4) we measured the spectrotemporal receptive fields (STRFs) of their hidden units (Fig. 2h and Fig. S2). To do this we took the trained model and gave it directly as input cochleagrams of 500 second duration whose activity at each frequency and time bin were independently

drawn from a Gaussian distribution with a mean of 0 and a standard deviation of 1. We then used the response of each model unit to this stimulus to calculate its STRF using an elastic-net regression model[79,80].

Units which a response of zero throughout this stimulus were declared non-active unit, having no STRF, and only active units were displayed in the figures. The temporal prediction model (Q=1) had 89 active units and the sparse autoencoder model (D=4) had 147 active units. We also measured the number of active units for the other models: temporal prediction model (Q=4), 96, temporal prediction model (Q=16), 102, sparse autoencoder model (D=1), 67, sparse autoencoder model (D=16), 59, frequency scrambled model, 157.

Decision Models

Behavioral Stimulus Set

The sound stimuli for the behavioral discrimination task were processed in a similar manner as for the sounds used for the encoding models. The behavioral task sound snippets used as input to the models consisted of a ~2 second snippets of the repeated sounds played to the mice in the Audiobox. We cosine ramped the first 5 ms of each snippet stimulus set to avoid spectral splatter. All snippets were resampled to 44,100 Hz and converted into cochleagrams. For stimuli with roving amplitude levels we used the average of the cochleagrams at the different levels.

Decision model

For each discrimination task, the decision model was trained with the corresponding behavioral responses of the safe and unsafe trials used as targets, while the associated safe and unsafe sounds were used as the training input. The behavioral responses consisted of an array of zeros and ones with each entry r_m indicating whether a mouse did not poke its nose for a particular trial m (r_m = 0: poke, r_m =1: no-poke) to the sound used in that trial, be it a safe sound or an unsafe sound. Mouse identity was ignored in training the decision model, and the responses of all the mice in the group for that task were collated together.

The safe and unsafe sounds were converted to cochleagrams, then put through an encoding model which had been pre-trained with natural sounds and then had its weights fixed. The encoding model hidden unit activity provided the input to the decision model. The decision model consisted of a logistic regression from the encoding model hidden unit

activity throughout the ~2 second snippet to the behavioral response to a single trial. The decision model was then trained to estimate the behavioral response to the full set of trials for the safe and unsafe sounds.

The behavioral response prediction, p, for the output unit for a given trial is provided by:

$$p = \mathrm{h}\left(b_o + \sum_{j=1}^{J}\sum_{t=1}^{T} w_{jt} a_j[t]\right)$$

Where w_{jt} is the weight from the activity of encoding-model hidden unit *j* at time *t* in the snippet onto the single output unit, $a_j[t]$ is the activity of encoding-model hidden unit *j* at time t in response to the sound snippet, be it the safe of unsafe sound. Also, J is the number of hidden units in the encoding model (J=1600), T is the duration of the snippet in time bins (T=~400) and b_o is the bias. A sigmoid activation function was used for the output unit:

$$\mathrm{h}(x) = \frac{1}{1 + e^{-x}}$$

The trainable parameters for this model, w_{jt} and b_o, as we fix all the weights and biases associated with the encoding model. We used the binary cross entropy as the cost function of the decision model, alongside L1-regularisation of the decision model weights:

$$E = \sum_{m=1}^{M} -r_m log_2(p_m) - (1 - r_m) log_2(1 - p_m) + \lambda \sum_{j=1}^{J}\sum_{t=1}^{T} |w_{jt}|$$

Here, r_m is the behavioral response for trial m and p_m is the model's response prediction for trial m. Also, λ is the regularization strength and M is the number of trials.

The training was performed using stochastic gradient descent on mini-batches of 100 training examples and the Adaptive Moment Estimation (ADAM) update function[76]. We also decreased the learning rate for ADAM in a step-wise manner throughout training. We

trained the model for 150 epochs, decreasing the initial set learning rate for ADAM (0.001) by a factor 10 every 15 epochs. Typically, the objective function reached a steady state within 50 epochs. The weights w_{jt} were initialized using the Glorot uniform initializer[77] and b_o was initialized to 0. For the behavioral tasks that had a second conditioning step to probe generalization, w_{jt} and b_o were initialized to the values learned by the model for the first conditioning.

For each encoding model type, we then performed the following process. For each task, we fit the decision model for a range of λ-values: 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}, and 0. Then for each λ-value we used the model to predict generalization responses to the novel sounds used in the task. For each novel sound this was done by producing its cochleagram and running it through the encoding model. The encoding model activities were then in turn put through the trained decision model to provide *p*, the model's prediction of average response of the mice, that is the average probability of no-nose-poke in response to the novel sound. After getting a *p* for each novel sound, we then measured the accuracy of the predictions over all the novel sounds in the task. We did this by measuring the mean squared error (MSE) between the model predictions and the actual responses. This then provided a plot of MSE as a function of λ for each task. We cannot simply choose the λ that gives us the lowest MSE on the novel sounds, as this would be akin to using the test dataset to choose the hyperparameters. Instead, for each task examined, we averaged together the plots of all the other tasks (except 2nd conditioning tasks), to provide a plot of the average MSE as a function of λ for all tasks except the one being examined, and the λ with the lowest average MSE on this plot was then chosen. This λ-value was then used for the task examined to provide the figures and the summary assessments associated with that task. For the discrimination tasks with a second conditioning phase, the same λ-value found for the first conditioning was used.

Cochleagram-only Model

The decision model for the cochleagram-only model was exactly the same as for the other full models except that the input to the decision model was the cochleagram rather than the encoding model hidden unit activity. Hence the decision model is given by:

$$p = \mathrm{h}\left(b_o + \sum_{f=1}^{F} \sum_{t=1}^{T} w_{ft} s_f[t] \right)$$

where w_{ft} is the weight from frequency channel f onto the sole output unit m at time step t, and $s_f[t]$ is the value of the cochleagram for frequency channel f at time step t. This model used the same cost function, training procedure, and hyperparameter selection procedure as the decision models that used the encoding model activity as input.

Modelling implementation

All code was implemented in MatLab and Python and run on NVIDIA GeForce GTX-1080 GPUs.

Figure legend

Fig. 1 Discrimination and generalization of frequency modulated sweeps (FMs) differing in two dimensions.

a, 4 pairs of spectrograms of FM sweeps differing in frequency range, direction, rate or duration. **b,** Spectrograms of stimulus sets varying along frequency range and sweep direction. The safe sound was a 9-18 kHz upward FM sweep (blue rectangle) and the unsafe sound was a 6-3 kHz downward FM sweep (red rectangle). **c,** Possible generalization gradients: generalisation occurring along sweep direction (left), frequency range (middle) or both dimensions (right). **d,** Schematic representation of the Audiobox. **e,** Schema of a safe/novel (left) and unsafe (right) visit. Nose-poking was followed by access to water (safe, left) or an air-puff (conditioning, right). **f,Top:** Experimental design. The task consists of three phases: the safe only phase with 100% safe visits (white), the conditioning phase with increasing probability of unsafe visits (red), and the generalization testing phase in which novel sounds were introduced (grey). **Bottom:** Mean daily response for mice trained with the combination of frequency range and sweep directions shown in **b.** Response was expressed as the fraction of visits without nose-pokes for different types of stimulus. Error bars represent standard error.

Fig. 2 The frequency range of a FM sweep dominates over its direction

a, Top: Spectrograms of the stimulus set used in the task. The animals were trained to discriminate the safe sound (blue ringed) and the unsafe sound (red ringed) that differed in frequency range and sweep direction. They were then tested on novel stimuli (the other six sounds). **Middle:** Generalization gradients showing the average response to each stimulus. Error bars represent standard error. **Bottom:** Display of the behavioral data as a heatmap. **b,** Similar to **a.**, for mice trained to discriminate FM sweeps with partially overlapping frequency range. **c,** Similar to **a.**, for mice trained to discriminate FMs with frequency-dependent preferred direction. Sound duration used for a-c. was 20 ms, and the rate of frequency modulation was 50 octaves/s. **d,** Similar to **c.**, for mice trained to discriminate FMs with duration of 40 ms and frequency modulation at 25 octave/s. **e,** The responses of the models to the safe (blue ringed) and unsafe sounds (red ringed) on which they were trained, and to the novel sounds. Results shown for the prediction (top row), compression (middle row), and cochleagram-only models (bottom row). Columns correspond to the task and behavioural data above in **a-d.**. **f,** Schematic of the modelling of the behavioural responses. The model has three stages: 1) The cochlear model converts the sound waveform to a spectrotemporal representation (a cochleagram). 2) The encoding model processes the cochleagram into a new representation, reflective of midbrain or cortex. The encoding model is pre-trained to predict or compress sounds, using diverse natural sounds for this pre-training. 3) The decision model maps from the encoding representation to the behavioural response (the no-nose poke probability). 3) The decision model is trained separately for each encoding model and for each task to predict the safe and unsafe stimuli. Once trained, the decision model is tested using the novel sounds. **g,** The encoding model used in the prediction model. This encoding model embodies temporal prediction and is a recurrent neural network trained to predict the immediate future of natural sound cochleagrams from their recent past. **h,** Example spectrotemporal receptive fields of prediction model units after being trained with natural sounds. Red – excitatory, blue – inhibitory.

Fig.3 Bi-dimensional generalization reveals hierarchical perception of FM sweep in mice.

a, Top: Spectrograms of the stimulus set used in the task. The safe sound (blue), is a 100 octave/s upward FM sweep, and the unsafe sound (red), is a 50 octave/s downward FM sweep. All FM sweeps were centered at 8 kHz. **Middle:** Generalization gradients showing the average response to each stimulus. Error bars represent standard error. **Bottom:** Display of the same behavioral data as in **middle**, but as a colour image arranged as in **top**. **b,** Similar to **a.**, for the combination of sweep direction and duration dimensions. Average generalization gradients (middle and bottom) revealed an effect on behavior of both dimensions. **c,** Similar to **a.**, for the combination of frequency range and duration dimensions. Average generalization gradients (middle and bottom) showed that both dimensions controlled the animals' behavior. **d,** The responses of the prediction (top row), compression (middle row), and cochleagram-only models (bottom row). Columns are different tasks, corresponding to **a-c.**.

Fig.4 Generalization along the non-preferred dimension was localised

a, Top, Spectrograms of FM sweeps used in the task. The safe sound (blue) is a 5 to 10 kHz upward FM sweep, and the unsafe sound (red), a 10 to 5 kHz downward FM sweep. **a, Middle,** Generalization gradients showing percentage of visits without nose pokes for each sound stimuli. Error bar represents standard error. **a,Bottom,** Display of generalisation gradients as a colour image. **b.** Same as **a.** for generalization after the 2nd conditioning during which the already tested downward FM sound (8-4 kHz downward FM sweep) was also conditioned. **c,** The responses of the prediction (top row), compression (middle row), and cochleagram-only models (bottom row). Columns are different tasks, corresponding to **a-b.**

Fig. 5 Behavioral control of sound dimensions is plastic and amenable to training

a, Left, Generalization gradients after initial training (1st conditioning) with safe (blue square) and unsafe (red) sounds combined differences in frequency range and sweep direction. **a, Middle,** Generalization gradient after subsequent conditioning to the non-dominant dimension, sweep direction (the 2nd conditioning, half the mice). **a, Right**, d' for the last 50 unsafe visits of the 1st and the 2nd conditioning for individual mice. **a,** Mice initially selectively generalised along frequency range (left), and were worst at direction discrimination (middle). **b,** Same as **a.** for mice initially trained with sweep direction and

rate (left), and subsequently trained to discriminate the non-dominant dimension, rate (middle). **c-d, Left,** Generalization gradients after the 1st conditioning with direction and duration. Mice initially generalised along both direction and duration dimensions. **c-d, Right,** Shift in generalization pattern after the 2nd conditioning to discriminate either direction (**c**) or duration (**d**). Responses to the safe and unsafe sound are marked as blue and red respectively. **e-f,** The responses of the prediction (top row), compression (middle row), and cochleagram-only models (bottom row). Columns are different tasks, corresponding to **c-d.**

Fig. 6 Bidirectional generalization of periodic sound

a, Top, Spectrograms of FM sweeps used in the task. **a, Bottom,** Generalization gradients plotting the response proportion to each of the stimuli presented in the task for the combination of tone frequency and AM sweep rate dimensions. **b,** Same as **a.** for the combination of tone frequency and repetition rate dimensions. For each figure, the safe and unsafe sounds are marked in blue and red, respectively. **c-d,** Comparison between behavioral data and responses by the prediction, the compression and the cochleagram-only models, for **a-b** respectively.

Fig. 7 Summary and Quantitative comparison of models

a, Bar plot shows the absolute value of the correlation coefficient between each dimension's tested variable space and performance of each mice. Star and circle represent the absolute value of correlation coefficient between dimensions and prediction by prediction model and compression model, respectively. Error bars represent standard error. **b,** Scheme shows the perceptual hierarchy inferred by the results. **c,** For each task, the mean squared error between the mouse responses to the novel sounds and the responses of the prediction, compression, cochleagram-only models, and prediction model trained with scrambled frequency channels. **d,** For each model type, the mean squared error from **c** averaged over all 17 tasks.

References:

1. Plato. *Phaedrus, section 265d-266a*. *Cambridge, MA: Harvard University Press* (1925).
2. Ehret, G. Frequency and intensity difference limens and nonlinearities in the ear of the housemouse (Mus musculus). *J. Comp. Physiol.* **102**, 321–336 (1975).
3. Fay, R. R. Auditory frequency discrimination in vertebrates. *J. Acoust. Soc. Am.* **56**, 206–9 (1974).
4. Kurt, S. & Ehret, G. Auditory discrimination learning and knowledge transfer in mice depends on task difficulty. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 8481–5 (2010).
5. Rohrbaugh, M., Brennan, J. F. & Riccio, D. C. Control of two-way shuttle avoidance in rats by auditory frequency and intensity. *J. Comp. Physiol. Psychol.* **75**, 324–30 (1971).
6. Syka, J., Rybalko, N., Brožek, G. & Jilek, M. Auditory frequency and intensity discrimination in pigmented rats. *Hear. Res.* **100**, 107–113 (1996).
7. Kobrina, A., Toal, K. L. & Dent, M. L. Intensity difference limens in adult CBA/CaJ mice (Mus musculus). *Behav. Processes* **148**, 46–48 (2018).
8. Terman, M. Discrinimation of auditory intensities by rats 1. *J. Exp. Anal. Behav.* **13**, 145–160 (1970).
9. Mercado, E., Orduña, I. & Nowak, J. M. Auditory Categorization of Complex Sounds by Rats (Rattus norvegicus). *J. Comp. Psychol.* **119**, 90–98 (2005).
10. Ohl, F. W., Scheich, H. & Freeman, W. J. Change in pattern of ongoing cortical activity with auditory category learning. *Nature* **412**, 733–736 (2001).
11. Rybalko, N., Šuta, D., Nwabueze-Ogbo, F. & Syka, J. Effect of auditory cortex lesions on the discrimination of frequency-modulated tones in rats. *Eur. J. Neurosci.* **23**, 1614–1622 (2006).
12. Wetzel, W., Wagner, T., Ohl, F. W. & Scheich, H. Categorical discrimination of direction in frequency-modulated tones by Mongolian gerbils. *Behav. Brain Res.* **91**, 29–39 (1998).
13. Schulze, H. & Scheich, H. Discrimination learning of amplitude modulated tones in Mongolian gerbils. *Neurosci. Lett.* **261**, 13–16 (1999).
14. Gaese, B. H., King, I., Felsheim, C., Ostwald, J. & von der Behrens, W. Discrimination of direction in fast frequency-modulated tones by rats. *J. Assoc. Res.*

Otolaryngol. **7**, 48–58 (2006).

15. Florentine, M. Intensity discrimination as a function of level and frequency and its relation to high frequency hearing. *J. Acoust. Soc. Am.* **74**, 1375–1379 (1983).
16. Heffner, H. & Masterton, B. Hearing in Glires: Domestic rabbit, cotton rat, feral house mouse, and kangaroo rat. *J. Acoust. Soc. Am.* **68**, 1584–1599 (1980).
17. Nabelek, I. & Hirsh, I. J. On the Discrimination of Frequency Transitions. *J. Acoust. Soc. Am.* **45**, 1510–1519 (1969).
18. Schouten, M. E. H. Identification and discrimination of sweep tones. *Percept. Psychophys.* **37**, 369–376 (1985).
19. Vyazovska, O. V. The effect of dimensional reinforcement prediction on discrimination of compound visual stimuli by pigeons. *Anim. Cogn.* 1–10 (2021). doi:10.1007/s10071-021-01526-z
20. Ahissar, M. & Hochstein, S. The reverse hierarchy theory of visual perceptual learning. *Trends Cogn. Sci.* **8**, 457–464 (2004).
21. Chalk, M., Marre, O. & Tkačik, G. Toward a unified theory of efficient, predictive, and sparse coding. *Proc. Natl. Acad. Sci.* **115**, 186–191 (2018).
22. Salisbury, J. M., Palmer, S. E., Salisbury, J. M. & Palmer, S. E. Optimal Prediction in the Retina and Natural Motion Statistics. *J. Stat. Phys*. **162**, 1309–1323 (2016).
23. Barlow, H. Sensory mechanisms, the reduction of redundancy, and intelligence. *Mech. thought Process.* (1959).
24. Attneave, F. Some informational aspects of visual perception. *Psychol. Rev.* **61**, 183–193 (1954).
25. Olshausen, B. A. & Field, D. J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **381**, 607–609 (1996).
26. Olshausen, B. A. & Field, D. J. Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Res*. **37**, 3311–3325 (1997).
27. Rao, R. P. N. & Ballard, D. H. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* **2**, 79–87 (1999).
28. Bialek, W., Nemenman, I. & Tishby, N. Predictability, Complexity, and Learning. *Neural Comput.* **13**, 2409–2463 (2001).
29. Singer, Y., Willmore, B. D. B., King, A. J. & Harper, N. S. Hierarchical temporal prediction captures motion processing from retina to higher visual cortex. *bioRxiv* 575464 (2019). doi:10.1101/575464

30. Singer, Y. *et al.* Sensory cortex is optimized for prediction of future input. *Elife* **7**, (2018).
31. Hage, S. R., Ehret, G. G., Nter Ehret, G. È. & Ehret, G. G. Mapping responses to frequency sweeps and tones in the inferior colliculus of house mice. *Eur. J. Neurosci.* **18**, 2301–2312 (2003).
32. Heil, P., Langner, G. & Scheich, H. Processing of frequency-modulated stimuli in the chick auditory cortex analogue: evidence for topographic representations and possible mechanisms of rate and directional sensitivity. *J. Comp. Physiol. A.* **171**, 583–600 (1992).
33. Zhang, L. I., Tan, A. Y. Y., Schreiner, C. E. & Merzenich, M. M. Topography and synaptic shaping of direction selectivity in primary auditory cortex. *Nature* **424**, 201–205 (2003).
34. Ehret, G. Categorical perception of mouse-pup ultrasounds in the temporal domain. *Anim. Behav.* **43**, 409–416 (1992).
35. Rahman, M., Willmore, B. D. B., King, A. J. & Harper, N. S. Simple transformations capture auditory input to cortex. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 28442–28451 (2020).
36. Rees, A. & Møller, A. R. Responses of neurons in the inferior colliculus of the rat to AM and FM tones. *Hear. Res.* **10**, 301–330 (1983).
37. Goodfellow, I., Bengio, Y. & Courville, A. Autoencoders. in *Deep Learning* 499–523 (MIT Press, 2016).
38. Lewicki, M. S. Efficient coding of natural sounds. *Nat. Neurosci.* **5**, 356–363 (2002).
39. Smith, E. C. & Lewicki, M. S. Efficient auditory coding. *Nature* **439**, 978–982 (2006).
40. Hage, S. R. & Ehret, G. Mapping responses to frequency sweeps and tones in the inferior colliculus of house mice. *Eur. J. Neurosci.* **18**, 2301–2312 (2003).
41. Kuo, R. I. & Wu, G. K. The Generation of Direction Selectivity in the Auditory System. *Neuron* **73**, 1016–1027 (2012).
42. Rosen, S. Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **336**, 367–73 (1992).
43. Hickok, G. & Poeppel, D. The cortical organization of speech processing. *Nat. Rev. Neurosci.* **8**, 393–402 (2007).
44. Leaver, A. M. & Rauschecker, J. P. Cortical Representation of Natural Complex

Sounds: Effects of Acoustic Features and Auditory Object Category. *J. Neurosci.* **30**, 7604–7612 (2010).

45. Chen, C., Krueger-Burg, D. & de Hoz, L. Wide sensory filters underlie performance in memory-based discrimination and generalization. *PLoS One* **14**, e0214817 (2019).
46. Kilgard, M. P. & Merzenich, M. M. Distributed representation of spectral and temporal information in rat primary auditory cortex. *Hear. Res.* **134**, 16–28 (1999).
47. Greenwald, A. G., McGhee, D. E. & Schwartz, J. L. Measuring individual differences in implicit cognition: the implicit association test. *J. Pers. Soc. Psychol.* **74**, 1464–80 (1998).
48. Polley, D. B. Perceptual Learning Directs Auditory Cortical Map Reorganization through Top-Down Influences. *J. Neurosci.* **26**, 4970–4982 (2006).
49. Ehret, G. & Riecke, S. Mice and humans perceive multiharmonic communication sounds in the same way. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 479–82 (2002).
50. Geissler, D. B. & Ehret, G. Time-critical integration of formants for perception of communication calls in mice. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 9021–5 (2002).
51. Carlin, M. A. & Elhilali, M. Sustained Firing of Model Central Auditory Neurons Yields a Discriminative Spectro-temporal Representation for Natural Sounds. *PLoS Comput. Biol.* **9**, e1002982 (2013).
52. Salisbury, J. M. & Palmer, S. E. Optimal Prediction in the Retina and Natural Motion Statistics. *J. Stat. Phys.* **162**, 1309–1323 (2016).
53. Berkes, P. & Wiskott, L. Slow feature analysis yields a rich repertoire of complex cell properties. *J. Vis.* **5**, 9–9 (2005).
54. Stone, J. V & Harper, N. Temporal constraints on visual learning: a computational model. *Perception* **28**, 1089–104 (1999).
55. Klein, D. J., König, P. & Körding, K. P. Sparse Spectrotemporal Coding of Sounds. *EURASIP J. Adv. Signal Process.* **2003**, 902061 (2003).
56. Carlson, N. L., Ming, V. L. & DeWeese, M. R. Sparse Codes for Speech Predict Spectrotemporal Receptive Fields in the Inferior Colliculus. *PLoS Comput. Biol.* **8**, e1002594 (2012).
57. Zhao, L. & Zhaoping, L. Understanding auditory spectro-temporal receptive fields and their changes with input statistics by efficient coding principles. *PLoS Comput. Biol.* **7**, (2011).
58. Kozlov, A. S. & Gentner, T. Q. Central auditory neurons have composite receptive

fields. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 1441–6 (2016).

59. Cusack, R. & Carlyon, R. Auditory perceptual organization inside and outside the laboratory. in *Ecological Psychoacoustics* (ed. G, N. J.) 15–48 (Elsevier, 2004).
60. Lotter, W., Kreiman, G. & Cox, D. A neural network trained for prediction mimics diverse features of biological neurons and perception. *Nat. Mach. Intell.* **2**, 210–219 (2020).
61. Chechik, G. *et al.* Reduction of Information Redundancy in the Ascending Auditory Pathway. *Neuron* **51**, 359–368 (2006).
62. Grill-Spector, K. & Malach, R. THE HUMAN VISUAL CORTEX. *Annu. Rev. Neurosci.* **27**, 649–677 (2004).
63. Hegdé, J. Time course of visual perception: Coarse-to-fine processing and beyond. *Prog. Neurobiol.* **84**, 405–439 (2008).
64. Hochstein, S. & Ahissar, M. View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron* **36**, 791–804 (2002).
65. Markov, N. T. & Kennedy, H. The importance of being hierarchical. *Curr. Opin. Neurobiol.* **23**, 187–194 (2013).
66. Ahissar, M. & Hochstein, S. Task difficulty and the specificity of perceptual learning. *Nature* **387**, 401–406 (1997).
67. Brugge, J. F. An Overview of Central Auditory Processing. in 1–33 (Springer, New York, NY, 1992). doi:10.1007/978-1-4612-2838-7_1
68. Brand, A., Urban, A. & Grothe, B. *Duration Tuning in the Mouse Auditory Midbrain.* (2000).
69. Langner, G., Albert, M. & Briede, T. Temporal and spatial coding of periodicity information in the inferior colliculus of awake chinchilla (Chinchilla laniger). *Hear. Res.* **168**, 110–130 (2002).
70. Clayton, K. K. *et al.* Auditory Corticothalamic Neurons Are Recruited by Motor Preparatory Inputs. *Curr. Biol.* **31**, 310-321.e5 (2021).
71. de Hoz, L. & Nelken, I. Frequency Tuning in the Behaving Mouse: Different Bandwidths for Discrimination and Generalization. *PLoS One* **9**, e91676 (2014).
72. Morrison, G. S., Rose, P. & Zhang, C. Protocol for the collection of databases of recordings for forensic-voice-comparison research and practice. *Aust. J. Forensic Sci.* **44**, 155–167 (2012).
73. MacWhinney, B. *The CHILDES Project: Tools for analyzing talk. CHILDES* (Lawrence Erlbaum Associates, 2000).

74. Beckman, M. & Edwards, J. PhonBank Japanese Paidologos Corpus. (2010).

75. Parsons, C. E., Young, K. S., Craske, M. G., Stein, A. L. & Kringelbach, M. L. Introducing the oxford vocal (OxVoc) sounds database: A validated set of non-acted affective sounds from human infants, adults, and domestic animals. *Front. Psychol.* **5**, 1–10 (2014).

76. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. 1–15 (2014). doi:http://doi.acm.org.ezproxy.lib.ucf.edu/10.1145/1830483.1830503

77. Glorot, X. & Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. *J. Mach. Learn. Res.* **9**, 249–256 (2010).

78. Srebro, N., Srebro, N. & Shraibman, A. Rank, trace-norm and max-norm. *Proc. 18th Annu. Conf. Learn. Theory* **14**, 545--560 (2005).

79. Willmore, B. D. B., Schoppe, O., Andrew, K. J., Schnupp, J. W. H. & Harper, N. S. Incorporating Midbrain Adaptation to Mean Sound Level Improves Models of Auditory Cortical Processing. (2016). doi:10.1523/JNEUROSCI.2441-15.2016

80. Willmore, B. Benlib-py. *GitHub* (2020). Available at: https://github.com/ben-willmore/benlib-py.

[i] Specifically, we use the following recordings from the Macauley Library: ML3429, ML3891, ML53181, ML55348, ML55350, ML59291, ML61494, ML62970, ML67895, ML70539, ML72995, ML80731, ML100723, ML100797, ML100857, ML111595, ML116302, ML116303, ML126289, ML129248, ML130909, ML132511, ML132526, ML136504, Ml164696, ML171372, ML191165, ML191178, ML191290, ML197064, ML199078, ML201215, ML205481, ML205774, ML207517, ML206448, ML207181, ML210645, ML213055, ML213356, ML217313, ML217490, ML220049, ML233387, ML516692, ML527184 and ML527292

Figure 1

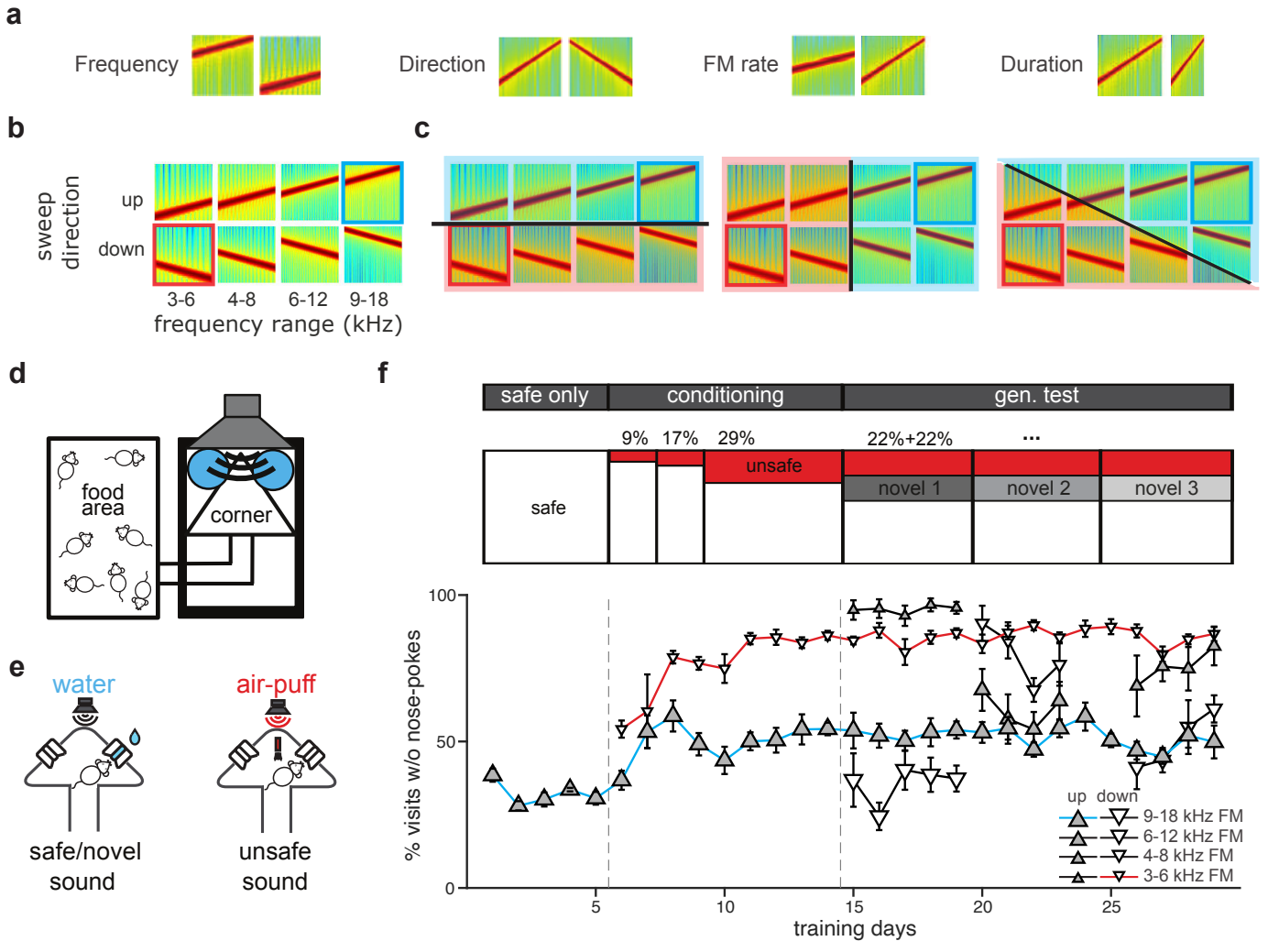

a
Frequency
Direction
FM rate
Duration
b
sweep direction
up
down
3-6
4-8
6-12
9-18
frequency range (kHz)
c
d
food area
corner
e
water
air-puff
safe/novel sound
unsafe sound
f
safe only
conditioning
gen. test
9%
17%
29%
22%+22%
...
safe
unsafe
novel 1
novel 2
novel 3
% visits w/o nose-pokes
100
50
0
5
10
15
20
25
training days
up
down
9-18 kHz FM
6-12 kHz FM
4-8 kHz FM
3-6 kHz FM

Figure 2

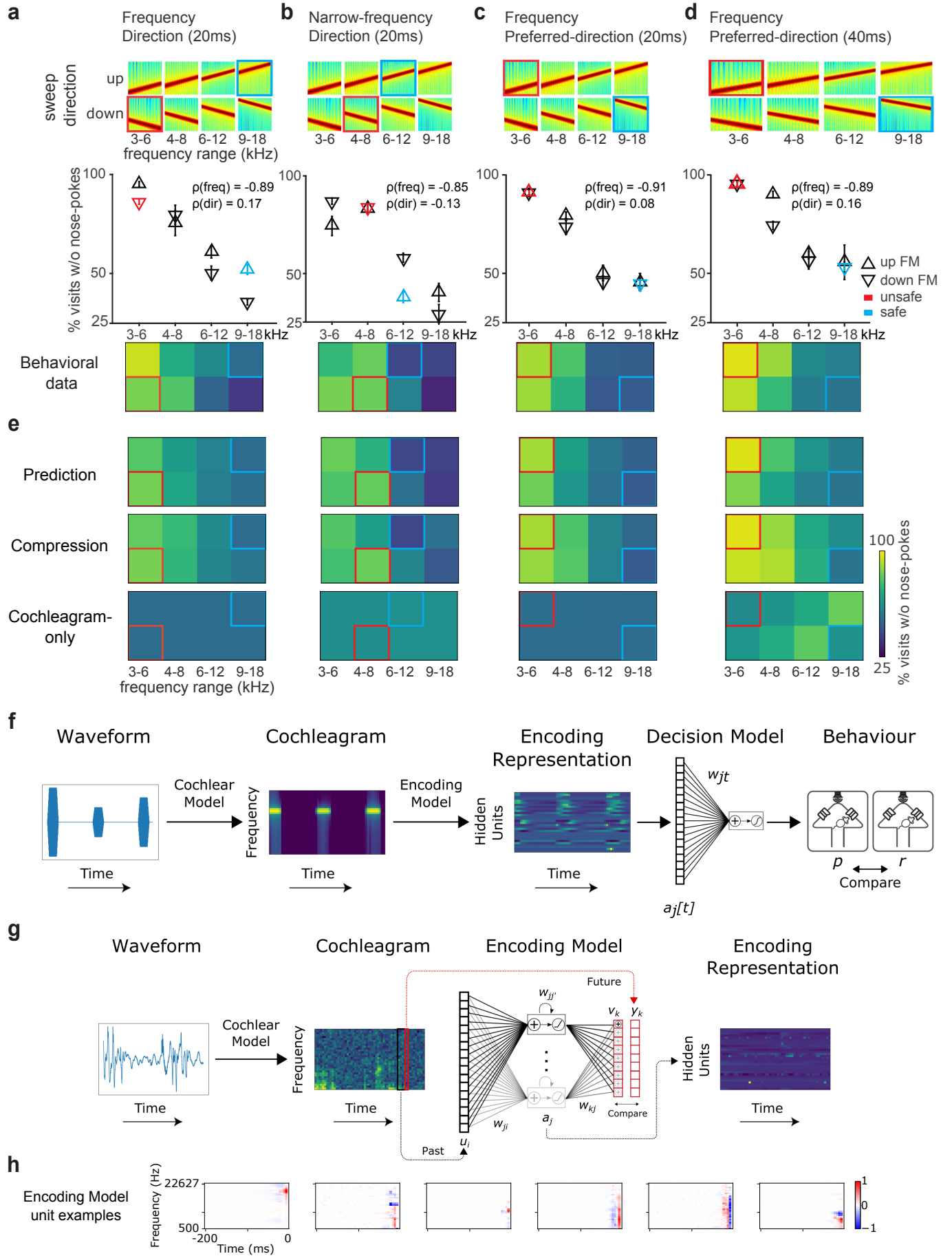

a
Frequency Direction (20ms)
b
Narrow-frequency Direction (20ms)
c
Frequency Preferred-direction (20ms)
d
Frequency Preferred-direction (40ms)
sweep direction
up
down
3-6 4-8 6-12 9-18
frequency range (kHz)
% visits w/o nose-pokes
100
50
25
ρ(freq) = -0.89
ρ(dir) = 0.17
ρ(freq) = -0.85
ρ(dir) = -0.13
ρ(freq) = -0.91
ρ(dir) = 0.08
ρ(freq) = -0.89
ρ(dir) = 0.16
3-6 4-8 6-12 9-18 kHz
up FM
down FM
unsafe
safe
Behavioral data
e
Prediction
Compression
Cochleagram-only
3-6 4-8 6-12 9-18
frequency range (kHz)
100
25
% visits w/o nose-pokes
f
Waveform
Cochleagram
Encoding Representation
Decision Model
Behaviour
Cochlear Model
Frequency
Encoding Model
Hidden Units
Time
w_{jt}
$a_j[t]$
p
r
Compare
g
Waveform
Cochleagram
Encoding Model
Encoding Representation
Cochlear Model
Frequency
Time
Future
Past
$w_{jj'}$
v_k
y_k
w_{kj}
w_{ji}
a_j
u_i
Compare
Hidden Units
h
Encoding Model unit examples
Frequency (Hz)
22627
500
-200
0
Time (ms)
1
0
−1

Figure 3

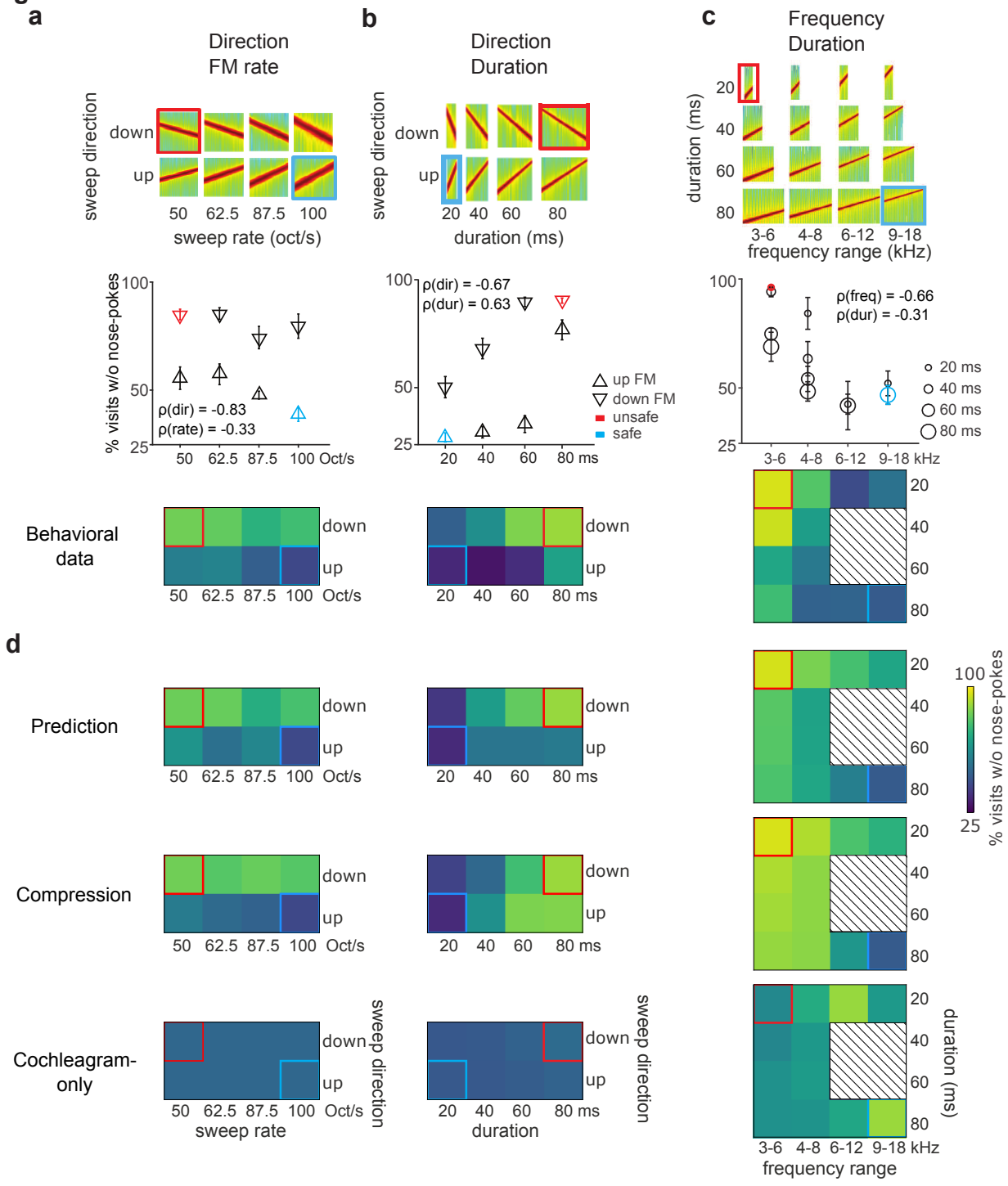

a
Direction
FM rate
sweep direction
down
up
50 62.5 87.5 100
sweep rate (oct/s)
b
Direction
Duration
sweep direction
down
up
20 40 60 80
duration (ms)
c
Frequency
Duration
duration (ms)
20
40
60
80
3-6 4-8 6-12 9-18
frequency range (kHz)
% visits w/o nose-pokes
100
50
25
ρ(dir) = -0.83
ρ(rate) = -0.33
50 62.5 87.5 100 Oct/s
ρ(dir) = -0.67
ρ(dur) = 0.63
20 40 60 80 ms
up FM
down FM
unsafe
safe
ρ(freq) = -0.66
ρ(dur) = -0.31
20 ms
40 ms
60 ms
80 ms
3-6 4-8 6-12 9-18 kHz
Behavioral data
down
up
50 62.5 87.5 100 Oct/s
20 40 60 80 ms
d
Prediction
Compression
Cochleagram-only
sweep direction
sweep rate
duration
frequency range
duration (ms)
% visits w/o nose-pokes

Figure 4

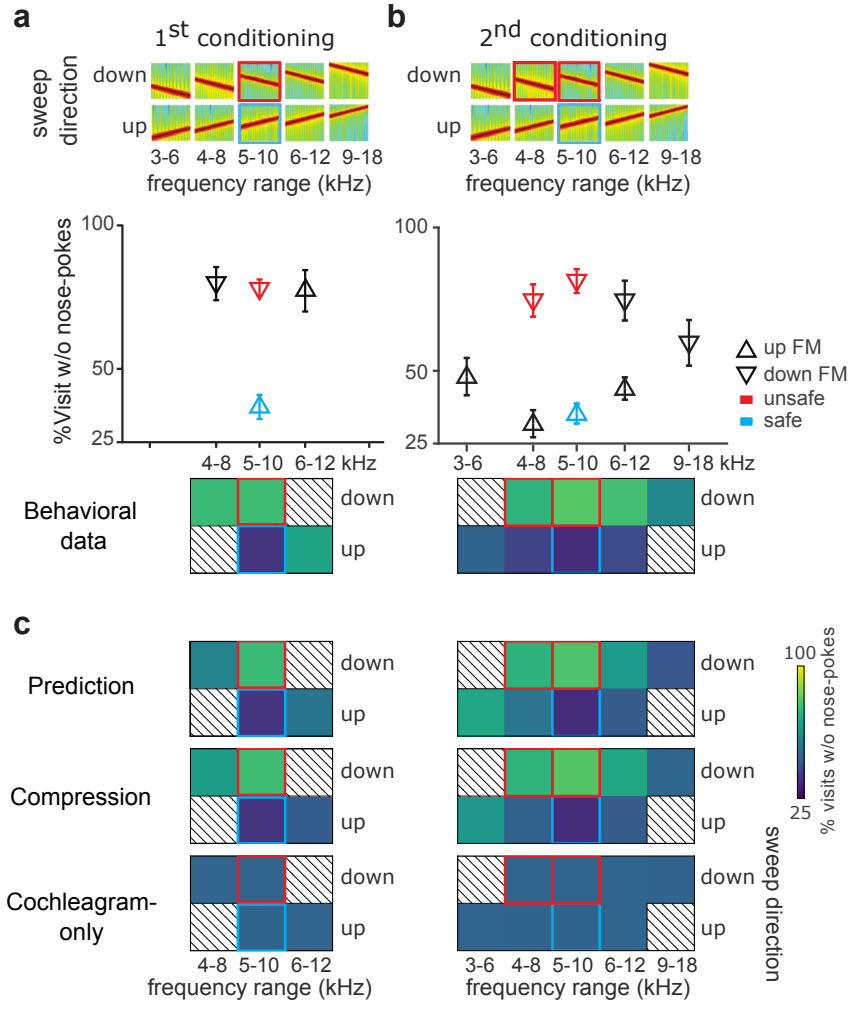

a
1st conditioning
sweep direction
down
up
3-6 4-8 5-10 6-12 9-18
frequency range (kHz)
b
2nd conditioning
down
up
3-6 4-8 5-10 6-12 9-18
frequency range (kHz)
%Visit w/o nose-pokes
100
50
25
4-8 5-10 6-12 kHz
3-6 4-8 5-10 6-12 9-18 kHz
up FM
down FM
unsafe
safe
Behavioral data
down
up
c
Prediction
Compression
Cochleagram-only
4-8 5-10 6-12
frequency range (kHz)
3-6 4-8 5-10 6-12 9-18
frequency range (kHz)
sweep direction
% visits w/o nose-pokes

Figure 5

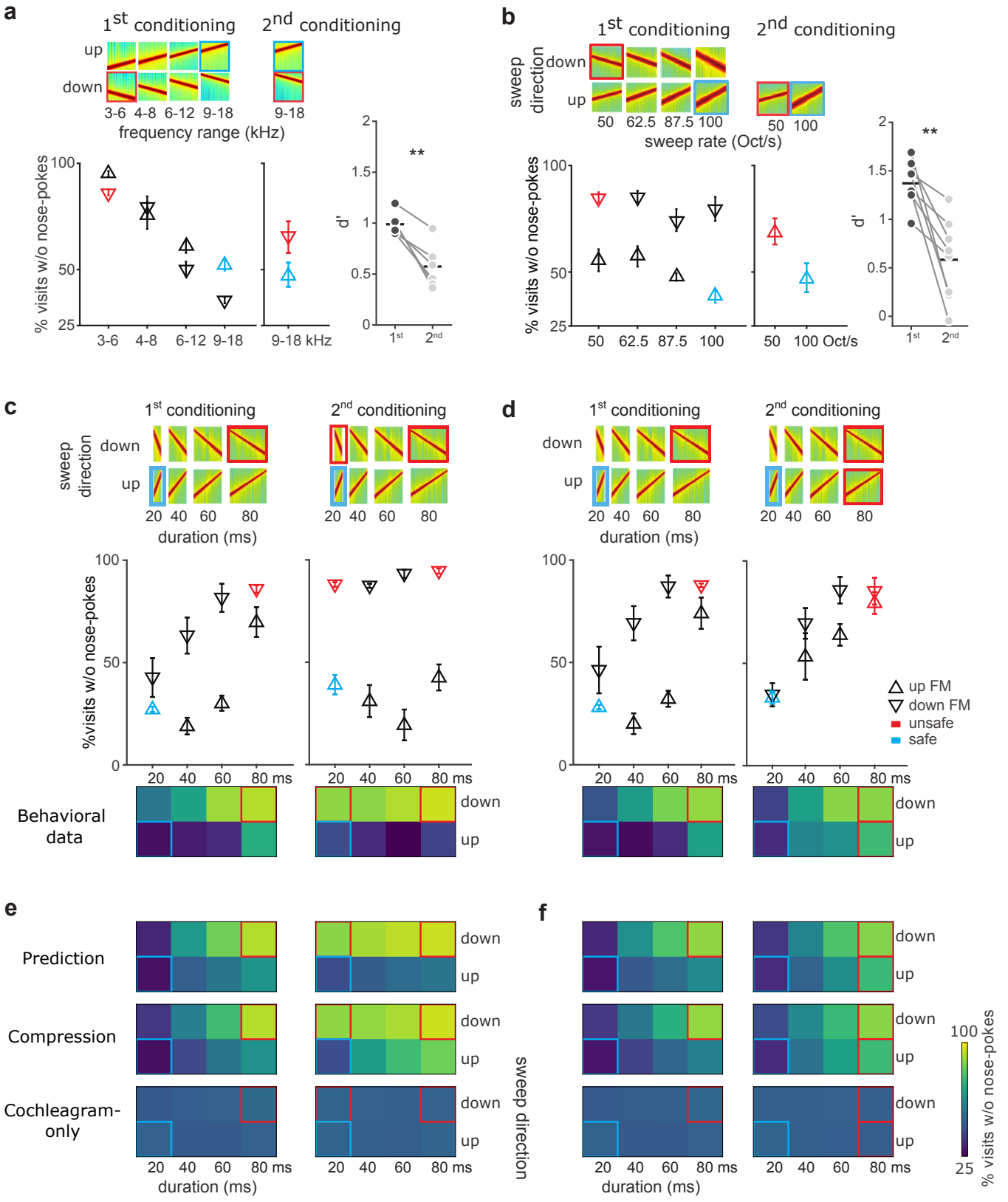

a
1st conditioning
2nd conditioning
up
down
3-6 4-8 6-12 9-18
9-18
frequency range (kHz)
% visits w/o nose-pokes
100
50
25
3-6 4-8 6-12 9-18
9-18 kHz
d'
2
1.5
1
0.5
0
**
1st
2nd
b
1st conditioning
2nd conditioning
sweep direction
down
up
50 62.5 87.5 100
50 100
sweep rate (Oct/s)
% visits w/o nose-pokes
100
50
25
50 62.5 87.5 100
50 100 Oct/s
d'
2
1.5
1
0.5
0
**
1st
2nd
c
1st conditioning
2nd conditioning
sweep direction
down
up
20 40 60 80
duration (ms)
%visits w/o nose-pokes
100
50
0
20 40 60 80 ms
Behavioral data
down
up
d
1st conditioning
2nd conditioning
down
up
20 40 60 80
duration (ms)
100
50
0
20 40 60 80 ms
up FM
down FM
unsafe
safe
down
up
e
Prediction
Compression
Cochleagram-only
down
up
20 40 60 80 ms
duration (ms)
f
sweep direction
down
up
20 40 60 80 ms
duration (ms)
100
25
% visits w/o nose-pokes

Figure 6

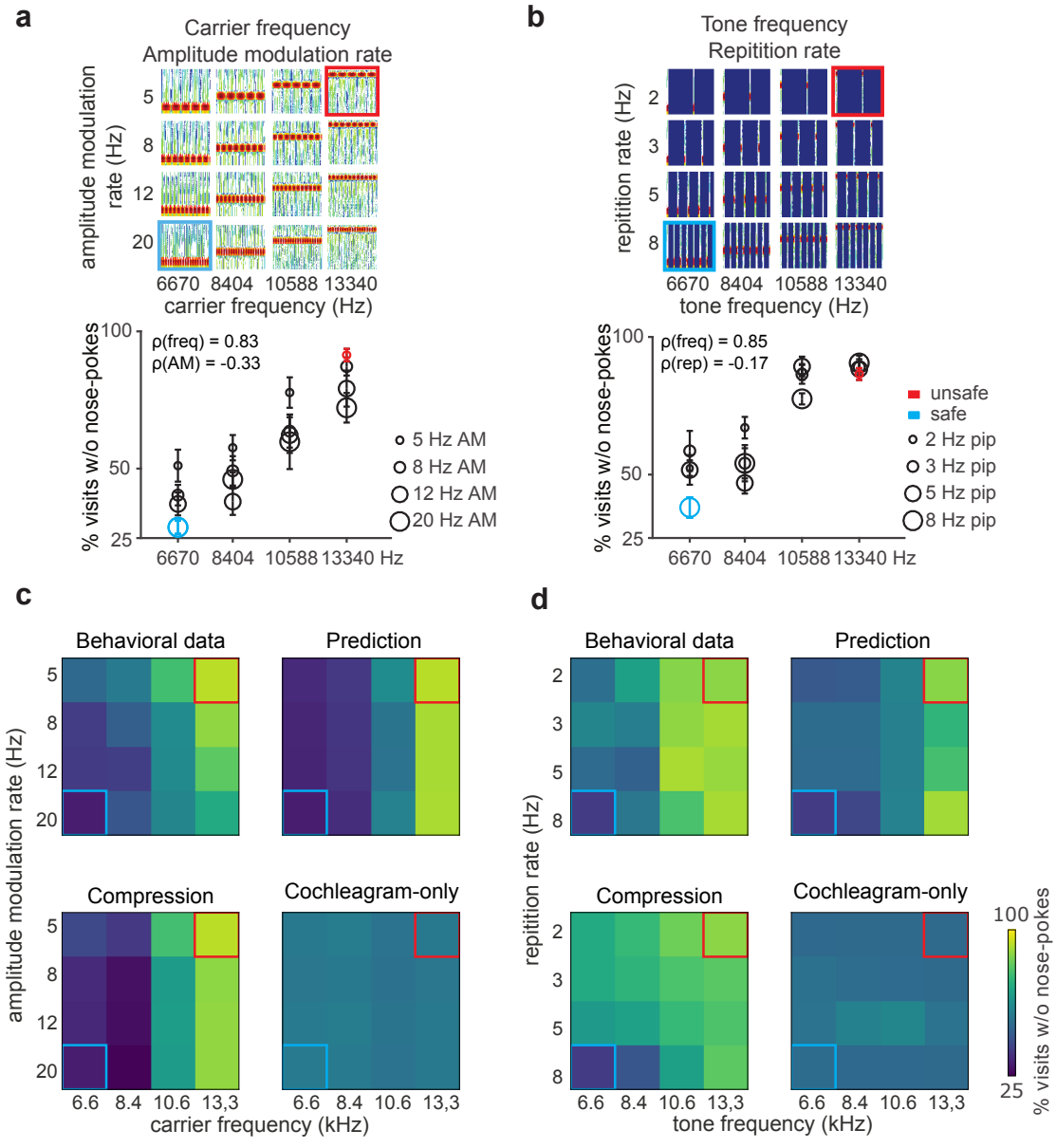
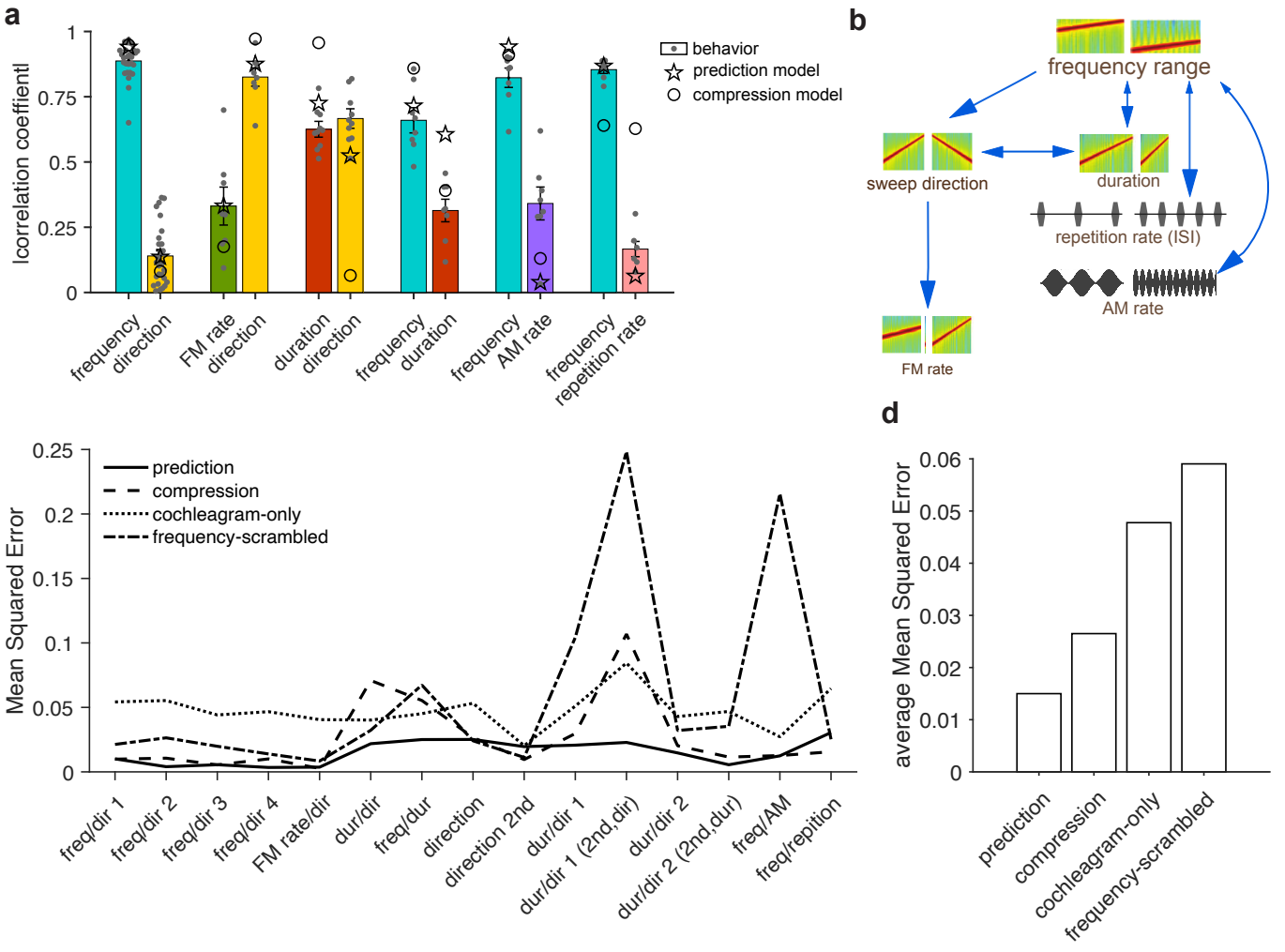

a
Carrier frequency
Amplitude modulation rate
amplitude modulation rate (Hz)
5
8
12
20
6670 8404 10588 13340
carrier frequency (Hz)
% visits w/o nose-pokes
100
50
25
ρ(freq) = 0.83
ρ(AM) = -0.33
5 Hz AM
8 Hz AM
12 Hz AM
20 Hz AM
6670 8404 10588 13340 Hz
b
Tone frequency
Repitition rate
repitition rate (Hz)
2
3
5
8
6670 8404 10588 13340
tone frequency (Hz)
% visits w/o nose-pokes
100
50
25
ρ(freq) = 0.85
ρ(rep) = -0.17
unsafe
safe
2 Hz pip
3 Hz pip
5 Hz pip
8 Hz pip
6670 8404 10588 13340 Hz
c
Behavioral data
Prediction
Compression
Cochleagram-only
amplitude modulation rate (Hz)
5
8
12
20
6.6 8.4 10.6 13,3
carrier frequency (kHz)
d
Behavioral data
Prediction
Compression
Cochleagram-only
repitition rate (Hz)
2
3
5
8
6.6 8.4 10.6 13,3
tone frequency (kHz)
100
25
% visits w/o nose-pokes

Figure 7


a
lcorrelation coeffientl
1
0.75
0.5
0.25
0
behavior
prediction model
compression model
frequency direction
FM rate direction
duration direction
frequency duration
frequency AM rate
frequency repetition rate
b
frequency range
sweep direction
duration
repetition rate (ISI)
AM rate
FM rate
Mean Squared Error
0.25
0.2
0.15
0.1
0.05
0
prediction
compression
cochleagram-only
frequency-scrambled
freq/dir 1
freq/dir 2
freq/dir 3
freq/dir 4
FM rate/dir
dur/dir
freq/dur
direction
direction 2nd
dur/dir 1
dur/dir 1 (2nd,dir)
dur/dir 2
dur/dir 2 (2nd,dur)
freq/AM
freq/repition
d
average Mean Squared Error
0.06
0.05
0.04
0.03
0.02
0.01
0
prediction
compression
cochleagram-only
frequency-scrambled