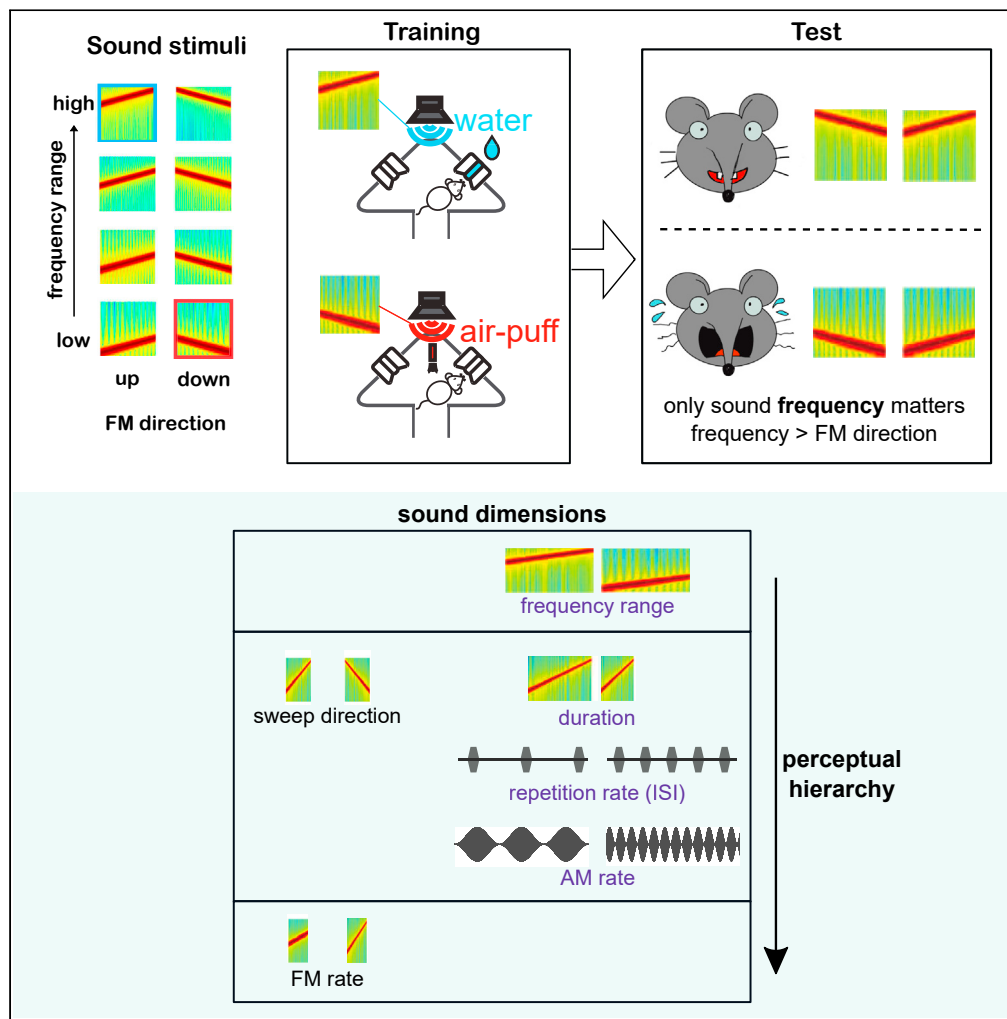


Article

# The perceptual categorization of multidimensional stimuli is hierarchically organized



Chi Chen, Livia de Hoz

livia.dehoz@charite.de

**Highlights**

A sound discrimination task uncovered a perceptual hierarchy of stimulus dimensions

This hierarchy was dominated by the sound's spectral composition

With training mice could learn to use dimensions low in the hierarchy, albeit rigidly

Stimuli are not perceived as a whole but as a combination of their features



## Article

## The perceptual categorization of multidimensional stimuli is hierarchically organized

Chi Chen<sup>1,2,3,4</sup> and Livia de Hoz<sup>1,4,5,6,\*</sup>

## SUMMARY

**As we interact with our surroundings, we encounter the same or similar objects from different perspectives and are compelled to generalize. For example, despite their variety we recognize dog barks as a distinct sound class. While we have some understanding of generalization along a single stimulus dimension (frequency, color), natural stimuli are identifiable by a combination of dimensions. Measuring their interaction is essential to understand perception. Using a 2-dimension discrimination task for mice and frequency or amplitude modulated sounds, we tested untrained generalization across pairs of auditory dimensions in an automatized behavioral paradigm. We uncovered a perceptual hierarchy over the tested dimensions that was dominated by the sound's spectral composition. Stimuli are thus not perceived as a whole, but as a combination of their features, each of which weights differently on the identification of the stimulus according to an established hierarchy, possibly paralleling their differential shaping of neuronal tuning.**

## 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),<sup>1</sup> by which our brains separate the world into objects that fall into different classes e.g., footsteps, motorbikes, human voices, the 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 multidimensional 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. Our auditory system is able to discriminate or categorize along sound frequency,<sup>2–6</sup> intensity,<sup>5–8</sup> frequency modulation direction,<sup>9–12</sup> or amplitude modulation frequency.<sup>4,13</sup> Yet stimulus discrimination along a single dimension can be influenced by stimulus features outside that dimension, as has been demonstrated in auditory processing.<sup>14–18</sup> While in the visual modality, dimension integration has been studied by training animals such as pigeons to use dimension combinations,<sup>19</sup> or by asking human subjects to identify pop-up stimuli of different level of complexity,<sup>20</sup> little is known about how multiple dimensions within any given modality are naturally integrated when classifying environmental stimuli.<sup>21,22</sup> To understand the interactions between stimulus dimensions, as well as the hierarchy of importance of the dimensions, we focused on the auditory domain and used sounds that varied in ranges and dimensions known to be of relevance to mice. 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 whether dimensions are perceptually categorized and which

<sup>1</sup>Department of Neurogenetics, Max Planck Institute for Experimental Medicine, Göttingen, Germany

<sup>2</sup>International Max Planck Research School for Neurosciences, Göttingen, Germany

<sup>3</sup>Göttingen Graduate School of Neurosciences and Molecular Biosciences, Göttingen, Germany

<sup>4</sup>Neuroscience Research Center, Charité Medical University, Berlin, Germany

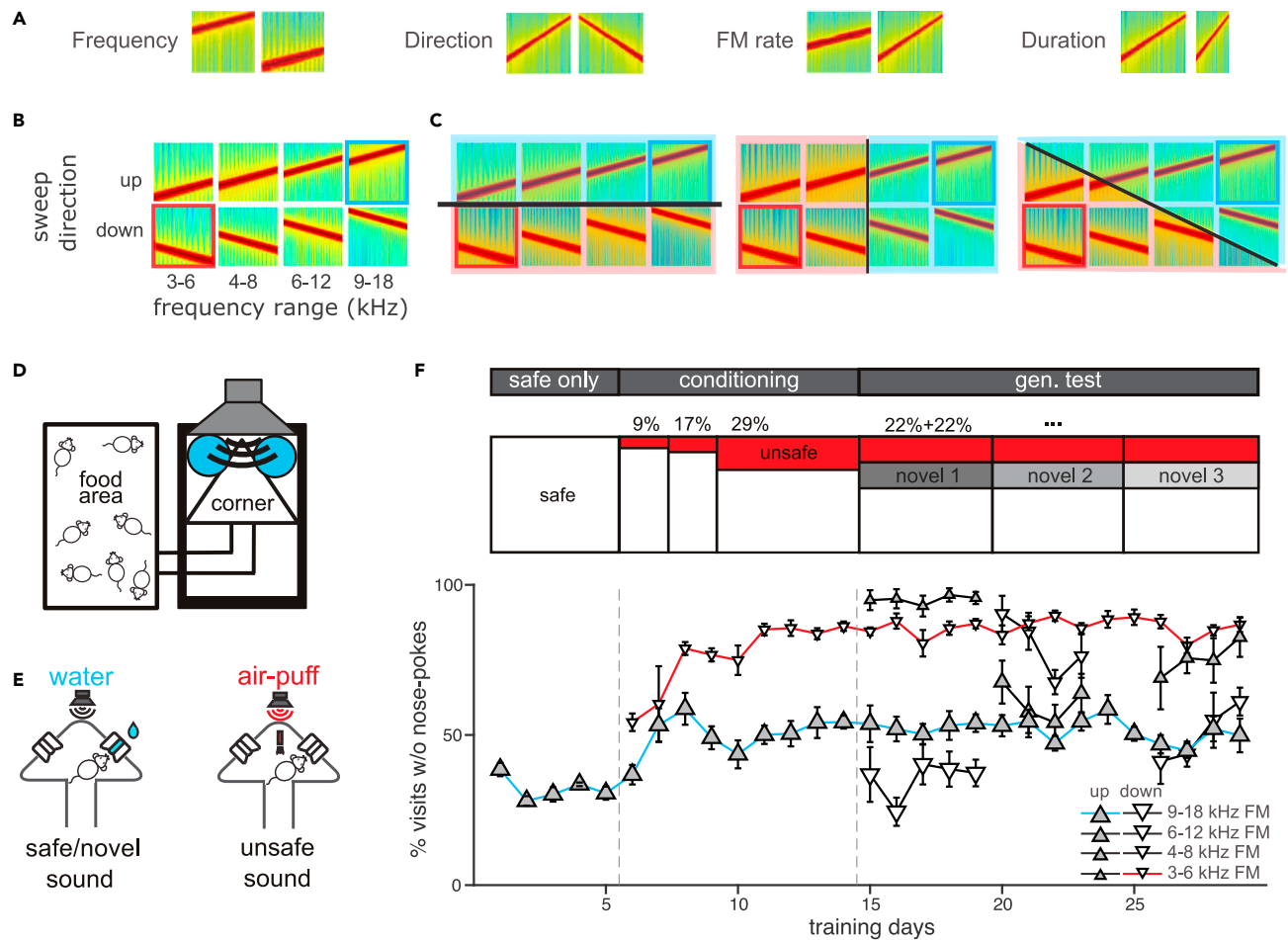
<sup>5</sup>Bernstein Center for Computational Neuroscience, Berlin, Germany

<sup>6</sup>Lead contact

\*Correspondence: livia.dehoz@charite.de

<https://doi.org/10.1016/j.isci.2023.106941>





**Figure 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: generalization 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 (gray). 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 stimuli. Error bars represent standard error.

dimensions tend to dominate in the discrimination of sounds (e.g., frequency or sweep direction). We used the Audiobox, a sister of the Intellicage, an automatic and naturalistic behavioral apparatus. It consists of an enriched multicompartmental environment in which mice live in groups for days at a time. Since neither food nor water is restricted, performance is *ad libitum*, mainly during the nocturnal phase. This apparatus thus, based on volitional training, permits animals to freely move and make choices about task engagement at their own will, making it ideal for testing of spontaneous generalization.

## 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 (frequency and amplitude), and duration (Figure 1A). For example, for sounds differing in frequency range and sweep direction, the safe sound could

be a 9–18 kHz upward-sweep (Figure 1B, blue frame), and the unsafe sound, a 6–3 kHz downward-sweep (Figure 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 (Figure 1B, other sounds in the 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 (Figure 1C).

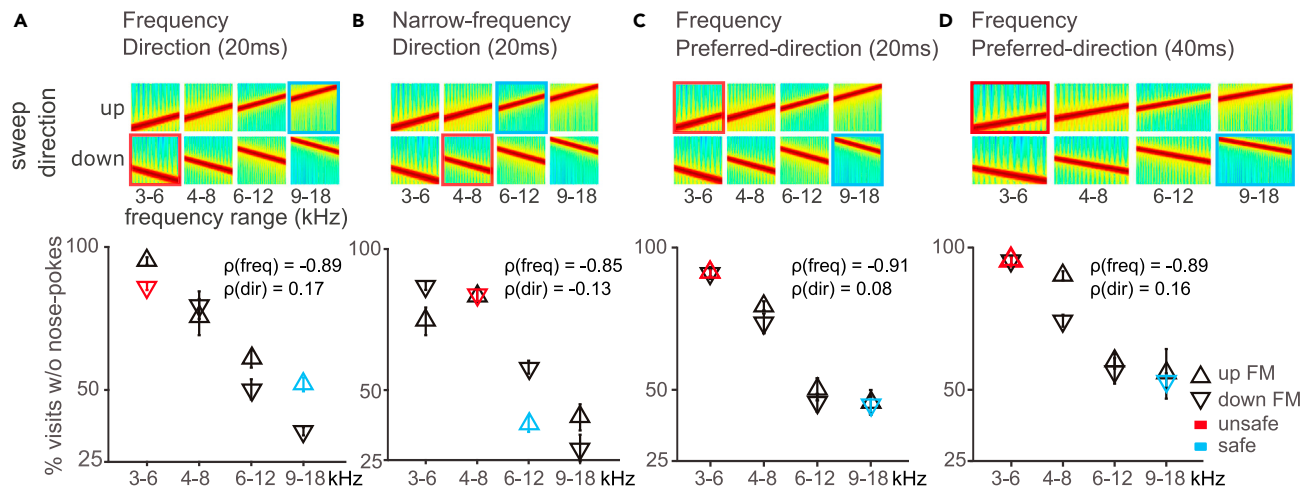
Each dimension pair was tested in a different group of mice. For all 4 dimensions, we chose a span of values known to elicit discriminative responses in neurons in the auditory system. The mice were trained in the Audiobox (Figures 1D, S1A, and S1B), an apparatus in which mice live in groups for the 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 (Figure 1E, left), whereas when the “unsafe sound” was played, nose-poking was followed by an aversive air puff and no access to water (Figure 1E, right; Figure S1C). The task, therefore, resembled a go/no-go paradigm. Importantly though, the time between a trial (visit) and the next, was defined by the mouse and was often several minutes,<sup>23,24</sup> such that decisions are, by necessity, based on long-term memory and not on the immediate comparison between two trials as is often the case in go/no-go paradigms. During the generalization phase, two novel sounds (appearing in 11% of the chamber visits each) were introduced and changed every 2 to 4 days (Figure 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 (Figure 1E, left). We measured a mouse’s response to a given sound as its level of nose-poke avoidance, i.e., the probability that it will not poke its nose in that visit (the “response”). This probability of not nose-poking was measured against a backdrop of about 100 visits per day per mouse (Figure S2A). The nose-poke avoidance in turn correlated with the length of the visit and was shorter for visit types in which the mouse avoided nose-poking (Figure S2B).<sup>23</sup>

### 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 (9–18 kHz upward vs. 6 to 3 kHz downward sweeps,  $n = 9$ ; Figure 1B). Each day, discrimination was measured as the difference between the response to the safe sound (Figure 1F bottom, blue line) and that to the unsafe sound (Figure 1F bottom, red line). Mice that showed stable discrimination performance for at least three consecutive days (See STAR Methods), were included in the analysis and then tested with novel sounds varying along both dimensions (Figure 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 (Figure 1F bottom, black lines).

The mean response for this first set of sounds, which varied in frequency range and modulation direction (safe, unsafe, and novel, Figure 2A, top), what we term the “generalization pattern,” is shown in the form of a performance plot of the average response to sounds presented during the generalization phase (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,  $\rho(\text{freq})$ , was  $-0.89$ ) and not its sweep direction ( $\rho(\text{dir}) = 0.17$ ). 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). Modeling animals’ nose-poking probability with a linear-mixed model with mouse as random effect, revealed a significant main effect of frequency range ( $p < 0.0001$ ), but no effect of sweep direction ( $p = 0.81$ ) or frequency  $\times$  direction interaction ( $p = 0.14$ ).

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 (6–12 kHz upward vs. 8 to 4 kHz



**Figure 2. The frequency range of an FM sweep dominates over its direction**

(A) Top: Spectrograms of the stimulus set used in the task. The animals were trained to discriminate between 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). Bottom: Generalization gradients showing the average response to each stimulus during the generalization phase.

(B) Similar to (A) for mice trained to discriminate FM sweeps with partially overlapping frequency range.

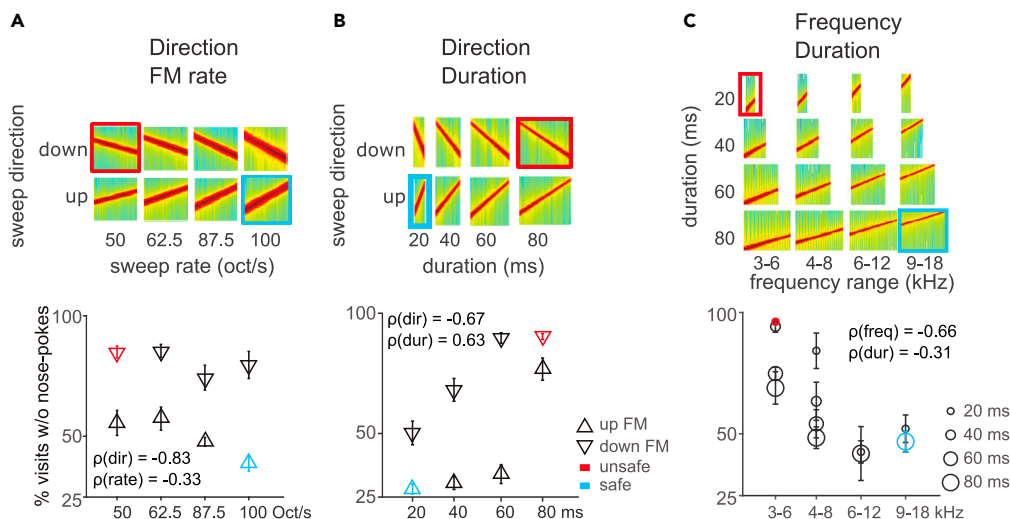
(C) Similar to (A) for mice trained to discriminate FMs with frequency-dependent preferred direction. The sound duration used for a-c was 20 ms, and the rate of frequency modulation was 50 octave/sec.

(D) Similar to (C) for mice trained to discriminate FMs with a duration of 40 ms and frequency modulation at 25 octave/sec. Error bars throughout represent standard error.

downward sweeps;  $n = 7$ ; Figure 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,  $\rho(\text{freq})$ , was  $-0.85$ ; between direction and behavior,  $\rho(\text{dir}) = -0.13$ ; Figure 2B, bottom). Modeling performance with a linear-mixed model revealed a significant main effect of frequency range ( $p < 0.0001$ ), but no effect of sweep direction ( $p = 0.11$ ) or frequency  $\times$  direction interaction ( $p = 0.21$ ).

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.<sup>25–27</sup> 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 (9–18 kHz downward vs. 6 to 3 kHz upward sweeps;  $n = 9$ ; Figure 2C, top). Our results showed that the mice again selectively attended to the frequency range of the novel sounds and ignored sweep direction (the average correlation coefficient between frequency and behavior,  $\rho(\text{freq})$ , is  $-0.91$ ; between direction and behavior,  $\rho(\text{dir}) = 0.08$ ; Figure 2C, bottom). Modeling performance with a linear-mixed model revealed a significant main effect of frequency range ( $p < 0.0001$ ), but no effect of sweep direction ( $p = 0.32$ ) or frequency  $\times$  direction interaction ( $p = 0.57$ ).

Up to this point we had used sweeps of 20 ms duration, based on discriminable durations in mice.<sup>28</sup> Since previous studies have shown that sweep direction discrimination can be limited by the sweep duration,<sup>14,17,18</sup> we now trained and tested mice with 40 ms long sweeps (9–18 kHz downward vs. 6 to 3 kHz upward sweeps;  $n = 7$ ; Figure 2D, top). We found that an increase in sweep duration did not change the generalization pattern (the average correlation coefficient between frequency and behavior,  $\rho(\text{freq})$ , is  $-0.89$ ; between direction and behavior,  $\rho(\text{dir}) = 0.16$ ; Figure 2D, bottom). Modeling performance with a linear-mixed model revealed a significant main effect of frequency range ( $p < 0.0001$ ), but no effect of sweep direction ( $p = 0.12$ ) or frequency  $\times$  direction interaction ( $p = 0.47$ ). Taken together, these results (Figures 2A–2D) show that sound frequency is perceived independently from sweep direction and treated as behaviorally more relevant.



**Figure 3. Bidimensional generalization reveals hierarchical perception of sound dimensions in mice**

(A) Top: Spectrograms of the stimulus set used in the task. The safe sound (blue) is a 100 octave/sec upward FM sweep, and the unsafe sound (red) is a 50 octave/sec downward FM sweep. All FM sweeps were centered at 8 kHz. Bottom: Generalization gradients showing the average response to each stimulus during the generalization phase.

(B) Similar to (A) for the combination of sweep direction and duration dimensions. Average generalization gradients (bottom) revealed an effect of both dimensions on behavior.

(C) Similar to (A) for the combination of frequency range and duration dimensions. Average generalization gradients (bottom) showed that both dimensions controlled the animals' behavior in a frequency-dependent manner. Error bars represent standard error throughout.

### The behavioral control of frequency modulation rate, direction, duration, and 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 frequency modulation rate and direction, sweep duration and direction, or sweep duration and frequency range (Figure 3).

We trained mice to discriminate sounds that differed in frequency 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;  $n = 8$ ; Figure 3A, top). We chose 50 and 100 octave/sec because these rates are within the range (25–250 octave/sec) that elicits discriminable neural responses.<sup>25</sup> As before, we tested responses to novel sounds during a generalization phase. Generalization patterns showed that mice responded to novel sounds according mainly to their sweep direction, and largely ignoring frequency modulation rate (the average correlation coefficient between direction and behavior,  $\rho(\text{dir})$ , was  $-0.83$ ; between rate and behavior,  $\rho(\text{rate}) = -0.33$ ; Figure 3A, bottom). Modeling performance with a linear-mixed model revealed a significant main effect of sweep direction ( $p = 0.0001$ ), a significant main effect of frequency modulation rate ( $p = 0.01$ ), but no interaction ( $F(3,56) = 1.09$ ,  $p = 0.19$ ). 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; this experiment) but be ignored when presented with a third (e.g. frequency; previous set of experiments). 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;  $n = 10$ ; Figure 3B, top). For the sweep duration, we used 20 and 80 ms.<sup>28</sup> Responses changed monotonically as a function of changes in both sound dimensions (the average correlation coefficient between direction and behavior,  $\rho(\text{dir})$ , was  $-0.67$ ; between duration and behavior,  $\rho(\text{dur}) = 0.63$ ; Figure 3B, bottom). It is worth noting that changing the duration of a sweep of fixed frequency range inevitably also alters the frequency modulation rate. Here, the frequency modulation rates for the safe and the unsafe sound were 100 and 25 octave/sec, respectively,

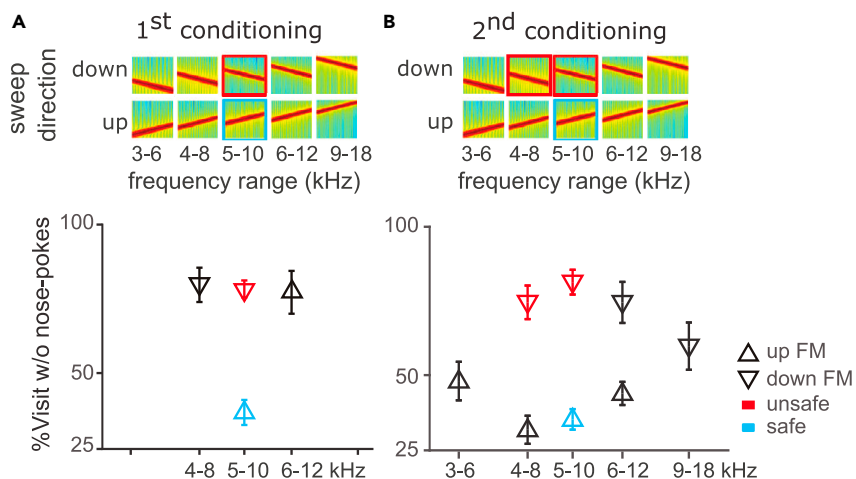
i.e., further apart than in the previous test, where they varied between 100 and 50 octave/sec. Since in the previous experiment the role of frequency modulation rate in discrimination was weak, it is unlikely that mice based their choice on the velocity dimension instead of the duration dimension. Modeling performance with a linear-mixed model revealed a significant main effect of sound duration ( $p < 0.0001$ ), a significant main effect of sweep direction ( $p < 0.0001$ ), but no duration  $\times$  direction interaction ( $p = 0.53$ ). This result indicates that both dimensions contribute equally to sound discrimination. The striking difference in the response to the 60 and 80 ms-long upward modulated sounds, with the former being treated as safe and the latter as unsafe, is interesting. Neither the difference in duration nor the difference in rate (33 and 25 octave/sec, respectively), warrant such dissociation based on behavioral and physiological findings.<sup>29–31</sup>

Finally, we investigated how a new set of mice generalized along the combined dimensions of frequency range and sweep duration (9–18 kHz 80 ms sweeps as safe vs. 3 to 6 kHz 20 ms sweeps as unsafe;  $n = 8$ ; Figure 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,  $\rho(\text{freq}) = -0.66$ ; between duration and behavior,  $\rho(\text{dur}) = -0.31$ ; Figure 3C, bottom). Modeling performance with a linear-mixed model revealed a significant main effect of frequency range ( $p < 0.0001$ ), a significant main effect of duration ( $p < 0.0001$ ) and a significant interaction between the two dimensions ( $p = 0.003$ ). This suggests that the influence of duration on animals' behavior is, in this task, frequency-dependent. The dominance of frequency is exemplified in the homogeneous responses to sounds with the highest frequency, where duration plays no role. The influence of duration over frequency judgment is revealed in the response to the lowest frequency sound, where a differentiation according to duration is made by the mice. Unlike in the previous task, sounds of 60 and 80 ms duration are bundled together here.

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

Up to this point we have tested, during a generalization phase, the untrained response to novel 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 explored how flexible these dynamics are. With this purpose, we first trained naive 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 between sweep pairs of opposite directions but same frequency range, so as to force direction discrimination (Figure 4A, 5–10 kHz upward sweeps as safe vs. 10–5 kHz downward sweeps as unsafe,  $n = 8$ ; Figure 4A, top). Discrimination learning was slow (Figure S3J) but the safe and unsafe sounds were well discriminated after 50 trials of training (Figure 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 frequencies is an upward sweep.<sup>26,27,32</sup> We expected, therefore, direction discrimination to be frequency-specific. Generalization to novel sounds was later measured with sweeps of successively different frequency ranges and directions (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 (Figure 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; Figure 4B, colored symbols) before testing generalization again on the various novel sounds (Figure 4B, black symbols). This conditioning did succeed in generating direction generalization across wider frequency ranges (Figure 4B, black symbols). The average correlation coefficient between direction and behavior was  $-0.82$ , whereas the value between frequency and behavior was  $0.12$  (Figure 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; Figure 4B outermost black symbols). Modeling performance with a linear-mixed model revealed a significant main effect of sweep direction ( $p = 0.004$ ), a weaker effect of frequency range ( $p = 0.02$ ), and interaction between the two dimensions ( $p = 0.02$ ). Therefore, generalization along the



**Figure 4. Generalization along the non-preferred dimension was localized**

(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) Bottom: Generalization gradients showing percentage of visits without nose-pokes for each sound stimulus during the generalization phase.

(B) Same as (A) for generalization after the second conditioning during which the already tested downward FM sound (8-4 kHz downward FM sweep) was also conditioned. Error bars represent standard error throughout.

non-preferred dimension could be trained but remained specific in frequency, possibly because plasticity occurred only within a specific region of the tonotopic map.

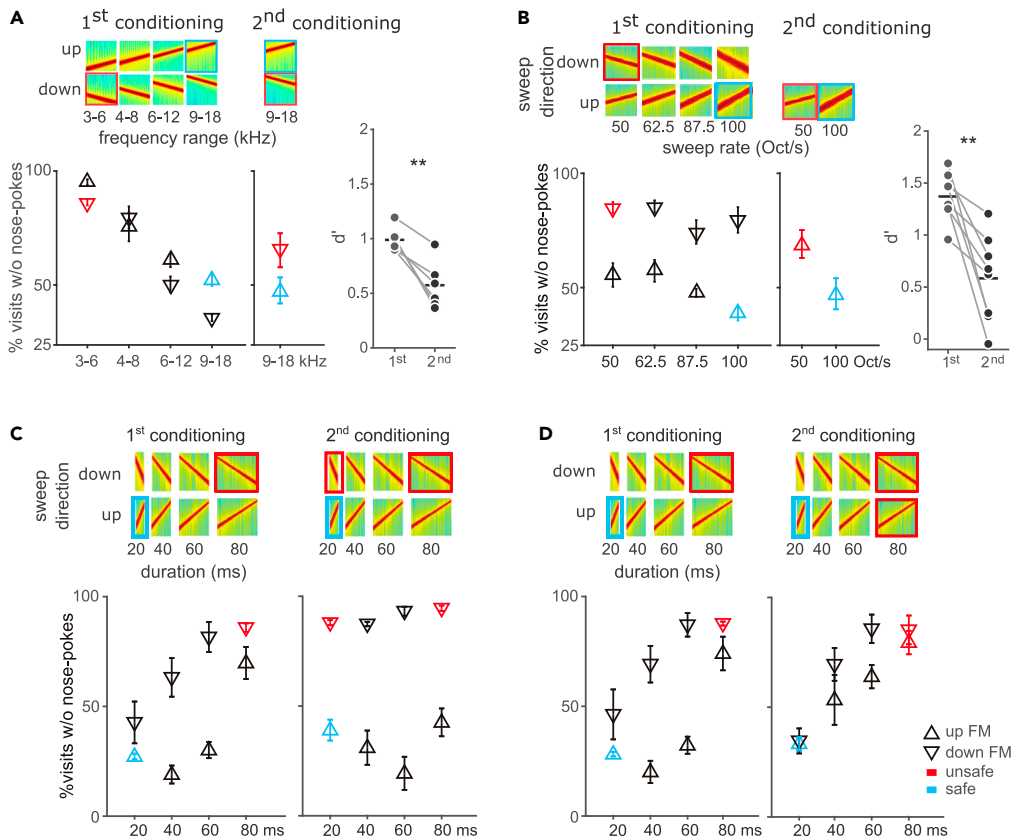
### Behavioral control of sound dimensions is plastic and amenable to training

We then tested whether we could train mice that had 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 (Figure 2A, repeated in Figure 5A left), where they showed a preference for the frequency dimension, while ignoring direction. These same animals were now given a second conditioning with the same safe sound but a high-frequency downward-sweep as the unsafe sound (Figure 5A, middle), to encourage them to discriminate along the direction dimension. The second 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 first conditioning even after 4 days (Figure 5A right; paired t-test for  $d'$ ,  $p = 0.002$ ). A similar result was observed in mice previously trained with the combination direction-rate (Figure 5B, left). When the upward 50 octave/sec 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 (Figure 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 (Figure 3B), were divided in two equally sized groups of comparable performances (Figures 5C and 5D, left). The first half were subsequently trained with an additional unsafe sound (the second unsafe tone) with the same direction as the initial unsafe sound, but the shorter duration of the safe sound (Figure 5C,  $n = 5$ ). Thus, the safe and the unsafe sounds differed now only in sweep direction. After this second conditioning, mice now avoided nose-poking in the presence of the added unsafe tone (Figure 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 (Figure 5C, right, black symbols). The discrimination gradient was now mainly along the direction dimension. Modeling the performance after retraining with a linear-mixed model revealed a significant main effect of sweep direction ( $p = 0.004$ ), effect of duration ( $p = 0.02$ ) and interaction between those two dimensions ( $p = 0.02$ ).

The other half of the mice (Figure 5D) was 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





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

(A) Left: Generalization gradients after initial training (first 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 second conditioning). (A) Right:  $d'$  for the last 50 unsafe visits of the first and the second conditioning for individual mice. Paired t-test was used to check the difference between performances after each conditioning.  $p$  values less than 0.01 are given two asterisks. (A) Mice initially selectively generalized along frequency range (left) and were worst at direction discrimination (middle).

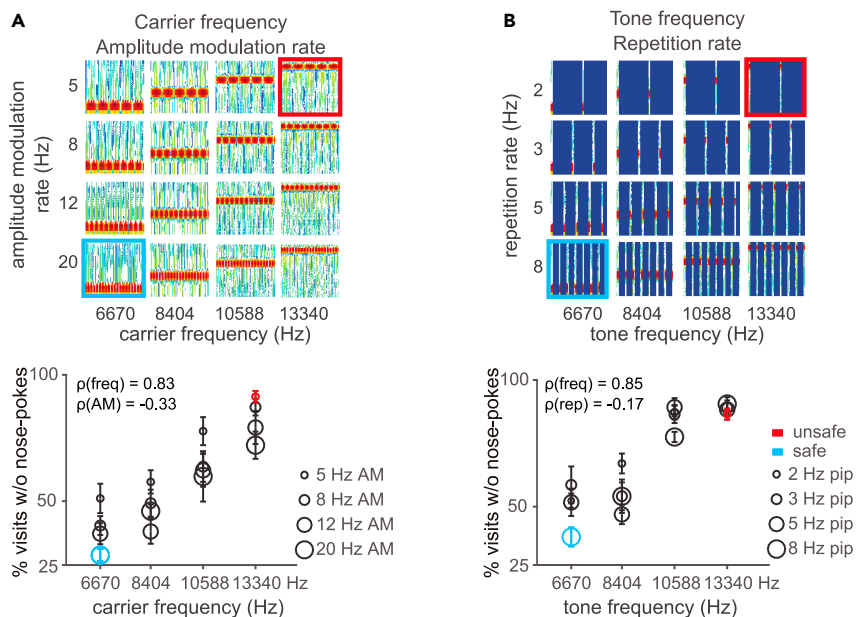
(B) Same as (A) for mice initially trained with sweep direction and rate (left), and subsequently trained to discriminate the non-dominant dimension, rate (middle).

(C and D) Left: Generalization gradients after the first conditioning with direction and duration. Mice initially generalized along both direction and duration dimensions. (C and D) Right: Shift in generalization pattern after the second conditioning to discriminate either direction (C) or duration (D). Responses to the safe and unsafe sound are marked as blue and red, respectively. Error bars represent standard error throughout.

only in sound duration (Figure 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%;  $d'$  second = 1.24), responses to novel sounds changed dramatically. Mice now mainly discriminated along the duration dimension (Figure 5D, right, black symbols). Modeling performance after retraining with a linear-mixed model revealed a significant main effect of sweep duration ( $p < 0.0001$ ), an effect of direction ( $p = 0.001$ ), but no interaction between those two dimensions ( $p = 0.89$ ). Thus, changes in task condition 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 the non-preferred dimension remained more difficult and did not generalize beyond narrow ranges.

### Perception of “envelope” features in mice

The “envelope” of the sound stream is important for auditory perception, for example in speech processing.<sup>33,34</sup> The experiments described so far focused on differences within the short frequency-modulation



**Figure 6. Bidirectional generalization of periodic sound**

(A) Top: Spectrograms of FM sweeps used in the task. (A) Bottom: Generalization gradients plotting the average response to each of the stimuli presented in the task for the combination of tone frequency and AM sweep rate dimensions during the generalization phase.

(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. Error bars represent standard error throughout.

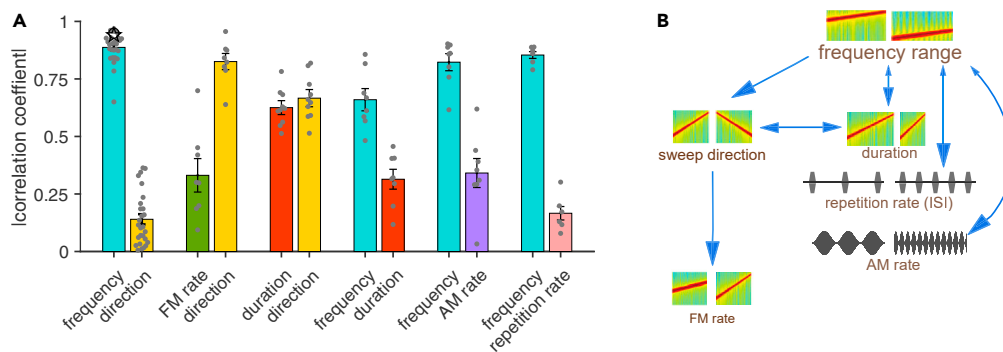
bouts presented always at a repetition rate of 3 Hz (Figures 1, 2, 3, 4, and 5). Next, we investigated how the “envelope”, i.e. the contour, influenced generalization patterns. Since amplitude modulations and the periodic envelope are important characteristics for acoustic communication signals,<sup>35</sup> we trained and tested mice with 100% sinusoidal amplitude-modulated (AM) stimuli (differing in dimensions of carrier frequency and amplitude modulation rate, Figure 6A).

We chose well discriminable carrier frequencies that differed by 1 octave<sup>36</sup> and amplitude modulations ranging between 5 and 20 Hz amplitude modulation rate<sup>4</sup> ( $\Delta F > 1$  octave). Figure 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,  $\rho(\text{freq})$ , was 0.83; between amplitude-amplitude modulation rate and behavior,  $\rho(\text{AM}) = -0.33$ ). Modeling performance with a linear-mixed model revealed a significant main effect of carrier frequency ( $p = 0.0001$ ), no significant effect of amplitude modulation rate ( $p = 0.16$ ), but no interaction between the two dimensions ( $p = 0.76$ ).

We then tested sounds differing in dimensions of tone frequency and repetition rate (Figure 6B) using rates ranging between 2 and 8 Hz ( $\Delta F$  of 2 octaves), since neurons in the auditory cortex selectively prefer stimulus rates ranging from 2 to 15 Hz.<sup>37</sup> Similar to AM stimuli, we found that the animals’ performance was influenced by both dimensions (the average correlation coefficient between frequency and behavior,  $\rho(\text{freq})$ , was 0.85; between repetition rate and behavior,  $\rho(\text{rep}) = -0.17$ ). Modeling performance with a linear-mixed model revealed a significant main effect of tone frequency ( $p < 0.0001$ ), a significant effect of repetition rate ( $p = 0.002$ ) and significant interaction between the two dimensions ( $p = 0.02$ ). Thus, although tone frequency exerts a stronger control over behavior, repetition rate also influences behavior in a frequency-dependent manner.

### Interim summary

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 (Figure 7A). The variability across



**Figure 7. Summary**

(A) Bar plot shows the absolute value of the correlation coefficient between each dimension's tested variable space and the performance of each mouse. Error bars represent standard error.

(B) Scheme shows the perceptual hierarchy inferred by the results.

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 (Figure 7B).

### Initial discrimination learning 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 (Figures S3 and S4, red). The average responses to the safe sound before (gray) and after conditioning (blue) are included. We did this for every dimension combination explored (Figures S3 and S4A–S4G). We also assessed learning for single-dimension discriminations of frequency, direction and duration (Figures S3H–S3K). Learning was quantified by measuring discrimination performance (the  $d'$ ) during the first 50 and the last 50 unsafe trials. Fast learning is reflected in relatively higher  $d'$  over the first 50 trials and smaller differences in  $d'$  between the first and last 50 trials.

Discrimination learning in all shown frequency-direction combinations was slow (Figures S3A–S3D) when compared to discrimination learning based on frequency information alone (Figures S3H and S3I; 2-sample t-test for  $d'$  over the first 50 trials,  $p < 0.0001$ ) but faster when compared to learning based on direction information alone (Figure S3J; 2-sample t-test for  $d'$  over the first 50 trials,  $p = 0.001$ ). For the frequency-direction combination, discrimination learning was progressively faster across tasks in the order shown in Figures S3A–S3D (2 sample t-test for  $d'$  over the first 50 trials, A vs. C  $p < 0.0001$ ; A vs. D  $p < 0.0001$ ; B vs. C  $p = 0.008$ ; B vs. D  $p < 0.0001$ ; C vs. D  $p = 0.0003$ ), and reflected in the final discrimination performance, measured in  $d'$  over the last 50 trials (Figure S3L; 2 sample t-test for  $d'$  over the last 50 trials, A vs. C  $p = 0.006$ ; A vs. D  $p < 0.0001$ ; B vs. C  $p = 0.1$ ; B vs. D  $p = 0.001$ ; C vs. D  $p = 0.04$ ). 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 (Figure 2A). For example, even though using the physiologically preferred direction<sup>25–27</sup> for the given frequency range speed up learning (compare Figures S3A and S3C), 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 in this task, we saw that sweep direction did not slow down learning when combined with duration (Figure S3F; 2 sample t-test for  $d'$  over the first 50 trials, dur/dir vs. dur  $p = 0.75$ ), despite it resulting in slow discrimination learning when alone (compared to duration; Figure S2K should be S3K; 2

sample t-test for  $d'$  over the first 50 trials, dir vs. dur  $p < 0.0001$ ), a pattern that was reflected in the final discrimination learning (Figure S3M; 2 sample t-test for  $d'$  over the last 50 trials, dur/dir vs. dur  $p = 0.42$ , dir vs. dur  $p = 0.0002$ ). 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, which had different effects during the initial training (Figures S4A–S4D,  $p = 0.02$  for freq/AM rate against freq alone;  $p < 0.0001$  for freq/repetition rate against freq alone) but similar pattern of interaction with frequency in the generalization phase (Figure 6).

## DISCUSSION

We set out to understand how sound dimensions are integrated during auditory perception. 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,<sup>38</sup> 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.

To maintain experimental control over the parameter space, we focused on pairwise comparisons across chosen dimensions. Natural sounds are likely to be characterized by multiple dimensions in a way that might depend on sound type. For example, vocalizations might be characterized by the presence of harmonics, as well as by their frequency and duration,<sup>29</sup> while for background sounds the length of individual frequencies might be less relevant than the overall summary statistics.<sup>39</sup> Further investigation is needed to better understand the connection between the statistical properties of multidimensional natural sounds and how animals perceive and react to sounds relevant to their ecological niche.

### 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 (Figure 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<sup>4</sup> or loudness.<sup>40</sup> In mice, maternal responsiveness to wiggling calls depends critically on both frequency and duration,<sup>29,40,41</sup> which matches our inference that frequency range and duration were at high levels of the perceptual hierarchy.

### Neurophysiological consideration of dimension perception

Is the hierarchical relevance of acoustic features reflected in the functional and anatomical organization of the auditory system? The notion that sensory systems might be organized hierarchically, i.e. into a series of stages that progressively increase abstract representations, have been a popular hypothesis for decades.<sup>42</sup> Hierarchical processing of sensory information has been intensively studied in the visual modality.<sup>43–46</sup> The influential reverse hierarchy theory proposed by Ahissar and Hochstein<sup>47</sup> posits that high-level abstract inputs are prioritized for perception along the sensory hierarchy and 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. Frequency-based neural selectivity is observed across all auditory stations, from the cochlea to the cortex, and is organized tonotopically.<sup>48</sup> Neural selectivity for sound duration, sweep direction, sweep rate, repetition rate, and frequency modulation rate all appear as early as the auditory midbrain.<sup>25,31,32,49,50</sup> 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 frequency modulation rate selectivity has been found in subpopulations of neurons in auditory cortex and midbrain.<sup>25,26</sup> Both these dimensions are in fact somewhat embedded in the tonotopic map.<sup>26,27,32</sup> 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 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 saliency 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 that 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 influenced the initial learning rate, and yet, these differences in learning rate had no real influence on 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 might be consistent with the relevance of these dimensions in natural sounds.

### Limitations of the study

To maintain experimental control over the parameter space, we focused on pairwise comparisons across chosen dimensions. Natural sounds are likely to be characterized by multiple dimensions in a way that might depend on sound type. For example, vocalizations might be characterized by the presence of harmonics, as well as by their frequency and duration,<sup>29</sup> while for background sounds the length of individual frequencies might be less relevant than the overall summary statistics.<sup>39</sup> The relationship between multidimensional natural sound statistics and an animal's coding of and behavioral response to niche-relevant sounds remains to be more thoroughly characterized.

### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- [KEY RESOURCES TABLE](#)
- [RESOURCE AVAILABILITY](#)
  - Lead contact
  - Materials availability
  - Data and code availability
- [EXPERIMENTAL MODEL AND SUBJECT DETAILS](#)
- [METHOD DETAILS](#)
  - Apparatus: The Audiobox
  - Sounds
  - Discrimination task
  - Multidimensional generalization measurement

- Analysis of performance in the audiobox
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2023.106941>.

## ACKNOWLEDGMENTS

We are grateful to members of the de Hoz lab for critical discussions, to Israel Nelken, Nicol Harper, and Freddy Trinh for insightful comments on earlier versions of this work, and to Klaus-Armin Nave for support. We are also grateful to three anonymous reviewers for their useful comments. CC was partially funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), grant number 327654276 (SFB 1315, A09) to LdH.

## AUTHOR CONTRIBUTION

C.C. and L.d.H. conceptualized and designed the experiments, C.C. run and analyzed the experiments, C.C. and L.d.H. wrote the paper.

## DECLARATION OF INTERESTS

The authors declare no competing interests. Chi Chen is currently an employee of Miltenyi Biotec, Bergisch Gladbach, Germany.

## INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: February 6, 2023

Revised: March 30, 2023

Accepted: May 18, 2023

Published: May 22, 2023

## REFERENCES

1. Plato. (1925). Phaedrus (Fowler), pp. 265d–266a.
2. Ehret, G. (1975). Frequency and intensity difference limens and nonlinearities in the ear of the housemouse (*Mus musculus*). *J. Comp. Physiol.* 102, 321–336.
3. Fay, R.R. (1974). Auditory frequency discrimination in vertebrates. *J. Acoust. Soc. Am.* 56, 206–209.
4. Kurt, S., and Ehret, G. (2010). Auditory discrimination learning and knowledge transfer in mice depends on task difficulty. *Proc. Natl. Acad. Sci. USA* 107, 8481–8485.
5. Rohrbaugh, M., Brennan, J.F., and Riccio, D.C. (1971). Control of two-way shuttle avoidance in rats by auditory frequency and intensity. *J. Comp. Physiol. Psychol.* 75, 324–330.
6. Syka, J., Rybalko, N., Brozek, G., and Jilek, M. (1996). Auditory frequency and intensity discrimination in pigmented rats. *Hear. Res.* 100, 107–113.
7. Kobrina, A., Toal, K.L., and Dent, M.L. (2018). Intensity difference limens in adult CBA/CaJ mice (*Mus musculus*). *Behav. Processes* 148, 46–48.
8. Terman, M. (1970). Discrimination of auditory intensities by rats. *J. Exp. Anal. Behav.* 13, 145–160.
9. Mercado, E., 3rd, Orduña, I., and Nowak, J.M. (2005). Auditory categorization of complex sounds by rats (*Rattus norvegicus*). *J. Comp. Psychol.* 119, 90–98.
10. Ohl, F.W., Scheich, H., and Freeman, W.J. (2001). Change in pattern of ongoing cortical activity with auditory category learning. *Nature* 412, 733–736.
11. Rybalko, N., Suta, D., Nwabueze-Ogbo, F., and Syka, J. (2006). Effect of auditory cortex lesions on the discrimination of frequency-modulated tones in rats. *Eur. J. Neurosci.* 23, 1614–1622.
12. Wetzel, W., Wagner, T., Ohl, F.W., and Scheich, H. (1998). Categorical discrimination of direction in frequency-modulated tones by mongolian gerbils. *Behav. Brain Res.* 91, 29–39.
13. Schulze, H., and Scheich, H. (1999). Discrimination learning of amplitude modulated tones in Mongolian gerbils. *Neurosci. Lett.* 261, 13–16.
14. Gaese, B.H., King, I., Felsheim, C., Ostwald, J., and von der Behrens, W. (2006). Discrimination of direction in fast frequency-modulated tones by rats. *J. Assoc. Res. Otolaryngol.* 7, 48–58.
15. Florentine, M. (1983). Intensity discrimination as a function of level and frequency and its relation to high-frequency hearing. *J. Acoust. Soc. Am.* 74, 1375–1379.
16. Heffner, H., and Masterton, B. (1980). Hearing in glires: domestic rabbit, cotton rat, feral house mouse, and kangaroo rat. *J. Acoust. Soc. Am.* 68, 1584–1599.
17. Nabelek, I., and Hirsh, I.J. (1969). On the discrimination of frequency transitions. *J. Acoust. Soc. Am.* 45, 1510–1519.
18. Schouten, M.E. (1985). Identification and discrimination of sweep tones. *Percept. Psychophys* 37, 369–376.
19. Vyazovska, O.V. (2021). The effect of dimensional reinforcement prediction on discrimination of compound visual stimuli by pigeons. *Anim. Cogn.* 24, 1329–1338.
20. Ahissar, M., and Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends Cogn. Sci.* 8, 457–464.
21. Melara, R.D., and Marks, L.E. (1990). Interaction among auditory dimensions: timbre, pitch, and loudness. *Percept. Psychophys* 48, 169–178.

22. Allen, E.J., and Oxenham, A.J. (2014). Symmetric interactions and interference between pitch and timbre. *J. Acoust. Soc. Am.* 135, 1371–1379.
23. de Hoz, L., and Nelken, I. (2014). Frequency tuning in the behaving mouse: different bandwidths for discrimination and generalization. *PLoS One* 9, e91676.
24. Cruces-Solis, H., Jing, Z., Babaev, O., Rubin, J., Gür, B., Krueger-Burg, D., Strenzke, N., and de Hoz, L. (2018). Auditory midbrain coding of statistical learning that results from discontinuous sensory stimulation. *PLoS Biol.* 16, e2005114.
25. Hage, S.R., and Ehret, G. (2003). Mapping responses to frequency sweeps and tones in the inferior colliculus of house mice. *Eur. J. Neurosci.* 18, 2301–2312.
26. Heil, P., Langner, G., and Scheich, H. (1992). 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.* 171, 583–600.
27. Zhang, L.I., Tan, A.Y.Y., Schreiner, C.E., and Merzenich, M.M. (2003). Topography and synaptic shaping of direction selectivity in primary auditory cortex. *Nature* 424, 201–205.
28. Ehret, G. (1992). Categorical perception of mouse-pup ultrasounds in the temporal domain. *Anim. Behav.* 43, 409–416.
29. Geissler, D.B., and Ehret, G. (2002). Time-critical integration of formants for perception of communication calls in mice. *Proc. Natl. Acad. Sci. USA* 99, 9021–9025.
30. Klink, K.B., and Klump, G.M. (2004). Duration discrimination in the mouse (*mus musculus*). *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* 190, 1039–1046.
31. Brand, A., Urban, R., and Grothe, B. (2000). Duration tuning in the mouse auditory midbrain. *J. Neurophysiol.* 84, 1790–1799.
32. Kuo, R.I., and Wu, G.K. (2012). The generation of direction selectivity in the auditory system. *Neuron* 73, 1016–1027.
33. Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 336, 367–373.
34. Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402.
35. Leaver, A.M., and Rauschecker, J.P. (2010). Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. *J. Neurosci.* 30, 7604–7612.
36. Chen, C., Krueger-Burg, D., and de Hoz, L. (2019). Wide sensory filters underlie performance in memory-based discrimination and generalization. *PLoS One* 14, e0214817.
37. Kilgard, M.P., and Merzenich, M.M. (1999). Distributed representation of spectral and temporal information in rat primary auditory cortex. *Hear. Res.* 134, 16–28.
38. Greenwald, A.G., McGhee, D.E., and Schwartz, J.L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *J. Pers. Soc. Psychol.* 74, 1464–1480.
39. McWalter, R., and McDermott, J.H. (2018). Adaptive and selective time averaging of auditory scenes. *Curr. Biol.* 28, 1405–1418.
40. Polley, D.B., Steinberg, E.E., and Merzenich, M.M. (2006). Perceptual learning directs auditory cortical map reorganization through top-down influences. *J. Neurosci.* 26, 4970–4982.
41. Ehret, G., and Riecke, S. (2002). Mice and humans perceive multiharmonic communication sounds in the same way. *Proc. Natl. Acad. Sci. USA* 99, 479–482.
42. Chechik, G., Anderson, M.J., Bar-Yosef, O., Young, E.D., Tishby, N., and Nelken, I. (2006). Reduction of information redundancy in the ascending auditory pathway. *Neuron* 51, 359–368.
43. Grill-Spector, K., and Malach, R. (2004). The human visual cortex. *Annu. Rev. Neurosci.* 27, 649–677.
44. Hochstein, S., and Ahissar, M. (2002). View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron* 36, 791–804.
45. Markov, N.T., and Kennedy, H. (2013). The importance of being hierarchical. *Curr. Opin. Neurobiol.* 23, 187–194.
46. Hegdé, J. (2008). Time course of visual perception: coarse-to-fine processing and beyond. *Prog. Neurobiol.* 84, 405–439.
47. Ahissar, M., and Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature* 387, 401–406.
48. Poon, P.F., and Brugge, J.F. (2012). *Central Auditory Processing and Neural Modeling* (Springer Science & Business Media).
49. Langner, G., Albert, M., and Briede, T. (2002). Temporal and spatial coding of periodicity information in the inferior colliculus of awake chinchilla (*Chinchilla laniger*). *Hear. Res.* 168, 110–130.
50. Clayton, K.K., Williamson, R.S., Hancock, K.E., Tasaka, G.-I., Mizrahi, A., Hackett, T.A., and Polley, D.B. (2021). Auditory corticothalamic neurons are recruited by motor preparatory inputs. *Curr. Biol.* 31, 310–321.
51. Lipp, H.-P., Litvin, O., Galsworthy, M., Vyssotski, D., Vyssotski, A., Zinn, P., Rau, A., Neuhäusser-Wespy, F., Würbel, H., Nitsch, R., et al. (2005). Automated behavioral analysis of mice using INTELLICAGE: Inter-laboratory comparisons and validation with exploratory behavior and spatial learning. In *Proceedings of Measuring Behavior 2005*, L.P.J.J. Noldus, F. Grieco, L.W.S. Loijens, and P.H. Zimmermann, eds. (Fifth International Conference on Methods and Techniques in Behavioral Research), pp. 66–69.
52. Maor, I., Shwartz-Ziv, R., Feigin, L., Elyada, Y., Sompolinsky, H., and Mizrahi, A. (2019). Neural correlates of learning pure tones or natural sounds in the auditory cortex. *Front. Neural Circuits* 13, 82.

## STAR★METHODS

## KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental models: Organisms/strains		
C57BL/6JOLAHsd female mice	Janvier, France	N/A
Deposited data		
Raw data	This paper	Mendeley Data: <a href="https://doi.org/10.17632/jgsv9xgwk8.1">https://doi.org/10.17632/jgsv9xgwk8.1</a>
Software and algorithms		
MatLab	mathworks	<a href="https://ch.mathworks.com/products/matlab.html">https://ch.mathworks.com/products/matlab.html</a>
Other		
Audiobox (intellicage)	TSE, Germany	

## RESOURCE AVAILABILITY

## Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Livia de Hoz ([livia.dehoz@charite.de](mailto:livia.dehoz@charite.de)).

## Materials availability

This study did not generate new unique reagents.

## Data and code availability

- Raw data have been deposited at Mendeley and are publicly available as of the date of publication.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

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 ( $21 \pm 1^\circ\text{C}$ ) on a 12 h light/dark schedule (7am/7pm) with access to food and water *ad libitum*. Females are typically used in the Intellicage<sup>51</sup> and Audiobox,<sup>23</sup> as well as the sister paradigm, the Educage,<sup>52</sup> to avoid male-associated dominance effects that can lead to fights in enriched environments. Within group variability is typically small in the Audiobox and estrous cycles were not assessed.

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.

## METHOD DETAILS

## Apparatus: The Audiobox

The Audiobox is 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. (0.1 ml/10 g) 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 (Figures S1A and S1B). 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 reads 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 1/4" 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 octave/sec as default). The sweeps were presented at roving-intensities (70 dB ± 3 dB). For the AM stimuli, tested dimensions were carrier frequency and amplitude 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 amplitude 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.<sup>9,12,25,27</sup> Mice show increased sensitivity for frequencies between 4 and 16 kHz,<sup>27</sup> and easily discriminate frequency distances of at least 0.20 octave,<sup>23,27,36</sup> 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/sec velocity of modulation was based on both physiological<sup>23</sup> and behavioral<sup>14</sup> 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 Figure 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,<sup>36</sup> mice were trained to discriminate between 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 (safe-only phase). 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 during the conditioning phase and a nose-poke during these visits was associated with an air puff and no access to water (Figure S1C). 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 (Figure 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, they 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 Figures S1D and S1E. 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 Figures S1F and S1G.

### 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  $Z(p)$ ,  $p \in [0, 1]$ , 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.

During discrimination learning, to calculate the  $d'$  over the first and last 50 unsafe trials, we used the percentage of all safe visits without nose-poke that happened during the corresponding unsafe trial periods (1st to 50th unsafe visit and last-50 to last unsafe visit, respectively).

### QUANTIFICATION AND STATISTICAL ANALYSIS

We tested the data 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 the 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 analyzed data with linear-mixed models with the mouse as a random effect. 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  $\pm$  SEM. Statistical significance was considered if  $p < 0.05$ .