Coding of unsupervised predictable sound-context associations in the auditory midbrain

Chi Chen[1,2,3,4], Alexandra Ertman[4,5], 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]International Graduate Program Medical Neurosciences, Charité Medical University, Berlin, Germany.

*Corresponding author: livia.dehoz@charite.de

Keywords: auditory, statistical learning, mouse, prediction

Abstract

The acoustic context is ubiquitous, highly informative and often predictable. Yet, we know little about where and how contextual sounds are coded in the brain. We have some understanding about the coding of reinforced sounds in the auditory cortex and subcortical midbrain. Contextual sounds, however, are learned in an unsupervised (non-reinforced) manner. Here we measured plasticity in the auditory midbrain, a hub of incoming and feedback auditory input, and found that it reflected learning of contextual information in a manner that depended on the predictability of the sound-context association and not on reinforcement. We found broad frequency-specific shifts in tuning in auditory midbrain neurons. These shifts were paralleled by an increase in response gain and correlated with an increase in neuronal frequency discrimination. The auditory midbrain, thus, codes for predictable and unsupervised sound-context associations. This points to a subcortical engagement in the detection of contextual sounds, a detection that facilitates the processing of relevant information described to occur in cortical auditory structures.

Introduction

The acoustic context that surrounds us is laden with information, some of which is predictable. Indeed, As animals move through their environment, they can often predict the sounds they will be exposed to. For example, when we enter a bar we have an expectation of what sounds we will hear, and this expectation is different when entering a swimming pool or moving through a forest. Being able to assess whether these sounds match our prediction and respond accordingly is important, and yet we know little about where these predictions are coded. Here, we investigated sound coding in subcortical auditory processing as mice learned contextual sounds. This study aimed to test whether neuronal activity in the auditory midbrain, a subcortical auditory structure at the interface between sensory and cognitive processing, reflected the learning of the predictable but not reinforced sound structure in the surrounding environment.

The acoustic context, like all structured information available in the environment, is often learned simply because it is there, independently of its immediate relevance. This statistical learning (Reber, 1967) is something the auditory system is optimized for, probably because of its omnidirectional sensitivity across a wide range of distances and attentional states. Yet, we know little about the mechanisms and circuits that underlie this form of learning. Learning of associations between sounds and reinforcers (punishments or rewards) leads to plasticity in auditory structures in the form of frequency-specific tuning shifts towards the reinforced tone, a process that has been particularly well described in the auditory cortex (Bieszczad & Weinberger, 2010; Blake et al., 2002; Letzkus et al., 2011) and, to a lesser extent, in the auditory midbrain (Gao & Suga, 2000; Yan et al., 2005; Zhang & Suga, 2005). Evidence for coding of unsupervised (non-reinforced) contextual sounds also exists. For example, cortical auditory structures show noise-invariant responses (Khalighinejad et al., 2019; Mesgarani et al., 2014; Rabinowitz et al., 2012), and prior listening exposure to the sounds in a room can improve speech understanding in listeners (Brandewie & Zahorik, 2018) and influence our perception of the world in a pre-attentive manner (McPherson & McDermott, 2020). Our own work determined that exposure over days to a context in which a sound is systematically played in a behaviourally irrelevant manner, a sound-context association, led to shifts in tuning in the auditory midbrain (specifically the inferior colliculus, IC) that seemed frequency-unspecific (Cruces-Solís et al., 2018).

Here we aimed to further study the role of different contextual factors, specifically the presence of a reinforcer, the predictability of the sound onset, and the sound exposure time, in the development of associative plasticity in the auditory midbrain.

Results

1. Contextual learning elicits IC plasticity independently of reinforcer

We first examined the role of reinforcement on the plasticity elicited by predictable sound-context associations in the auditory midbrain. We separately exposed three groups of mice in the Audiobox (Fig 1A), an enriched and automatic apparatus made of 2 main areas connected by corridors in which mice live in groups for days with little interaction with the experimenter. Each group had a different sound-context pairing experience in the sound corner (see Methods, Fig 1A and B). Importantly, none of the animals were conditioned and food and water were always *ad libitum*. Exposed mice, in the “predictable” and “noW-predictable” exposure groups (noW: no water; Fig 1A middle and right), heard 16 kHz tone pips for the duration of each visit to the sound corner. The control group (Fig 1A left), on the other hand, lived in the Audiobox for the same length of time but was not exposed to any sound in the sound corner. Two water bottles were typically in the corner and mice drank for about 60% of the time they were in the corner in about 70% of the visits (Cruces-Solís et al., 2018; de Hoz & Nelken, 2014). Unlike the other two groups, the “noW-predictable” group had no access to water in the corner and, instead, water bottles were in the food area, where no sound was ever played. This ensured that no association was made between the sound and the water. Then, to ascertain auditory midbrain plasticity, we acutely recorded from the left inferior colliculus (IC) of anesthetized animals that had been sound-exposed in the Audiobox for 6 to 12 days. For this experiment, we recorded multiunit activity using NeuroNexus (NeuroNexus, MI, USA) multi-electrode arrays with 16 sites linearly arranged and separated by 50 µm (Fig 1B). The electrode was dorsoventrally inserted in IC along the tonotopic axis and was placed in two consecutive positions to cover a range of depths between 0 and 1250 µm (see Methods; Fig 1B). To characterize frequency tuning, we presented sweeps of 30 ms long pure tones of different frequency-intensity combinations (see Methods) and built tuning curves for each recording site.

The plastic changes we observed in the predictable group replicated and extended those we reported in Cruces-Solís et al. (Cruces-Solís et al., 2018). We again saw an increase in response gain and a shift in frequency tuning in this group compared to the control group. The increase in sound-evoked responses was evident when comparing the tuning curves of predictable and control animals, shown in Fig 1C across example depths of 350, 550, and 700 mm (best frequency of 16 kHz for control animals) (Fig 1C; blue *vs.* red). The increase in gain was observed across all frequency regions (Supplementary Fig 1G; blue *vs.* red). The shift in best frequency (BF) is also evident in the tuning curves and quantified in Fig 1D. In addition, we could ascertain that the upward shift, while visible across many depths beyond the exposed frequency band, did not extend throughout the inferior colliculus as had previously assumed (Cruces-Solís et al., 2018). In fact, below 1100 µm in depth the shift seems to be downwards.

The key result, however, is that the pattern was similar for animals that were exposed but had no water in the corner. The increase in response gain was also present in the noW-predictable group (Fig 1C). While the presence of water in the corner did not imply a temporal association between the sound and the water, since mice only drank sporadically during a given visit and they often had visits with no drinking (de Hoz & Nelken, 2014), the results confirm that the change in response gain was not caused by an association between the 16 kHz contextual sound and the presence of water in the corner. Quantification of average evoked response across tone frequencies confirmed that the enhancement of evoked responses in both normal and no-water exposure animals relative to control was homogenous along the IC tonotopic axis (Supplementary Fig 1G; linear mixed model [two factors: group × depth]: group $p = 0.016$, depth $p = 0.001$). The increase in evoked responses at higher depths is likely related to the relative shank width of the probe (A1x16-3mm-50-177) at different depths (Claverol-Tinture & Nadasdy, 2004). Nevertheless, probe positioning was consistent across different animals and cannot explain the differences between the groups. Further quantification of changes in BF relative to the control across recording depths between 100 and 750 µm also confirmed that the shift in BF was not dependent on the presence of water (Fig 1D; linear mixed effect model [two factors: group × depth]: overall group $p = 0.03$, ctrl *vs.* pred. $p = 0$, pred. *vs.* no-water $p = 0.0002$, ctrl *vs.* no-water $p = 0.0005$). The BF shift in the noW-predictable group was not as strong as in the predictable group, indicating that while plasticity is not dependent on the water association, its presence might strengthen it.

These results replicated and substantially extended our previous finding (Cruces-Solís et al., 2018). We confirmed that exposure to sound-context associations led to an enhancement in overall response gain and a global shift in BF. Importantly, this plasticity was unsupervised since it did not depend on the association between sound exposure and access to water.

2. Predictability in the contextual sounds triggers plasticity

We next tested the effect of predictability by contrasting our results with the effect of unpredictable exposure. We focused on the predictability of the onset of sound exposure, which had previously been controlled by the mouse's entry into the corner. We trained a new group of mice in the Audiobox and exposed them again to a pure tone of 16 kHz. This time, however, the tone pips were not, as before, triggered by the onset of the visit but appeared randomly 5-25 seconds after visit onset and only in 80% of the visits (Fig 1E, Supplementary Fig 1A). This means that sound was not controlled by the mouse movement and also not present for the visit's duration. Compared to the control group that was trained in parallel, the magnitude and BF across depths in the unpredictable group did not differ from those in the control group (Supplementary Fig 1B, C, H). Quantification of the average difference in BF with respect to the control group across the depths between 100 and 750 mm confirmed that there was no shift in BFs in the unpredictable group (Fig G-F; Supplementary Fig 1C inserted; linear mixed model, $p = 0.69$). Therefore, the unpredictable onset of sound exposure, coupled with a weaker sound-context association, did not elicit plasticity in the IC.

We next examined whether the absence of an effect was caused by the shorter sound-context association that resulted from a delayed sound onset, rather than from its unpredictability. We trained two additional groups of mice in the Audiobox. In both groups the sound was initiated by the mouse as she entered the corner but, while one group was exposed to the predictable sound, as before, for the duration of each visit, the other was exposed only for the first 3 seconds ("entry-only" exposure, Fig 1E; Supplementary Fig. 1D). The entry-only group showed sound-evoked responses in the IC that were comparable to those of the predictable exposure group trained in parallel (Supplementary Fig. 1E, F, I). The average BFs across depths were also comparable (Fig G-F; Supplementary Fig. 1F; linear mixed model, $p = 0.08$).

When combined, the comparison across groups trained with different sound-context association patterns (Fig 1E, 2F) revealed that predictability (sound triggered by visit onset) and neither reinforcement (water in the corner) nor the exposure time, determined the presence of plasticity in the IC (Fig 1F; linear mixed model on 3 factors [water × onset × context], a significant effect of onset $p < 0.0001$). This effect was also visible in the tuning curves (Fig. 2G).

3. The BF shift is broadly frequency-specific

To better understand the frequency-specificity of the BF shift, we directly compared the effect of predictable and unpredictable sound-context associations, in the absence of water, using Neuropixels probes (Imec, Belgium) that sampled the IC in depth (Fig 2A). We trained two new groups of mice, a predictable and an unpredictable exposure group, both without water in the corner. As expected, mice in the noW-predictable group showed an upwards shift in BF across a wide range of depths compared to the noW-unpredictable group (Fig 2B-C; linear mixed model [two factors: group × depth]: group $p = 0.016$, depth $p = 0.002$, interaction $p = 0.013$), confirming again that predictability of the sound-context association is the key variable determining the presence of plasticity. As is evident in the plots (Fig 2B-C), the shift is constrained in depth, and therefore in frequency coding range. The BF shift in the predictable groups (predictable, noW-predictable, entry-only) is similar in magnitude for frequencies below 16 kHz, and extends for frequencies above 16 kHz until 25 kHz.

We investigated how BF shifted along the tonotopic axis by plotting the change in BF compared to the control group as a function of the BF of the control group (Fig 2D). Modeling a linear mixed effect model to predict the change in BF with two factors [(water + onset + context) × BF] reveals a significant effect of onset ($p < 0.0001$) and interaction between onset and BF ($p < 0.0001$). A similar pattern of change was observed when comparing the noW-predictable with the noW-unpredictable group (Fig 2E; linear mixed model [two factors: group × BF]: group $p < 0.0001$, BF $p = 1$, interaction $p < 0.0001$).

To sum up, we exposed six groups of mice in the Audiobox with different sound exposure paradigms and examined the effect of three possible factors, sound-water association, sound-entry predictability, and context-specificity, on plasticity induced by sound exposure in the Audiobox (Fig. 2F). Again, it is important to note that none of the groups were conditioned, only

exposed. Behaviorally, consistent with previous data (Cruces-Solís et al., 2018), sound exposure, independently of its duration and predictability (predictable, unpredictable, and entry-only), did not change the animals' behavior when compared to the control group (Fig 2G-H). These groups showed comparable daily time spent in the corner (Fig. 2G, blue *vs.* red, purple and dark green; one-way ANOVA reveals no significant effect of the group, $F(3,56)=2$, $p = 0.12$). Interestingly, the no-water groups still visited the corner, despite the absence of water, albeit for a relatively shorter time compared to the other exposure groups (Fig. 2C, red *vs.* pink, and dark green vs light green; two-sample t-test, $p < 0.0001$ for both comparisons). This resulted in a relatively short time of sound exposure (Fig. 2H). What this quantification revealed is that plasticity, measured here through the magnitude of the BF shift (Fig 2D-E) could not be explained through differences in the length of sound exposure, since groups with different exposure time (the entry-only group and the predictable group, for example) showed similar shift magnitudes. To conclude, sound exposure elicited a broad frequency-specific shift in BFs along the IC tonotopic axis only when the sound was predicted by visit onset and this shift was independent of the presence of water in the corner or the length of the sound exposure (Fig. 2E).

4. Increased frequency discriminability coupled to BF shifts

We next explored how the change in coding that was so dramatically reflected in the BF shift affects frequency representation. First, we looked at the excitation profiles, the response to a given frequency across the structure, a concept first described in the auditory nerve and then referred to as structural tuning by our previous work (Cruces-Solís et al., 2018). We feel that excitation profiles better reflect the ongoing responses to sounds in IC. For example, when the 16 kHz sound is presented, cells at several depths in the IC will respond with different strength (Fig 3A). Using all the data recorded with Neuronexus probes across groups, we observe that these frequency-depth plots illustrate well the increasing BF with depth characteristic of the tonotopic map of IC, but they also reveal other interesting features. For example, at higher depths neurons often respond to low frequencies in addition to the depth-appropriate frequencies. This is what in the auditory nerve would be described as the long tail (Kiang & Moxon, 1974), which can also be observed in the low frequency cluster at higher depths in the Neuropixels recordings (Fig 2B). We then quantified the discriminability between responses to

nearby frequencies (Fig 3B) and found increased separation in the predictable groups compared to the unpredictable group (Fig 3C) in a delta F dependent manner (Fig 3D). Modeling a linear fixed effect model to predict the Euclidean distance with three factors [(water + onset + context) × frequency × delta F], reveals significant effect of onset ($p < 0.0001$), context ($p = 0.01$), frequency ($p < 0.0001$), delta F ($p < 0.0001$) and interactions between onset and frequency ($p < 0.0001$), frequency and delta F ($p < 0.0001$), onset, frequency and delta F ($p < 0.0001$). Discriminability not surprisingly increased with delta F, but it did so more in the predictable groups. This increase was not limited to comparisons with the exposed frequency and, like the BF shift, extended above and below. Independently of delta F, the discriminability in all the groups, including the control, was larger for frequencies near and above 16 kHz but weaker for frequencies below. This pattern was more patent in the predictable groups.

Discussion

We set out to determine whether changes in frequency representation in mouse auditory midbrain could be the result of unsupervised learning of predictable contextual sounds, a form of statistical learning. We found that exposure in an enriched environment to predictable sound-context associations led to an upward and broadly frequency-specific shift in best frequency and an increase in response gain, that paralleled increased neuronal frequency discrimination. These changes were independent of reinforcer since they developed also in mice exposed to sound-context associations in the absence of water. Plasticity was observed in all exposure conditions in which the sound was activated as the mouse entered the corner, independently of whether it was present for the duration of the visit, but not when the sound was activated at random times after visit onset. We conclude that the auditory midbrain is sensitive to the predictability of contextual sounds and discuss the consequences this might have for the processing of background and foreground acoustic information.

The results expand previous findings from the lab where we found an increase in gain and, what we thought was a frequency-unspecific BF shift in animals exposed to predictable sounds in the presence of water (Cruces-Solís et al., 2018). The sampling of a wider depth range with both Neuronexus and Neuropixels probes in the current study, as well as the comparison between several exposure treatments allows us to conclude that plasticity is associated specifically with

the unsupervised learning of predictable sounds, and that the BF shift, although broad, is frequency-specific.

Electrophysiology was performed acutely in anesthetized mice that were removed from the enriched environment in which they were exposed, at time points between 6 and 12 (Neuronexus measures) or 19 to 21 days (Neuropixels) from the beginning of sound exposure. The narrow corridors and ceilings in the Audiobox made it difficult to take measurements *in situ*. Anesthesia is unlikely to affect the different groups differently, but might influence the magnitude of some of the patterns observed. Comparison between the current findings and data obtained from awake mice will be important.

When mice in the Audiobox are exposed to sounds that do not predict a punishment, as in our study, they act as if the sound was not there (see also Cruces-Solis). Thus, animals exposed to a predictable or an unpredictable sound, or no sound at all, spent comparable daily time lengths in the corner. The presence or absence of water in the corner did influence the motivation to go to the corner, but the absence of water did not abolish the natural curiosity of the mice, who still performed several visits per day independently of whether the sound was predictable or not. But this lack of effect on the behavior should not be confused with lack of learning. We have observed across many studies that exposure in the Audiobox without reinforcement leads to latent learning that is only expressed during subsequent reinforced learning (Cruces-Solís et al., 2018; de Hoz & Nelken, 2014), which has different time courses than reinforced learning in the same apparatus (Chen et al., 2019, 2021; de Hoz & Nelken, 2014), and influences reinforced associative plasticity in the auditory cortex (de Hoz et al., 2022). Unsupervised exposure does lead to unsupervised learning and we believe that the changes in midbrain auditory responses are a correlate of this learning. Indeed, we found that neuronal generalization during unsupervised exposure paralleled behavioral generalization measured through prepulse inhibition of the startle reflex (Cruces-Solís et al., 2018).

In the experiments described here, the sound is controlled by the mouse entering the corner. In the predictable groups this entry activates the sound, for the length of the visit or only for 3 seconds. In the unpredictable groups, this entry starts the clock that will determine at which of 5 random intervals, ranging from 5 to 25 seconds, the sound will begin. A BF shift was only

observed in animals exposed to a predictable sound even when the sound was only present for the first 3 seconds from visit onset. Overall the data suggest that predictability is defined by the movement of the mouse through the environment. It is not unreasonable to think that the mice learn that entering the corner will be coupled with sound onset.

Frequency-specific shifts in tuning are a hallmark of auditory plasticity (Bakin & Weinberger, 1990; Blake et al., 2002; Yan et al., 2005; Zhang & Suga, 2005). Typically, when an animal is trained in a frequency conditioning paradigm, a shift in BF is observed that is localized to the population of neurons that respond to the conditioned frequency. This shift enables and parallels an increase in the representation of the relevant frequency (Blake et al., 2002). What is surprising in our results is that the shift occurs also in populations that do not respond to the exposed frequency (16 kHz) and is always an upwards shift, independently of whether the BF is below or above 16 kHz. In our initial study, we thought the shift was frequency-unspecific, perhaps covering the whole depths of the IC (Cruces-Solís et al., 2018). In the current work, sampling through a wider range of depths, we conclude that the shift is frequency-specific but broad, extending to about ⅔ of an octave above and more than 1 octave below the exposed frequency (Fig 2D-E). The shift is likely tightly related to the observed changes in frequency representation in the form of increased frequency discriminability for the predictable groups. The mechanism thus seems to differ from that associated with reinforced learning, although both revolve around frequency shifts. While reinforced learning leads to a local shift towards the relevant frequency, with a concomitant decrease in discriminability for nearby frequencies (Bao et al., 2013; de Villers-Sidani et al., 2007), unsupervised learning results in a broad upward shift that leads to overall increased frequency discriminability.

Finally, we observe these changes in the inferior colliculus, an auditory midbrain hub previously associated with short lasting (up to 60 minutes) plasticity that is dependent on cortical feedback (Gao & Suga, 2000; Yan et al., 2005). The plasticity we observe after at least 6 days and up to 21 days of predictable contextual sound exposure might have different roles. Evidence for an early detection of context can be deduced from noise-invariance in cortical structures. Indeed, using naturalistic background sounds, several labs have found evidence for noise-invariant responses in cortical structures (Khalighinejad et al., 2019; Mesgarani et al., 2014; Norman-Haignere & McDermott, 2018; Rabinowitz et al., 2012), supporting the idea that context is coded

separately. One possibility is that predictable acoustic context is learned through cortico-subcortical loops and encoded in subcortical structures for the rapid processing that would facilitate noise-invariant responses higher up.

Methods

Animals

Female C57BL/6JOlaHsd (Janvier, France) mice were used for experiments. All mice were 5-6 weeks old (Neuronexus experiments) or 8 weeks old (Neuropixels experiments) at the beginning of the experiment. Animals were housed in groups and 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 and performed in accordance with the Niedersächsisches Landesamt für Verbraucherschutz und Lebensmittelsicherheit, project license number 33.14-42502-04-10/0288 and 33.19-42502-04-11/0658; and in accordance with the Landesamt für Gesundheit und Soziales (LaGeSo), project license number G0140/19.

Apparatus: The Audiobox

The Audiobox is a device developed for auditory research and based on the Intellicage (NewBehavior, Switzerland). Mice were kept in groups of 6 to 10 animals. At least one day before experimentation, each mouse was lightly anesthetized with Avertin i.p. (0.1ml/10g), Ketamin/Xylazin (13 mg/mL / 1 mg/mL; 0.01 mL/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. Thus, each animal was individually identifiable through the use of the implanted transponder. The Audiobox served both as living quarters for the mice and as their testing arena.

The Audiobox was placed in a soundproof room which was temperature regulated and kept in a 12 h dark/light schedule. The apparatus consists of three parts, a food area, a ‘corner’, and a long corridor connecting the other two parts (Fig. 1A). The food area serves as the living quarter, where the mice have access to food *ad libitum*. Water is available either in two bottles situated in the corner or in the food area. The presence of the mouse in the ‘corner’, a ‘visit’, is

detected by an antenna located at the entrance of the corner. The antenna reads the unique transponder carried by each mouse as it enters the corner. A heat sensor within the corner senses the continued presence of the mouse. An antenna located at the entrance of the corner detects the transponder in the back of the mouse as it enters. The mouse identification is then used to select the correct acoustic stimulus. Once in the ‘corner’, specific behaviors (nose-poking and licking) can be detected through other sensors. All behavioral data is logged for each mouse individually. A loudspeaker (22TAF/G, Seas Prestige) is positioned right above the ‘corner’, for the presentation of the sound stimuli. During experimentation, cages and apparati were cleaned once a week by the experimenter.

Sound exposure in the Audiobox

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.

All mice had *ab libitum* food in the home cage, as well as free access to water through nose-poking in the corner, except the ‘no-water’ groups (see below) which had water in the food area instead. There was no conditioning in the Audiobox at any stage. All groups were first habituated to the Audiobox for 3 days without sound presentation. After the habituation phase, except for the control group, mice were exposed to 16 kHz pure tone pips in the corner with different exposure protocols.

The ‘predictable exposure’ group had the same sound exposure paradigm as in a previous study (Cruces-Solís et al., 2018). The mice were exposed to 16 kHz tone pips for the duration of each visit, regardless of nose-poke activity and water intake.

The ‘no-water predictable’ group was treated the same as the predictable exposure group, i.e., it was exposed to 16 kHz tone pips for the duration of each visit. However, there was no water in the corner, and water was accessible in the home cage.

For the ‘entry-only’ group, 16 kHz tone pips were triggered by the entry into the corner, but only lasted for 3 sec. For visits with a duration of less than 3 sec, the sound stopped once the mouse left the corner.

For the ‘unpredictable’ group, 16 kHz tone pips were presented in 80% of the visits with delay random between 5 to 25 seconds from the visit onset. Once the sound started, it continued until the mouse left the corner. Thus, the starting of the sound could not be predicted by the animal.

The control group lived in the Audiobox for the same amount of time as the other groups, but without sound presentation. All animals can hear environmental sounds, i.e., their own vocalization and the sound of the ventilation fan. All animals were in the exposure phase for at least 5 days. All animals were in the Audiobox in the exposure phase for at least 5 days: 6-12 days in the Neuronexus data set and 19-21 days in the Neuropixels data set.

Electrophysiology

After at least 5 days in the exposure phase, animals were removed from the apparatus one at a time and immediately anesthetized for electrophysiology. Animals were initially anesthetized with Avertin (0.15ml/10 g; Neuronexus recordings) or Ketamine/Xylazine (13 mg/mL / 1 mg/mL; 0.01 mL/g; Neuropixels recordings). Additional smaller doses of drug were delivered as needed to maintain anesthesia during surgery. The surgical level of anesthesia was verified by the pedal-withdrawal reflex. Body temperature was maintained at 36 □ with a feedback-regulated heating pad (ATC 1000, WPI, Germany).

Anesthetized mice were fixed with blunt ear bars on a stereotaxic apparatus (Kopf, Germany). Vidisic eye gel (Bausch + Lomb GmbH, Germany) was used to protect the eyes from drying out. Lidocaine (0.1 mL) was subcutaneously injected at the scalp to additionally numb the region before incision. The scalp was removed to expose the skull. Periosteum connective tissue that adheres to the skull was removed with a scalpel. The bone surface was then disinfected and cleaned with hydrogen peroxide. Bone suture junctions Bregma and Lamda were aligned. A metal head-holder was glued to the skull 1.0 mm rostral to Lambda with methyl methacrylate resin (Unifast TRAD, GC), or a combination of Histoacryl (B. Braun) and dental cement (Lang Dental and Kulzer). A craniotomy of 0.8 mm × 1.0 mm with the center 0.85 mm to the midline and 0.75 caudal to Lambda was made to expose the left inferior colliculus. After the craniotomy, the brain was protected with Saline (B. Braun, Germany). The area which is posterior to the transverse sinus and anterior to the sigmoid sinus was identified as the inferior colliculus.

The borders of the left inferior colliculus became visible after the craniotomy. The electrode was placed in the center of the inferior colliculus after measuring its visible rostrocaudal and mediolateral borders. The aim was to target the same position of the inferior colliculus across animals. Before lowering the electrode, a ground wire was connected to the neck muscle as a reference for the recordings. Neuronexus electrodes (single shank 16-channel silicon probe; 177 μm^2 area/site and 50 µm spacing between sites; Neuronexus, USA) were vertically lowered with a micromanipulator (Kopf, Germany) and slowly advanced (2-4 µm/sec, to minimize damage to the tissue) to a depth of 750 µm from the brain surface. The final depth was double-checked by making sure that the most dorsal channel was aligned with the dura (Fig. 1B). After one recording session, the electrode was further advanced 500 µm to a depth of 1250 µm. A final set of experiments was performed in Berlin using Neuropixels probes (Imec, Belgium, 1.0 with cap). These were inserted 3000 µm deep into the central nucleus of IC measured from the surface of the dura, at a speed of 62 µm per second using an electronic manipulator (Luigs & Neumann). A silver ground was attached to the probe and positioned at the caudal end of the skin incision.

During data acquisition with Neuronexus probes, the electrophysiological signal was amplified (HS-18-MM, Neuralynx, USA), sent to an acquisition board (Digital Lynx 4SX, Neuralynx, USA), and recorded with a Cheetah 32 Channel System (Neuralynx, USA). The voltage traces were acquired at a 32 kHz sampling rate with a wide band-pass filter (0.1±9,000 Hz). Neuropixels recordings were performed using a PXIe-6341 board (National instrument), sampled at 30 kHz, and recorded with open-source software SpikeGLX (http://billkarsh.github.io/SpikeGLX/).

Acoustic stimulation during recording

The sound was synthesized using MATLAB, produced by a USB interphase (Octa capture, Roland, USA), amplified (Portable Ultrasonic Power Amplifier, Avisoft, Germany), and played with a free-field ultrasonic speaker (Ultrasonic Dynamic Speaker Vifa, Avisoft, Germany). The speaker was 15 cm away from the right ear. The sound intensity was calibrated with a Bruël & Kjaer microphone. For measuring the frequency response area, we used sound stimuli which consisted of 30 ms pure tone pips with 5 ms rise/fall slopes repeated at a rate of 2 Hz. Thirty-two frequencies were used (3.4 kHz to 49.4 kHz, 0.125-octave spacing) at different intensities (0 dB to 80 dB with steps of 5 or 10 dB) played in a pseudorandom order. Each frequency-intensity

combination was played 5 times. For Neuropixels recordings, 51 frequencies were used (6.5 to 65 kHz), played at 70 dB SPL, each repeated 5 times in a given sweep. Two or three sweeps were played per recording.

Analysis of electrophysiological recordings

The recorded voltage signals were high-pass filtered at 500 Hz to remove the slow local field potential signal. To improve the signal-to-noise ratio of the recording, the common average reference was calculated from all the functional channels and subtracted from each channel (Ludwig et al., 2009).

For Neuronexus data multi-unit analysis, spikes were detected as local minima below a threshold of 6 times the median absolute deviation of each channel and later summed at 1ms bins. If the calculated value was higher than -40 µV, the threshold was set to -40 µV. Multi-unit sites were included in the analysis only if they were classified as sound-driven, defined by significant excitatory evoked responses (comparison of overall firing rate in 80 ms windows before and after stimulus onset for all frequencies and intensities; paired t-test, $p < 0.05$).

To analyze the sound-driven responses, signals were aligned at 200 ms before each stimulus onset. The temporal pattern of responses was compared using peri-stimulus time histograms (PSTHs). Multiunit spikes were averaged over an 80 ms window before and after each stimulus onset to calculate the spontaneous activity and evoked firing rates respectively.

Sound-driven responses at a given intensity as a function of frequency were used to generate the iso-intensity tuning curves. The combined frequency-intensity responses were used to generate the tonal receptive field (TRF).

The best frequency (BF) was selected as the frequency which elicited the highest responses when averaging across all intensities. In the rare cases where more than one frequency elicited the highest response, BF was calculated as the mean of those frequencies.

The excitation profile was characterized by computing the normalized response to a given frequency across different depths. The normalization was done by rescaling the tuning curve of each depth with a 0-max normalization.

Neural activities at the depth from 500 µm to 750 µm were measured in two recording sessions, first when the probe tip was at 750 µm depth and then when it was at 1250 µm. When analyzing responses at a given depth, recordings for the same depth were averaged.

Spike trains from the Neuropixels data set were sorted via the open-source platforms Kilosort3 and CatGT. Frequency tuning was derived using custom scripts built in MATLAB 2020a. Evoked activity was calculated as all tuned post-trigger spikes summed over an 80 ms time window. Only activity which was significantly different from an 80 ms pre-trigger window was considered evoked. A set of selection criteria were applied to these evoked units: 1) minimum spike amplitudes ≥ 18 μV, ~3x baseline noise (Steinmetz et al., 2018); 2) a minimum spike count of 25 across responses to all frequencies within a frequency sweep; 3) smooth tuning, meaning no more than 4 crossings of the median spike count line.

To remove the tail units from Neuropixels datasets in order to construct the mean responses plotted in Figure 3C, a two-cluster Gaussian mixture model (GMM; *fitgmdist*, MATLAB) initialized with the k-means algorithm was applied.

Statistical analysis

Group comparisons were made using multiple-way ANOVAs after testing for normality distribution using the Shapiro-Wilk test. Samples that failed the normality test were compared using a Kruskal-Wallis test or Wilcoxon signed rank test for paired datasets. 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 ANOVA. All multiple comparisons used critical values from a t distribution, adjusted by Bonferroni correction with an alpha level set to 0.05.

Certain analyses were performed using linear mixed-effects models (*fitlme*, MATLAB), as these models are appropriate for data with repeated measurements from the same subject. The mouse was used as the random effect. We checked whether the residual of the linear mixed-effects model was normally distributed. A significant test was performed by comparing the full model with a similar model with only random effects using the standard likelihood ratio test.

Means are expressed ± SEM. Statistical significance was considered if $p < 0.05$.

Acknowledgements. We are grateful to Markus Krohn, head of the technical workshop at the MPI-EM, and his team, as well as Harry Scherer, for technical support.

References

Bakin, J. S., & Weinberger, N. M. (1990). Classical conditioning induces CS-specific receptive field plasticity in the auditory cortex of the guinea pig. *Brain Research*, *536*(1-2), 271–286.

Bao, Y., Szymaszek, A., Wang, X., Oron, A., Pöppel, E., & Szelag, E. (2013). Temporal order perception of auditory stimuli is selectively modified by tonal and non-tonal language environments. *Cognition*, *129*(3), 579–585.

Bieszczad, K. M., & Weinberger, N. M. (2010). Representational gain in cortical area underlies increase of memory strength. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(8), 3793–3798.

Blake, D. T., Strata, F., Churchland, A. K., & Merzenich, M. M. (2002). Neural correlates of instrumental learning in primary auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(15), 10114–10119.

Brandewie, E. J., & Zahorik, P. (2018). Speech intelligibility in rooms: Disrupting the effect of prior listening exposure. *The Journal of the Acoustical Society of America*, *143*(5), 3068.

Chen, C., Krueger-Burg, D., & de Hoz, L. (2019). Wide sensory filters underlie performance in memory-based discrimination and generalization. *PloS One*, *14*(4), e0214817.

Chen, C., Trinh, F., Harper, N., & de Hoz, L. (2021). Natural sound characteristics explain perceptual categorization. In *bioRxiv* (p. 2021.09.29.462467). https://doi.org/10.1101/2021.09.29.462467

Claverol-Tinture, E., & Nadasdy, Z. (2004). Intersection of Microwire Electrodes With Proximal CA1 Stratum-Pyramidale Neurons at Insertion for Multiunit Recordings Predicted by a 3-D Computer Model. In *IEEE Transactions on Biomedical Engineering* (Vol. 51, Issue 12, pp.

2211–2216). https://doi.org/10.1109/tbme.2004.834274

Cruces-Solís, H., Jing, Z., Babaev, O., Rubin, J., Gür, B., Krueger-Burg, D., Strenzke, N., & de Hoz, L. (2018). Auditory midbrain coding of statistical learning that results from discontinuous sensory stimulation. *PLoS Biology*, *16*(7), e2005114.

de Hoz, L., Barniv, D., & Nelken, I. (2022). Prior unsupervised experience leads to long lasting effects in sensory gating during discrimination learning. In *bioRxiv* (p. 2022.09.30.510296). https://doi.org/10.1101/2022.09.30.510296

de Hoz, L., & Nelken, I. (2014). Frequency tuning in the behaving mouse: different bandwidths for discrimination and generalization. *PloS One*, *9*(3), e91676.

de Villers-Sidani, E., Chang, E. F., Bao, S., & Merzenich, M. M. (2007). Critical period window for spectral tuning defined in the primary auditory cortex (A1) in the rat. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *27*(1), 180–189.

Gao, E., & Suga, N. (2000). Experience-dependent plasticity in the auditory cortex and the inferior colliculus of bats: role of the corticofugal system. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(14), 8081–8086.

Khalighinejad, B., Herrero, J. L., Mehta, A. D., & Mesgarani, N. (2019). Adaptation of the human auditory cortex to changing background noise. *Nature Communications*, *10*(1), 2509.

Kiang, N. Y., & Moxon, E. C. (1974). Tails of tuning curves of auditory-nerve fibers. *The Journal of the Acoustical Society of America*, *55*(3), 620–630.

Letzkus, J. J., Wolff, S. B. E., Meyer, E. M. M., Tovote, P., Courtin, J., Herry, C., & Lüthi, A. (2011). A disinhibitory microcircuit for associative fear learning in the auditory cortex. *Nature*, *480*(7377), 331–335.

Ludwig, K. A., Miriani, R. M., Langhals, N. B., Joseph, M. D., Anderson, D. J., & Kipke, D. R. (2009). Using a common average reference to improve cortical neuron recordings from microelectrode arrays. *Journal of Neurophysiology*, *101*(3), 1679–1689.

McPherson, M. J., & McDermott, J. H. (2020). Time-dependent discrimination advantages for

harmonic sounds suggest efficient coding for memory. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(50), 32169–32180.

Mesgarani, N., David, S. V., Fritz, J. B., & Shamma, S. A. (2014). Mechanisms of noise robust representation of speech in primary auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(18), 6792–6797.

Norman-Haignere, S. V., & McDermott, J. H. (2018). Neural responses to natural and model-matched stimuli reveal distinct computations in primary and nonprimary auditory cortex. *PLoS Biology*, *16*(12), e2005127.

Rabinowitz, N. C., Willmore, B. D. B., Schnupp, J. W. H., & King, A. J. (2012). Spectrotemporal contrast kernels for neurons in primary auditory cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *32*(33), 11271–11284.

Reber, A. S. (1967). Implicit learning of artificial grammars. In *Journal of Verbal Learning and Verbal Behavior* (Vol. 6, Issue 6, pp. 855–863). https://doi.org/10.1016/s0022-5371(67)80149-x

Steinmetz, N. A., Koch, C., Harris, K. D., & Carandini, M. (2018). Challenges and opportunities for large-scale electrophysiology with Neuropixels probes. *Current Opinion in Neurobiology*, *50*, 92–100.

Yan, J., Zhang, Y., & Ehret, G. (2005). Corticofugal shaping of frequency tuning curves in the central nucleus of the inferior colliculus of mice. *Journal of Neurophysiology*, *93*(1), 71–83.

Zhang, Y., & Suga, N. (2005). Corticofugal feedback for collicular plasticity evoked by electric stimulation of the inferior colliculus. *Journal of Neurophysiology*, *94*(4), 2676–2682.

Figure legends

Fig. 1: Exposure to predictable sound, independently of reinforcer, elicits IC plasticity

A, Schematic representation of the Audiobox design for the control (left), predictable (middle), and no-W (no-water) predictable exposure group (right), respectively. Food was *ad libitum* in the food area. Water was available in the corner (control and predictable group) or in the food area (no-W group). **B,** Schematic representation of recording using NeuroNexus 16-electrode arrays from two positions. Insert: Schematic representation of the positioning of the most superficial recording site, aligned to the dura. **C,** Average tuning curves of simultaneously recorded sound-evoked activity at 70 dB for the depth of 350, 550, and 700 µm in the IC of the control (blue), predictable (red), and no-W predictable (pink) group. **D,** Left: Average best frequency (BF) as a function of recording depths (two-way ANOVA, group $F(2, 761) = 6.88$, $p = 0.001$, depth $F(25, 761) = 211.4$, $p = 0$, interaction $F(59, 761) = 0.85$, $p = 0.77$). Right: Difference in BF relative to the average BF of the control group across the depths between 100 and 900 µm (two-way ANOVA, group $F(2, 525) = 35.64$, $p = 0$, depth $F(16, 525) = 0.55$, $p = 0.92$, interaction $F(32, 525) = 0.52$, $p = 0.99$. Corrected pair comparisons: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$). **E,** Schematic representation of the sound exposure protocol. **F,** Average tuning curves for the depth of 550 µm in the IC. **G,** Average BF for selected depths in the IC. Error bars represent standard error.

Fig. 2: Predictability of the contextual sound triggers plasticity

A, Schematic representation of recording using Neuropixels probes. Insert: Illustration of high-density Neuropixels probes. **B,** BF of individual units recorded from different depths of the IC. **C,** Average BF as a function of depths for units along the tonotopic axis. Linear mixed effect model with two factors [group × depth]: group $p = 0.016$, depth $p = 0.002$, interaction $p = 0.013$. **D,** Average change in BF relative to the average BF of the control group was frequency specific. **E,** Similar to **D.** for recording using Neuropixels probes. **F,** Table illustrating the sound-context association pattern in different groups depending on whether there was water in corner (top), sound triggered by visit onset (middle), sound for most of the visit (bottom). **G,** Average daily time spent in the corner of individual mice. Groups without water in the corner spent less time in the corner compared to groups with water in the corner (two-sample t-test, ****$p < 0.0001$). **H,** Average daily length of sound exposure of individual mice.

Fig. 3: BF Shifts were accompanied by an increase in frequency discrimination ability

A, Relative response magnitude to a given frequency across the IC. Sound-evoked responses were rescaled with a 0-max normalization based on the tuning curve of each depth. **B,** Frequency discrimination ability measured by Euclidean distance between the excitatory profile of frequency pairs. **C,** Frequency discrimination ability for a given frequency to other frequencies. **D,** frequency discrimination ability for a given delta F as a function of sound frequency.

Supplementary Fig. 1: Predictable sound-context association induced a homogeneous increase in response gain.

A, Schematic representation of the sound exposure protocol for the control and unpredictable group. **B,** Average tuning curves of simultaneously recorded sound-evoked activity at 70 dB for the depth of 350, 550, and 700 µm in the IC of the control (blue) and unpredictable (dark green) group. **C,** Average best frequency (BF) as a function of recording depths (linear mixed effect model with the group and depth factor: $p_{group} = 0.63$, $p_{depth} < 0.0001$, $p_{group,\ depth} = 0.64$). Insert: Difference in BF relative to the average BF of the control group (two-sample t-test, $p = 0.75$). **D,E,F,** Similar to **A,B,C,** for the comparison between the predictable and entry-only group. There was no shift in BF (linear mixed effect model with the group and depth factor: $p_{group} = 0.39$, $p_{depth} < 0.0001$, $p_{group,\ depth} = 0.04$) and no significant difference in BF change (two-sample t-test, $p = 0.81$). **G,** Average evoked responses to frequencies at 70 dB as a function of recording depth (linear mixed effect model with the group and depth factor: $p_{group} = 0.02$, $p_{depth} = 0.0003$, $p_{group,\ depth} = 0.80$). **H,I,**Similar to **G,** show the response gain for the control vs. unpredictable (linear mixed effect model with the group and depth factor: $p_{group} = 0.53$, $p_{depth} < 0.0001$, $p_{group,\ depth} = 0.72$) and predictable vs. entry-only (linear mixed effect model with the group and depth factor: $p_{group} = 0.69$, $p_{depth} < 0.0001$, $p_{group,\ depth} = 0.006$), respectively.

Figure 1

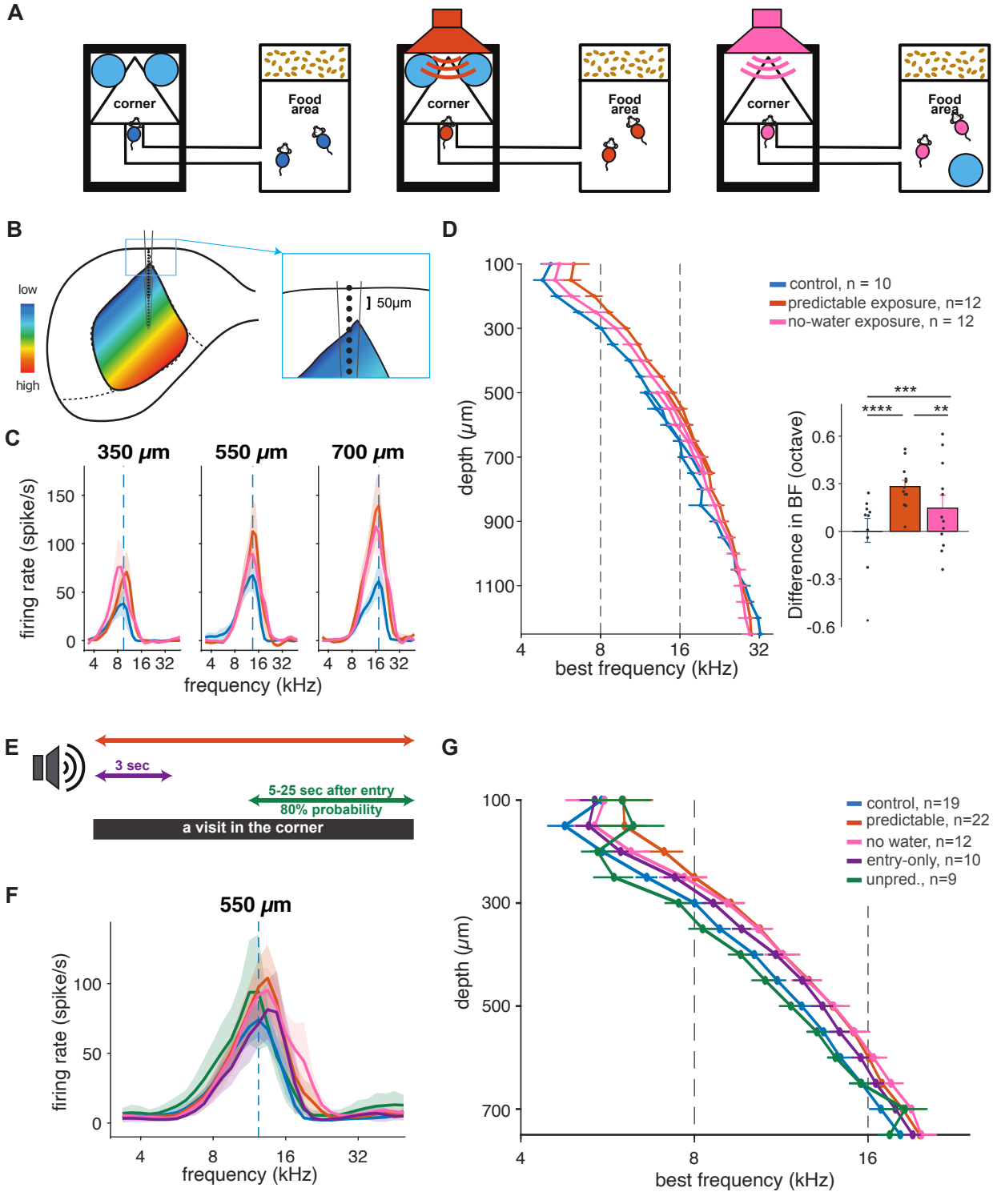

A
corner
Food area
corner
Food area
corner
Food area
B
low
high
50µm
C
350 µm
550 µm
700 µm
firing rate (spike/s)
150
100
50
0
4 8 16 32
frequency (kHz)
D
control, n = 10
predictable exposure, n=12
no-water exposure, n = 12
depth (µm)
100
300
500
700
900
1100
best frequency (kHz)
Difference in BF (octave)
0.6
0.3
0
-0.3
-0.6
***
****
**
E
3 sec
5-25 sec after entry
80% probability
a visit in the corner
F
550 µm
firing rate (spike/s)
100
50
0
frequency (kHz)
G
control, n=19
predictable, n=22
no water, n=12
entry-only, n=10
unpred., n=9
depth (µm)
100
300
500
700
best frequency (kHz)

Figure 2

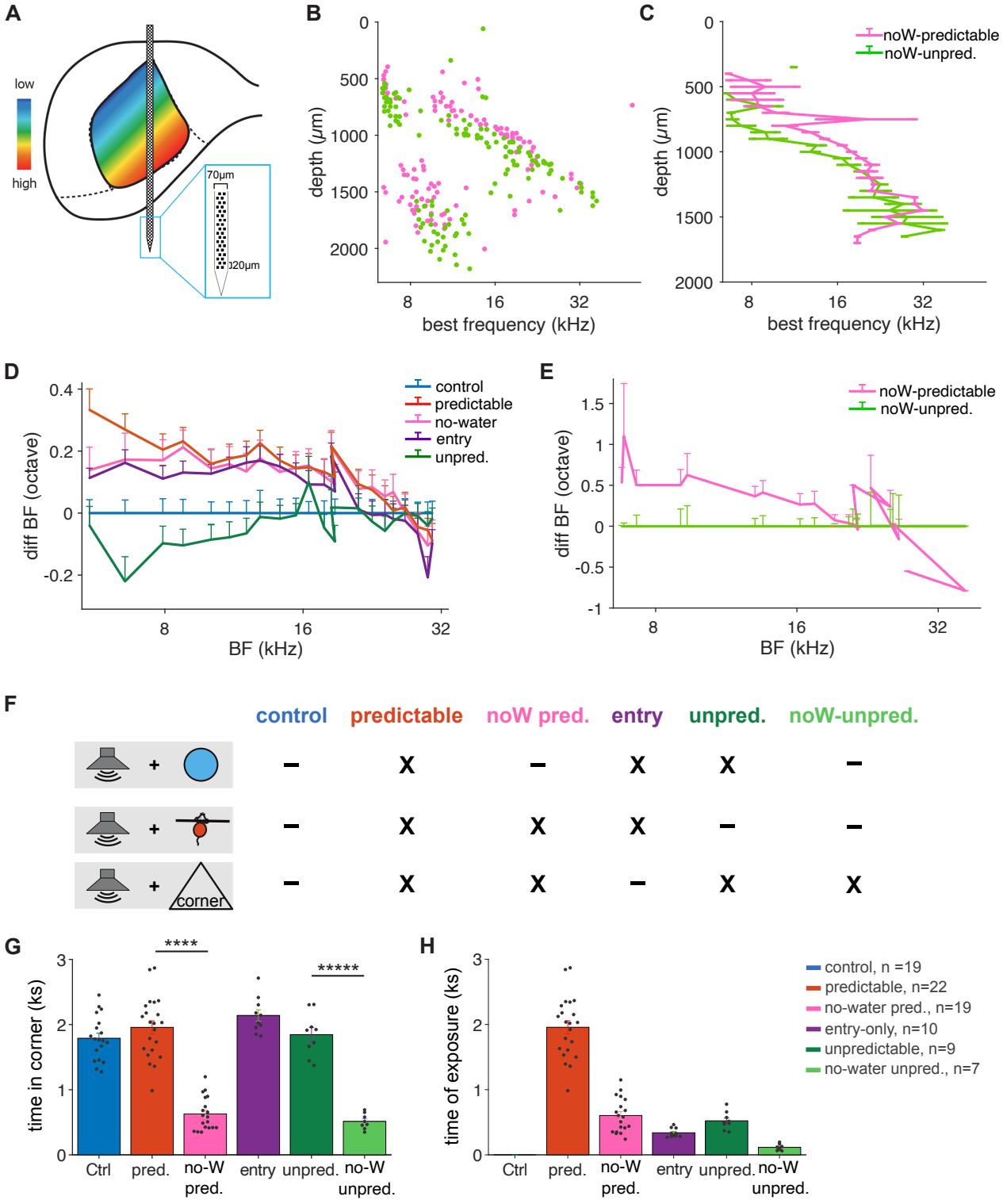
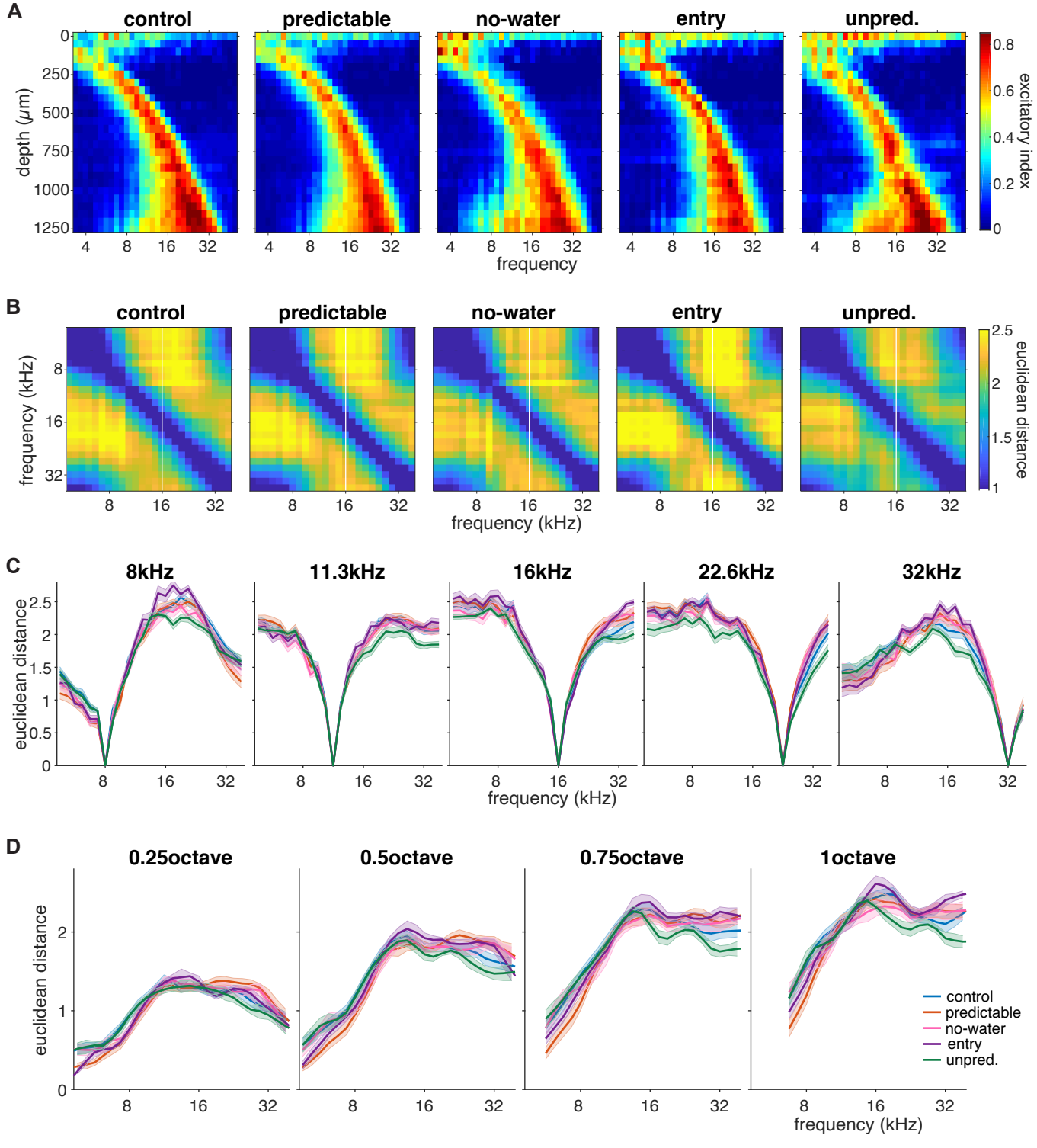

A
low
high
70µm
20µm
B
depth (µm)
best frequency (kHz)
C
noW-predictable
noW-unpred.
depth (µm)
best frequency (kHz)
D
control
predictable
no-water
entry
unpred.
diff BF (octave)
BF (kHz)
E
noW-predictable
noW-unpred.
diff BF (octave)
BF (kHz)
F
control
predictable
noW pred.
entry
unpred.
noW-unpred.
corner
G
time in corner (ks)
****
*****
Ctrl
pred.
no-W pred.
entry
unpred.
no-W unpred.
H
time of exposure (ks)
control, n =19
predictable, n=22
no-water pred., n=19
entry-only, n=10
unpredictable, n=9
no-water unpred., n=7
Ctrl
pred.
no-W pred.
entry
unpred.
no-W unpred.

Figure 3


A
control
predictable
no-water
entry
unpred.
depth (μm)
0
250
500
750
1000
1250
4
8
16
32
frequency
excitatory index
0.8
0.6
0.4
0.2
0
B
control
predictable
no-water
entry
unpred.
frequency (kHz)
8
16
32
frequency (kHz)
euclidean distance
2.5
2
1.5
1
C
8kHz
11.3kHz
16kHz
22.6kHz
32kHz
euclidean distance
2.5
2
1.5
1
0.5
0
8
16
32
frequency (kHz)
D
0.25octave
0.5octave
0.75octave
1octave
euclidean distance
2
1
0
8
16
32
frequency (kHz)
control
predictable
no-water
entry
unpred.