

Novel Tools and Methods

Automated Behavioral Experiments in Mice Reveal Periodic Cycles of Task Engagement within Circadian Rhythms

ID **Nikolas A. Francis, Kayla Bohlke, and ID Patrick O. Kanold**<https://doi.org/10.1523/ENEURO.0121-19.2019>

Department of Biology, University of Maryland, College Park, Maryland 20742

Abstract

High-throughput automated experiments accelerate discovery in neuroscience research and reduce bias. To enable high-throughput behavioral experiments, we developed a user-friendly and scalable automated system that can simultaneously train hundreds of mice on behavioral tasks, with time-stamped behavioral information recorded continuously for weeks. We trained 12 cages of C57BL/6J mice (24 mice, 2 mice/cage) to perform auditory behavioral tasks. We found that circadian rhythms modulated overall behavioral activity as expected for nocturnal animals. However, auditory detection and discrimination accuracy remained consistently high in both light and dark cycles. We also found a periodic modulation of behavioral response rates only during the discrimination task, suggesting that the mice periodically reduce task engagement (i.e., take “breaks”) when task difficulty increases due to the more complex stimulus–response paradigm for discrimination versus detection. Our results highlight how automated systems for continuous high-throughput behavioral experiments enable both efficient data collection and new observations on animal behavior.

Key words: auditory; circadian; high throughput; home cage; operant conditioning

Significance Statement

Automated high-throughput behavioral experiments in mice promise researchers the ability to quickly and reliably assess the behavior of large animal populations, while also minimizing experimenter-induced bias. However, the technical complexities of automation have limited widespread adoption of high-throughput behavioral methods. Here, we present a new tool for behavioral research, the ToneBox, which allows both novice and expert behaviorists to automatically test hundreds of mice simultaneously on different behavioral tasks. We provide manufacturing specifications and detailed documentation of system operation. Using C57BL/6J mice, we show that the ToneBox tracks circadian cycles via behavioral response rates, and that task difficulty modulates duty cycles of task engagement.

Introduction

Targeted investigation of the links between genetics, the brain, and behavior has seen rapid advancement because of powerful new tools for high-throughput experiments (Rose et al., 2016; Sofroniew et al., 2016; Kuchib-

hotla et al., 2017; Francis et al., 2018). Automated behavioral experiments (Gess et al., 2011; Schaefer and Claridge-Chang, 2012; Poddar et al., 2013; de Hoz and Nelken, 2014; Francis and Kanold, 2017; Balzani et al., 2018) in mice will enable the advancement of neurosci-

Received March 27, 2019; accepted August 26, 2019; First published September 5, 2019.

Author contributions: N.A.F. and P.O.K. designed research; N.A.F. and K.B. performed research; N.A.F. contributed unpublished reagents/analytic tools; N.A.F. analyzed data; N.A.F. and P.O.K. wrote the paper.

This research was supported by Department of Health and Human Services (DHHS) | National Institutes of Health (NIH) | National Institute on Deafness and Other Communication Disorders Grant RO1-DC-009607; DHHS | NIH | National Institute of Neurological Disorders and Stroke Grant U01-NS-090569; and DHHS | NIH | National Institute of Neurological Disorders and Stroke Grant U19-NS-107464.

ence research by efficiently producing large high-quality datasets that bypass the limitations of manual behavioral methods, such as daily animal handling at suboptimal times during the circadian rhythm—which may stress animals and increase behavioral variability across experiments (Balcombe et al., 2004). Automation removes experimenter-induced variability and makes more efficient use of both human and animal resources. However, technical complexities such as calibrated stimulus presentation, scalable automation, and big data analysis have limited the widespread adoption of high-throughput automated behavioral methods across neuroscience laboratories.

Given the potential for great scientific advancement, we developed the “ToneBox,” a scalable and user-friendly automated system for 24/7 behavioral training in mice within their home cage. The ToneBox does not require new users to have deep technical knowledge to build and operate it, yet the ToneBox also enables accurate stimulus calibration and informative analysis of large behavioral datasets. ToneBox specifications and operation are detailed within both the Materials and Methods and an on-line repository. The repository contains a manual with parts lists and comprehensive step-by-step instructions for assembly and operation. We also provide models for 3D printing and manufacturing of printed circuit boards (PCBs), as well as software for system operation and data analysis.

Using the ToneBox, we found that while circadian rhythms modulated overall behavioral activity, as expected for nocturnal animals such as mice, auditory detection and discrimination accuracy remained consistently high in both light and dark cycles. We compared task performance during an easy task (pure-tone detection) versus a hard task (pure-tone frequency discrimination) and found a periodic modulation of behavioral response rates only during the more difficult task.

Materials and Methods

Animals

Twenty-four C57BL/6J mice (12 male, 12 female) were trained simultaneously on auditory tasks in 12 home cages (two mice per cage). Mice were received at postnatal day 30 (P30). All mice were maintained on a 12 h (5:00 P.M. to 5:00 A.M.) light cycle and a 12 h (5:00 A.M. to 5:00 P.M.) dark cycle, and were provided food *ad libitum*. The mice were visually inspected every day (see below), and cages were cleaned weekly. All experimental procedures were approved by the University of Maryland Animal Care and Use Committee.

Correspondence should be addressed to Patrick O. Kanold at Pkanold@umd.edu.

<https://doi.org/10.1523/ENEURO.0121-19.2019>

Copyright © 2019 Francis et al.

This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

Hardware

Each ToneBox (Fig. 1) consists of the following seven components: (1) a computer with Matlab and a Wi-Fi router; (2) a ToneBox central control unit (CCU) with Raspberry Pi and custom ToneBox shield; (3) a USB (universal serial bus) sound card; (4) a behavioral interface (BI) that contains a speaker and a touch-responsive waterspout; (5) a sound attenuating box; (6) a home cage; and (7) a water delivery system with solenoid valves. A detailed list of parts, 3D models, a PCB file, and assembly instructions are available in the on-line public repository.

In brief, the CCU connects to both the BI and water delivery system via 3.5 mm TRS (tip, ring, and sleeve) cables. The Raspberry Pi is powered by a micro USB cable, and connected to a computer using a local Wi-Fi network. The speaker in the BI is driven by a 2.5 W class D audio amplifier (catalog #PAM8302, Adafruit) that receives audio from the USB audio device connected to the CCU. Water delivery is controlled by a normally closed 12 V solenoid valve (catalog #BCBI6683, American Science & Surplus) that is connected to a water reservoir through a manual valve. The solenoid connects to the CCU using a TRS cable. Since USB 2.0 only provides 5 V power, a 12 V input to the CCU is connected to a relay switch controlled by the Raspberry Pi. The CCU also contains capacitive touch sensors (catalog #AT42QT1012, Adafruit) that receive input from a TRS cable connected to the waterspout in the BI. The BI consists of two main components: a water-resistant speaker for sound delivery (catalog #AS02708CO-WR-R, PUI Audio), and a custom 3D printed stainless steel waterspout. The BI is hung on the side of a standard 13 × 7 × 5 inch home cage using a clip on the back of the BI. A small rectangle the size of the BI must be cut out of the cage lid to fit the BI in the home cage. The BI speaker was calibrated *in situ* using a Brüel & Kjær 4944-A microphone. Calibration was performed by whitening the recording of a 1–45 kHz noise played through the speaker, then applying the whitening filter to pure tones used in the task (Fig. 1c). The BI, CCU, and home cage sit within a sound-attenuating chamber. This is simply a cabinet with a clear door and a ventilating fan. The cabinets provide ≥20 dB attenuation between ToneBoxes for tones >1 kHz, allowing simultaneous auditory training to be possible for multiple cages. Greater attenuation is possible with additional sound-proofing customization of the cabinets. The waterline is run from the source, into the cabinet, and then finally down to the waterspout inside the BI. We placed 12 ToneBoxes on a metal rack inside our animal housing facility, so that experiments could run 24 h/d.

Software

The ToneBox is a hardware system that can be controlled by any software that has drivers for the Raspberry Pi 3B. Here, we ran the ToneBox on a Windows 10 PC with a custom GUI (graphical user interface; “ToneControl”) written in the Matlab programming language (MathWorks). The user can select up to 16 ToneBoxes, although the software could be customized for more ToneBoxes. After selecting a ToneBox, the user chooses a training phase: habitua-

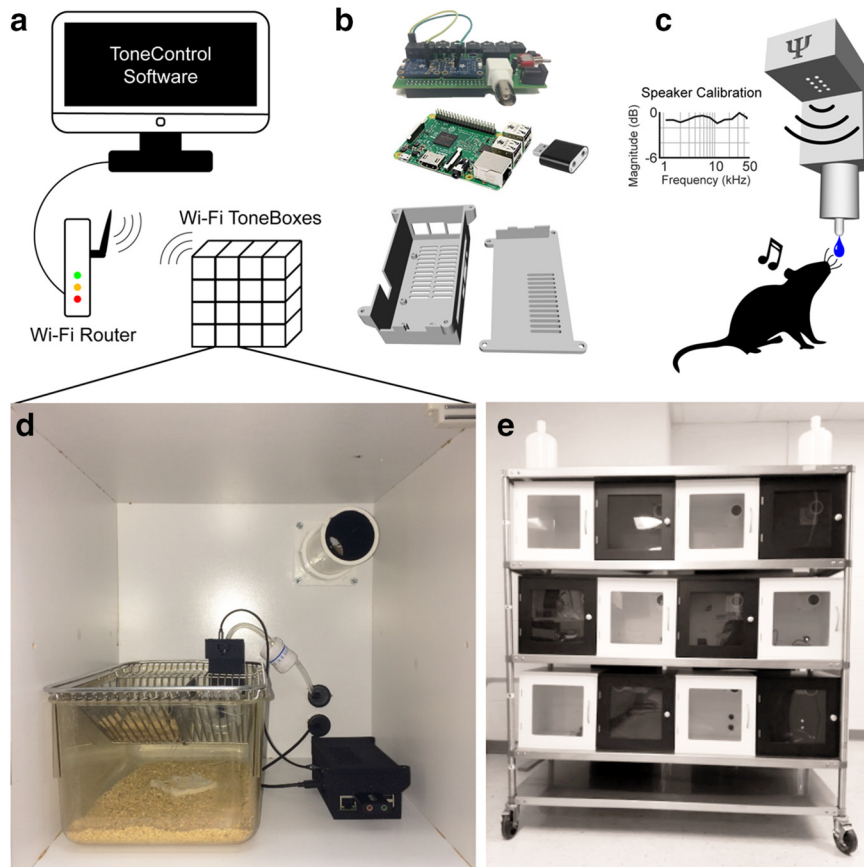


Figure 1. The ToneBox system for automated auditory operant conditioning in the mouse home cage. **a**, ToneControl software runs from the Matlab environment on a desktop computer. A Wi-Fi router is used to connect the computer to each Wi-Fi ToneBox. **b**, The ToneBox CCU. Bottom, A 3D model of the CCU case. Inside each CCU is a Raspberry Pi 3+ (middle) that is connected to a custom shield (top) with two capacitive touch sensors used to monitor behavioral responses. A USB sound card (middle) is used for audio input and output. **c**, The ToneBox BI: 3.5 mm audio cables connect the CCU to the BI within each home cage. An overhead speaker in the BI presents sounds. Pure tones of 1–45 kHz were calibrated to <1 dB magnitude variability. The waterspout at the base of the BI is connected to the CCU capacitive touch sensor to detect licks. **d**, The CCU, BI, and home cage are placed within an actively ventilated enclosure. The BI is mounted on the side of the cage. The BI is designed to place the waterspout inside the cage, while keeping the speaker outside to avoid damage from exploratory behavior by the mice. Water tubing enters through the back of the enclosure, connecting a flow regulator to the waterspout. **e**, Twelve enclosures with clear doors are shown stacked together on a mobile rack, with water supplies.

tion, shaping, detection, or discrimination. Next, the user selects the target and nontarget (for discrimination) pure-tone frequencies and levels. Silent probe trials can also be selected. Before starting experiments, the user can test the water and sound presentation using GUI buttons. In addition, the user can choose to calibrate the speaker.

We also developed a “ToneGraph” GUI for data analysis and for monitoring system status. The user can plot all data from each cage separately or take averages across cages. The user can also select subsets of trials for analysis. ToneControl, ToneGraph, and a software manual are available in the on-line public repository together with a test dataset to learn its function.

Behavioral training

For our initial experiments, we trained 12 cages of mice (two mice per cage) simultaneously. Individual mice, and a cage of 4 mice were also trained. Each group of cage mates was trained using positive rein-

forcement to detect a pure tone to receive water. Cages were housed inside sound-attenuating chambers that were arranged on shelves in our animal housing facility. Training consisted of the following four sequential phases: habituation, shaping, detection, and discrimination. Water bottles were removed from each cage, and the mice were water restricted for 23 h before habituation. To motivate task acquisition, water was made available only through task performance. Training ran on a 1 h duty cycle, 24 h/d. Each behavioral response (i.e., a lick) was recorded with a 30 ms resolution (~33 Hz sampling rate).

Habituation

Habituation is used to acclimate the mice to drinking from the BI waterspout without sounds being presented. Behavioral trials occurred at random intervals between 30 and 300 s. Water was released from the waterspout for 5 s during each trial. Habituation typically lasted 4 d.

Shaping

After the average lick rate for a cage showed sustained licking during habituation, they were moved on to shaping, in which the mice begin to associate hearing pure tones and licking the waterspout to receive water. Each trial was 4 s. The first second was silent, followed by a 1 s 11.3 kHz pure tone, and ending with 2 s of silence. Intertrial intervals (ITIs) were randomized between 5 and 9 s. The mice were required to refrain from licking for at least 5 s during the ITI to initiate the next trial. The first lick made during a 3 s behavioral response window after a tone onset (i.e., a hit response) was rewarded with water. On 20% of trials water was automatically released 0.5 s after the tone onset. Water was released on the remaining 80% of trials only after a behavioral hit. Shaping typically lasted 3–4 d.

Detection

Once the average lick rate began to increase after tone presentation, and once the lick latency distribution showed first licks regularly occurring after tone onset, we advanced the mice to a pure-tone detection task. The trial structure remained the same as in shaping; however, the freely available water on 20% of trials was removed and a response made during the 1 s silence before the tone (i.e., “early” responses) resulted in a punishment. If a mouse licked the waterspout during the 1 s silence, then the tone was still presented, but no water was delivered for subsequent licks, and a 20 s timeout was added to the ITI. The mice were also required to refrain from licking for at least 5–25 s during the ITI to initiate the next trial. The punishment of early licks and the 5–25 s wait period teach the mice to carefully control licking as to indicate tone detection. We also collected audiograms by randomizing the pure-tone frequency (4–45 kHz) and level (30–60 dB SPL).

Discrimination

We tested pure-tone frequency discrimination by training mice to lick in response to a target tone frequency (5 kHz) and to refrain from licking in response to a nontarget tone frequency (2 kHz). The mice were punished with a 20 s time-out for licking after the onset of a nontarget tone. All other task contingencies remained the same as in detection. This version of a go/no-go discrimination task functions as a stop-signal task, wherein the false alarm rate indicates the ability of the mouse to inhibit licking the waterspout in response to hearing the nontarget. An auditory reversal task could also be implemented using the ToneBox discrimination task. Reversal tasks are important for measuring cognitive flexibility. The user simply reverses the frequency parameters for the target and nontarget tones.

Behavioral response rates

Early rates were calculated as the percentage of total trials in which the first lick was during the 1 s prestimulus silence. Hit and miss rates were calculated as the percentage of target tone trials with or without a first lick during the behavioral response window, respectively. False alarm and correct rejection rates were calculated as the percentage of nontarget tone trials with or without a

first lick during the behavioral response window, respectively.

Health monitoring

During detection and discrimination trials, the mice received 0.075 ml/hit. Given that a cage of two mice does ~64 trials/d (Fig. 2b), each mouse received ~2.4 ml of water/d, if divided evenly between the pair of mice in a cage. Since our system does not track individual mice, we cannot determine the exact distribution of water per mouse. We inspected the mice daily to ensure normal hydration by monitoring skin turgor, appearance, and weight loss. Skin turgor and appearance were rated daily on a scale of 1–5 as follows: (1) normal skin turgor and posture; (2) skin turgor present (<1 s); (3) hunched posture, piloerection, skin turgor 1–2 s; (4) eyes sunken, severe fuzzy facial fur, skin turgor up to 3 s; and (5) failure to right itself or grasp cage bars, acts paralyzed. Intervention to correct the health of the animal would occur after a rating of 3; however, we did not observe any rating >2 throughout the study. All mice were visually inspected every day and appeared to be healthy and active. To confirm the health of the mice, we weighed each mouse in each cage once a day for 5 d of detection trials after a few months of continuous training. The average weights (\pm SD) for each day were as follows: 30.3 \pm 6.6, 30.5 \pm 6.6, 30.9 \pm 6.7, 31.0 \pm 6.5, and 31.0 \pm 6.5 g. The average weight over 5 d was 30.7 \pm 6.0 g/mouse, which is typical for adult C57BL/6J mice. Since the mice remained healthy during our study, and task performance was the only means of getting water, we presume that all animals performed the task.

Statistical comparisons

All statistical testing was done in Matlab software using a two-tailed nonparametric Kolmogorov–Smirnov (KS) test ($n = 12$ cages). The KS test is appropriate since we only sought nonparametric pairwise comparisons of empirical behavioral response distributions.

Code accessibility

The code/software described in the article is freely available on-line at <https://github.com/KanoldLab/Tone-Box>. The code is available as [Extended Data 1](#). The results of this study were obtained using the Windows 10 operating system on a PC.

Results

The ToneBox uses freely programmable sounds to train mice on operant conditioning tasks (Fig. 1). For our current use, we programmed the system to present pure tones (Materials and Methods). We used auditory tasks because the auditory system is crucial for human and animal communication, auditory deficits are associated with many sensory and cognitive diseases (Corcoran et al., 2002; Ewing and Grace, 2013; Zhou et al., 2015), and mice are readily trained on auditory tasks (Kurt and Ehret, 2010; Gess et al., 2011; Poddar et al., 2013; de Hoz and Nelken, 2014; Francis and Kanold, 2017). The ToneBox can easily be augmented to allow for different stimulus modalities using the additional general purpose

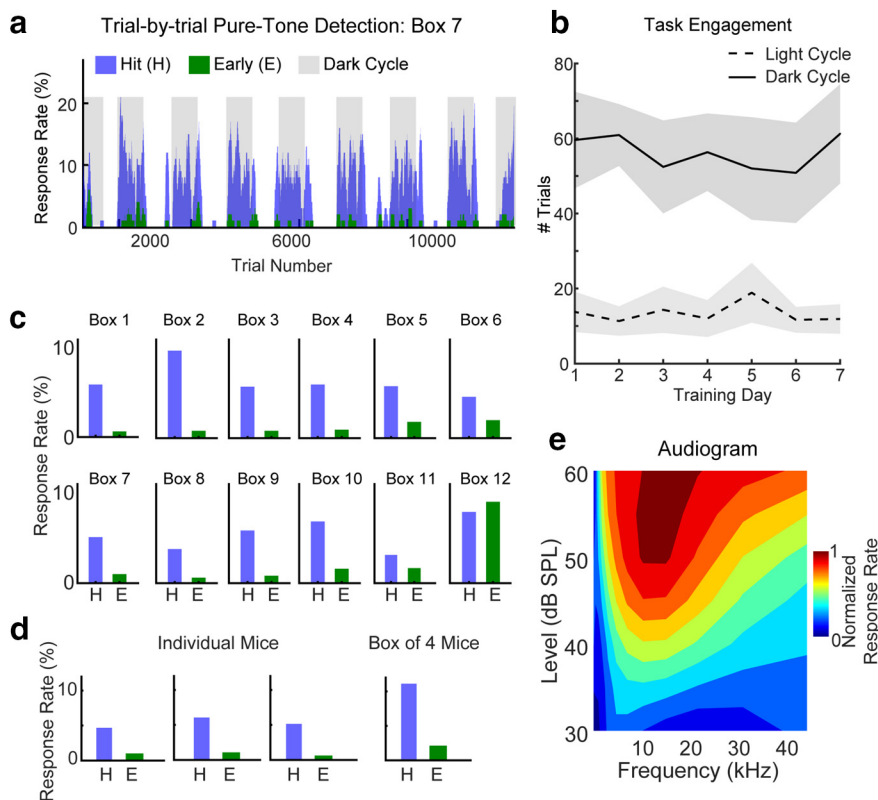


Figure 2. Auditory detection task performance in mice trained using ToneBoxes. **a**, Example of behavioral data collected over 9 consecutive days in a single box of two mice. Response rates were calculated using a 25-trial sliding window. **b**, Task engagement was defined as the average number of trials with at least one lick per trial. The solid and dotted lines show the dark and light cycles, respectively. The Shading shows ± 2 SEM. **c**, Hit and early response rates for each of the 12 tested boxes, color coded as in **a**. **d**, Hit and early response rates for three individual mice that were initially trained in pairs, then isolated and trained alone (left three panels), and a box of four mice initially trained together (right). Data color coded as in **a**. **e**, Pure-tone detection with roving tone frequency and level. Hot and cool colors indicate high and low hit rates, respectively.

digital output ports to control, for example, an LED. Each ToneBox is composed of a sound-attenuating box (≥ 20 dB attenuation between boxes for tones > 1 kHz) that contains the home cage, a BI, and a CCU (Fig. 1*a–d*). The water delivery system routes a water line into the waterspout of each BI. The ToneBox was intentionally designed for scalability and use in animal facilities; thus, it is low cost, contains Wi-Fi connectivity, and allows easy integration with existing home cages. Thus, we set up 12 ToneBoxes in our animal housing facility (Fig. 1*e*). For initial experiments (Figs. 2–4), we placed two P30 C57BL/6J same-sex littermates into each ToneBox, totaling 24 mice (12 male, 12 female).

The ToneBox is operated with ToneControl software in Matlab. ToneControl uses a private wireless network to detect and interact with individual CCUs. A flat acoustic spectrum is critical for accurate testing of auditory function. We thus included automatic speaker calibration in ToneControl. Calibration only requires the user to have a microphone and an amplifier. We used Brüel & Kjær 4944A and 1704-A-001 equipment, but most ultrasonic amplifiers with an analog output will work. Ultrasonic calibration is possible up to ~ 45 kHz using the specified BI and CCU components, and a sound card with at least a 96 kHz sampling rate.

A primary goal of automation is to acquire reliable results when testing large animal populations, which is, for example, required to identify potentially subtle effects of pharmacological, genetic, or environmental manipulations. Here we compared task performance during light versus dark cycles in the animal housing facility for both pure-tone detection and frequency discrimination tasks. In brief, during each trial of pure-tone detection, the trial began with a 1 s silence, followed by a 1 s tone, and then a 2 s silence. The intertrial interval was randomized between 5 and 9 s. If a mouse licked the waterspout after the target tone onset (i.e., a “hit” response), then water was delivered for 2 s. If the mouse licked during the 1 s silence before the tone (i.e., an “early” response), the target was still presented and licks were still recorded, but no water was delivered and the mouse was given a 20 s time-out. Between trials, the mouse was required to refrain from licking the waterspout for 5–25 s before the next trial began. The distinction between the ITI and the lick-refrain period was only that licking during the ITI per se did not affect when the next trial began. For pure-tone frequency discrimination tasks, either a target or a nontarget tone was presented with equal likelihood (i.e., 50%) in each trial. A lick after a nontarget (i.e., a “false alarm” response) was punished with a 20 s time-out, while all other task

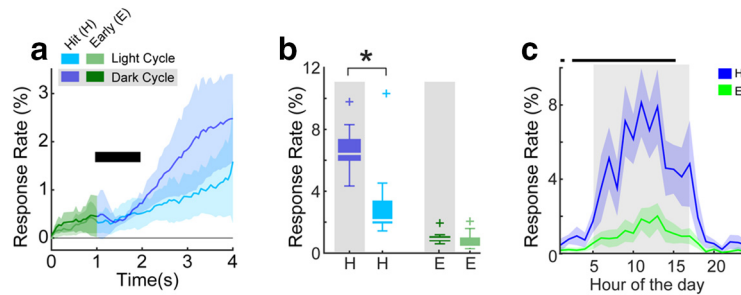


Figure 3. Statistical distributions of tone detection behavioral response rates across 12 ToneBoxes. **a**, Behavioral response rate histograms across time in a trial for hits during the dark cycle (dark blue) and light cycle (light blue), and for early responses also during the dark cycle (dark green) and light cycle (light green). Shading shows ± 2 SEMs. **b**, Response rate–based task performance accuracy box plots. The star indicates that the dark cycle hit rate was significantly above the light cycle hit rate ($p = 0.019$, KS test; $n = 12$ cages). Data color coded as in **a**. The + marks indicate data points outside of the 25th and 75th percentiles. **c**, Hit and early rates shown for each hour of the day. The bin for each hour begins at tick marks. The dark cycle is shown in the shaded region. The thick black line shows when the hit rate was significantly above the early rate ($p < 0.01$, KS test; $n = 12$ cages). Shading shows ± 2 SEMs.

contingencies were the same as for target trials. The ToneBox also allows users to randomize tone levels across trials and include silent probe trials.

The user selects one of the following four training phases: habituation, shaping, detection, and discrimination. Here we focus on the core phases of detection and discrimination. Figures 2–4 show the results of training 12 cages of 24 mice on pure-tone detection tasks. The training session was run continuously for 9 d while the data were analyzed on-line with the provided ToneGraph software (Materials and Methods). Figure 2a shows an example of trial-by-trial response rates (i.e., hit and early rates) for a single cage. Since each trial of the task was time stamped, we were able to identify the circadian rhythm of the mouse through behavior, which is evidenced by the concentration of behavioral activity during the dark cycle of the room (Fig. 2a, gray). White overhead lights were kept on from 5:00 P.M. to 5:00 A.M. during the light cycle. The white lights were turned off, and red overhead lights were turned on from 5:00 A.M. to 5:00 P.M. during the dark cycle. The time stamping of trials also allowed us to quantify task engagement at different timescales across 7 full days of training (Fig. 2b). The average number of trials per circadian cycle with at least one lick per trial was greater during the dark versus light cycle (54.1 ± 10.5 vs 13.0 ± 3.3 , $p < 0.001$). The average number of trials per minute per day was 0.07 ± 0.01 . Our results for the light versus dark cycles indicate that the ToneBox system can be used to study how circadian rhythms affect performance across the day, without artificially disrupting the wakefulness of animals, as is done when removing animals from their home cage for manual behavioral training.

For each of the 12 boxes, we calculated hit and early response rates across the training period of 9 d. Figure 2c shows that hit rates tended to be greater than early rates for all boxes except box 12. These data indicate that the ToneBox is highly effective at automatically training large populations of mice to detect sounds.

Our initial experiments on tone detection used two mice per cage. However, some researchers may prefer to train individually housed mice or perhaps multiple cage-mates together. The latter scenario, together with the ease of

ToneBox scalability, would allow for the simultaneous automated training of hundreds of mice per study. We find that individual mice (two males, one female) initially trained in pairs perform well when separated and continued training individually (Fig. 2d, three left columns). In addition, larger groups of four mice initially trained together in one home cage (four females) also collectively perform well on the task (Fig. 2d, right column). Thus, one need not train mice individually if one requires only a population of trained mice with a statistical distribution of expertise across cage mates. Individual mice from groups could be tested individually when needed. Given that other commercial behavioral systems are orders of magnitude more expensive than the ToneBox, our system remains scalable for individual or groups of mice. Our data show that mice learn to perform the tone detection task equally well when trained as individuals, pairs, and in larger groups.

Behavioral phenotyping may require multiple behavioral tasks to test the full range of sensation in a given modality. The ToneBox system allows different kinds of behaviors to be tested simultaneously in separate cages. For example, while most boxes continued to do the ongoing detection task with one pure tone at a fixed frequency and level, we stopped the training in a single box to allow for testing of auditory detection sensitivity by randomizing by tone level and frequency (Fig. 2e). Using this procedure, we are able to determine the behavioral audiograms of the mice. Hot colors in Figure 2e indicate better task performance, which was higher for more intense tones and best near 12 kHz, which is consistent with previous findings in healthy mice obtained using classic manual methods (Radziwon et al., 2009).

Our data suggest differences in task engagement across the light and dark cycles (Fig. 2a). To quantify how task engagement varied during light and dark cycles, we averaged tone detection response rates separately within the dark and light cycles (Fig. 3). In general, we found that hit rates were higher during dark versus light cycles ($6.5 \pm 1.0\%$ vs $3.9 \pm 1.8\%$, $p = 0.017$), and early rates were less than hit rates (dark: $2.08 \pm 1.60\%$, $p < 0.001$; light: $1.06 \pm 0.51\%$, $p = 0.002$; Fig. 3b). Early rates were not differ-

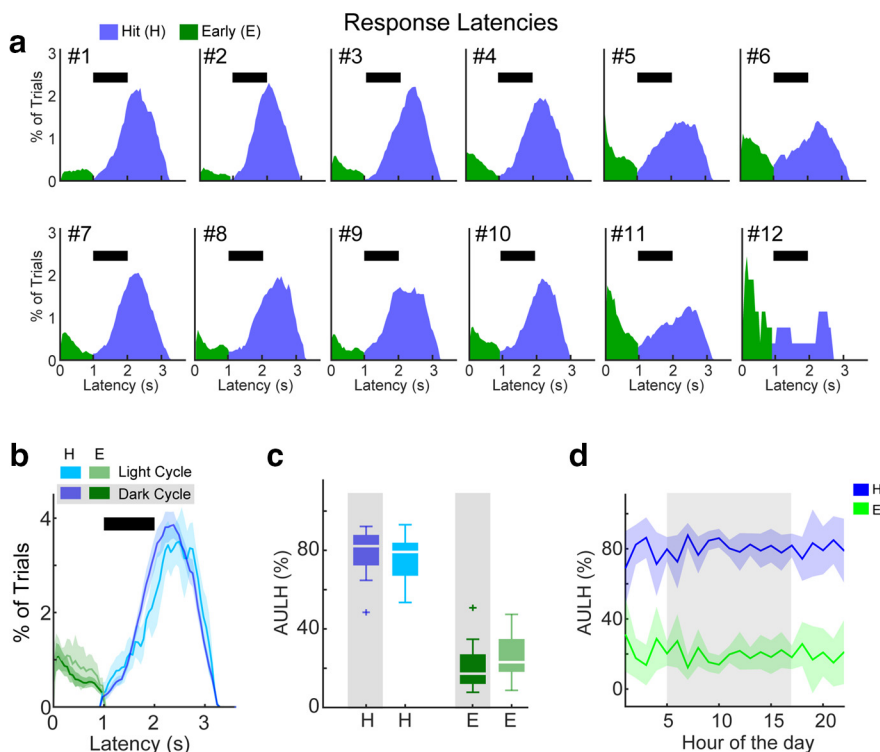


Figure 4. Pure-tone detection behavioral response latencies (i.e., the delay from trial onset until the first lick of the waterspout in each trial). **a**, Latency histograms for each cage. The black bar shows when the 1 s tone was presented. **b**, Average response latency distributions across 12 ToneBoxes. **c**, Latency-based task performance accuracy box plots. Performance accuracy was defined here by the AULH (**a**) for hit and early responses. Shading shows ± 2 SEMs. Data color coded as in **b**. **d**, Hit and early latency-based task performance accuracy shown for each hour of the day. The dark cycle is shown in the shaded region. Shading shows ± 2 SEMs. Data color coded as in **b**.

ent during dark versus light cycles ($p = 0.24$). Figure 3c shows the average hit and early rates across each hour of the day. Response rates increased just before the onset of the dark cycle (Fig. 3c, gray region), peaked near the middle of the dark cycle, then decreased and stayed low for the remainder of the day. Hit rates were significantly above early rates both during the hours leading up to the dark cycle and during most of the dark cycle ($p < 0.01$). Hit rates peak between 9:00 A.M. and 1:00 P.M. Toward the end of the dark cycle, and during most of the following light cycle, hit rates and early rates were not significantly different. These analyses show that task engagement is not uniform across the light/dark cycle. Rather, there is an optimal time window for task engagement that peaks midway through the dark cycle.

Each behavioral response (i.e., a lick) is recorded with a 30 ms resolution (~ 33 Hz sampling rate), which allows fine time analysis of behavioral responses to quantify behavioral accuracy. For example, Figure 4a shows histograms of behavioral response latencies (i.e., the delay from trial onset until the first lick of the waterspout for each cage in each trial). For 9 of 12 cages, the response latency functions peaked near the end of the pure tone (mean peak latency for all cages, 1.75 ± 0.59 s), indicating that most mice tended to lick only after the onset of the pure tone. We note that the response latencies may include a delay to approach the waterspout that depends

on where the mouse was in the cage when it heard the tone.

Figure 4b shows the average response latency histograms for light versus dark cycles across cages. The functions appeared quite similar for light and dark cycles, suggesting that a response latency-based metric of task performance accuracy may be less influenced by overall behavioral activity than behavioral response rates. We thus quantified auditory detection accuracy by taking the percentage of the area under the latency histogram (AULH) for hit and early responses (Fig. 4a, blue and green curves, c, d). We found that the AULH was greater for hit versus early responses for both dark and light cycles (dark: $79 \pm 7.4\%$ vs $20.4 \pm 7.3\%$, $p < 0.001$; light: $75.0 \pm 6.7\%$ vs $24.7 \pm 6.6\%$, $p < 0.001$), indicating good task performance accuracy. Moreover, we found that task performance accuracy for dark versus light cycles was similar ($\text{hit}_{\text{dark}} - \text{hit}_{\text{light}} = 4.0 \pm 5.5\%$, $p = 0.43$; $\text{early}_{\text{dark}} - \text{early}_{\text{light}} = -4.3 \pm 5.4\%$, $p > 0.05$). Task performance accuracy might vary within the dark or light cycle. We thus reanalyzed the latency histograms in hourly time bins. Figure 4d shows the AULH analysis across each hour of the day. This analysis revealed that detection accuracy was consistently high across both light and dark cycles. This similarity in detection accuracy is in marked contrast to the light/dark cycle dependence of response rates (Fig. 3c). Together, our results indicate that most mice in cages

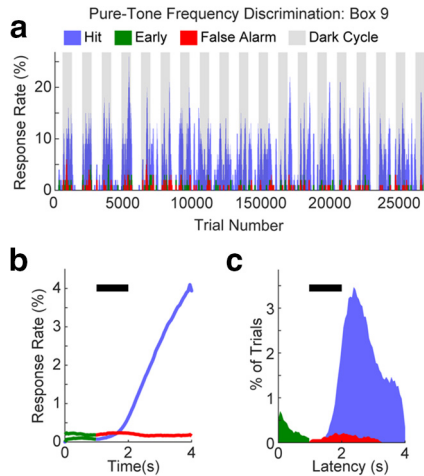


Figure 5. Pure-tone frequency discrimination task performance. **a**, Example of behavioral data collected over 19 consecutive days in a single ToneBox of two mice. **b**, Hit, early, and false alarm response rate histograms across time in a trial, color coded as in **a**. **c**, Hit, early, and false alarm latency histograms across time in a trial, color coded as in **a**.

were well trained (i.e., the mice usually refrained from licking unless the pure tone was detected). Significant increases and decreases in hit rates anticipated the change to dark and light cycles in the room, respectively. The overall lower hit rate during the light cycle compared with the dark cycle most likely reflects water satiation and that the mice were likely sleeping, as is expected from nocturnal animals (de Hoz and Nelken, 2014). Despite the circadian rhythmic modulation of overall behavioral activity (Fig. 3c), the latency of behavioral responses during pure-tone detection indicate that task performance accuracy was similar for dark and light cycles (Fig. 4d). Thus, the performance in an auditory detection task is independent of the circadian rhythm.

We reasoned that task difficulty might affect task engagement. Thus, having established a behavioral baseline for auditory detection, we sought to study how increasing task difficulty affected behavioral activity. We thus trained mice in two ToneBoxes on a pure-tone frequency discrimination task (Figs. 5–7). The discrimination task is more difficult than tone detection because the mouse must classify tones as a target versus nontarget before making a behavioral choice—which requires a behavioral inhibition that is absent in the detection task. Figure 5a shows an example of trial-by-trial behavioral activity for hit, early, and false alarm trials across 19 continuous days of training for an individual box of two mice. As in tone detection, most behavioral activity was concentrated during the dark cycle for tone discrimination, and most responses were hits. Figure 5, b and c, shows the response rates and latency histograms for this cage. Both the response rate and latency functions peak after the tone onset, indicating that the mice were able to discriminate target versus nontarget tones. Figure 6a shows the average response rates for the two cages—confirming that the mice were more active during the dark cycle.

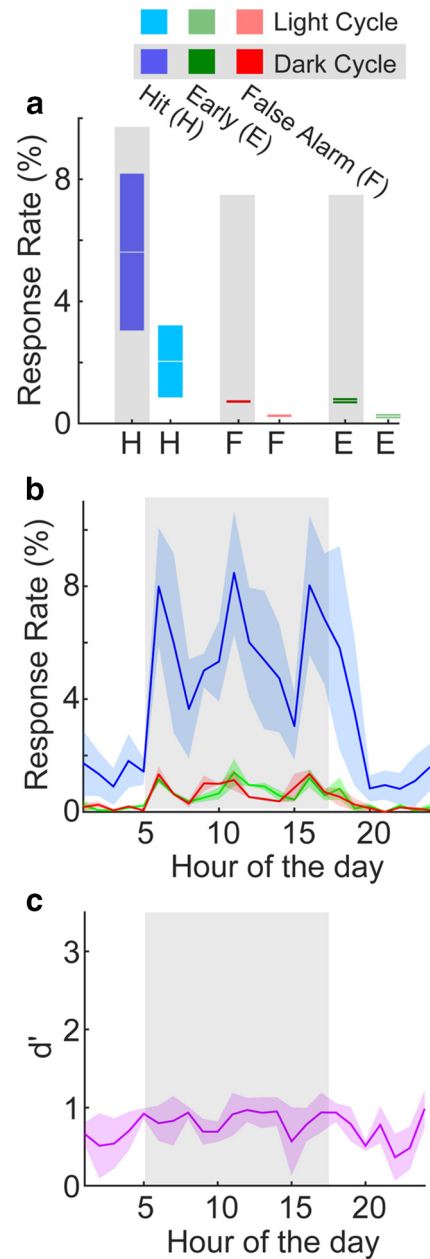


Figure 6. Statistical distributions of tone discrimination response rates across two ToneBoxes. **a**, Response rate-based task performance accuracy box plots for hits during the dark cycle (dark blue) and light cycle (light blue), false alarms during the dark cycle (dark red) and light cycle (light red), and for early responses also during the dark cycle (dark green) and light cycle (light green). **b**, Hit, false alarm, and early rates shown for each hour of the day. The dark cycle is shown in the shaded region. Data color coded as in **a**. Shading shows 2 SEMs. **c**, Response rate-based d' values for each hour of the day. Shading shows ± 2 SEMs.

A comparison of the results from the detection task (Fig. 3c) with those of the discrimination task (Fig. 6b) revealed task-dependent differences in the hourly response rates. Only for the discrimination task was there a periodic modulation in both tested cages, with response rates repeatedly rising and falling approximately every 3–4 h

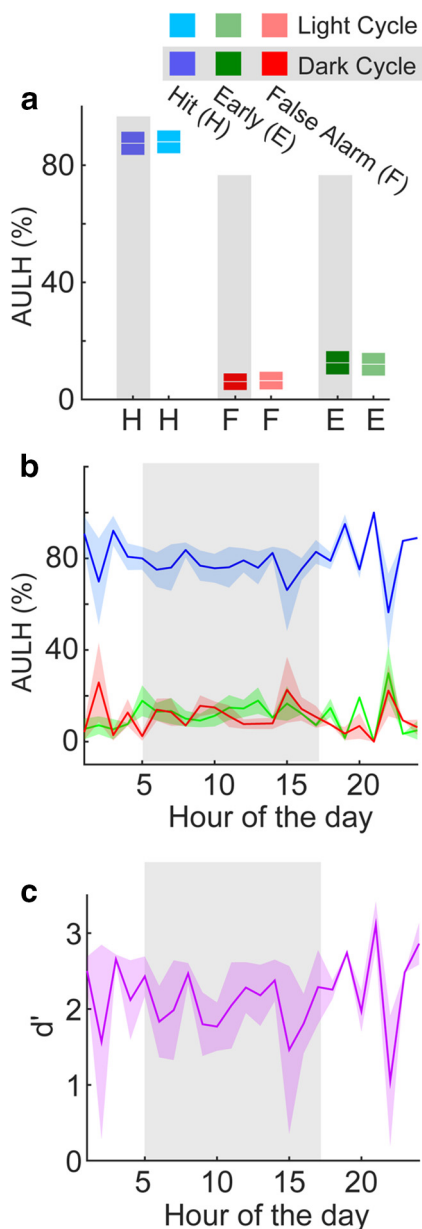


Figure 7. Statistical distributions of tone discrimination response latencies across two ToneBoxes. **a**, Latency-based task performance accuracy box plots. Performance accuracy was defined here by the AULH for hit and early responses. **b**, Response latency-based task performance accuracy shown for each hour of the day, and for hit, false alarm, and early responses. The dark cycle is shown in the shaded region. Shading shows ± 2 SEMs. Data color coded as in **a**. **c**, Response latency-based d' values for each hour of the day. Shading shows ± 2 SEMs.

during the dark cycle. These data suggest that when task difficulty increases, the mice tend to take periodic breaks in task performance. However, it is possible that the periodic modulation may be an effect of continued training on a go/no-go task, rather than due to task difficulty per se.

We next investigated whether there were variations in discrimination task accuracy across the dark cycle. The discrimination task provides both hit and false-alarm tri-

als, and we thus computed the discrimination sensitivity (d') values to quantify discrimination sensitivity (Fig. 6c). In contrast to the response rates per se (Fig. 6b), we found that the d' value was generally consistent across the circadian cycle. This hourly d' analysis of the discrimination task supports our finding from our analysis of response latency: once engaged with the task, the mice show consistent performance accuracy throughout the circadian cycle. Moreover, AULH analysis of hit, early, and false alarm responses during the discrimination task (Fig. 7) also show that task performance accuracy was consistently high for both light and dark cycles. Thus, while task difficulty can cause variations in task engagement during the dark cycle, task accuracy remains constant.

Here we demonstrate the ToneBox for use in tone detection and tone discrimination tasks. In addition to testing basic sensation, frequency discrimination as implemented here can be used to test cognitive flexibility and impulsivity in a reversal or stop-signal task (Logan et al., 2014), respectively (Materials and Methods).

Discussion

These results show that the ToneBox system can simultaneously train many mice on different tasks and can thus be used to efficiently phenotype animal behavior. Benefits of the ToneBox system include that (1) it fits within existing commercial home cages—which reduces cost and eliminates the need for additional laboratory space to train mice; (2) it is open-source, and thus fully customizable; (3) it can produce ultrasonic sounds and provides automated acoustic calibration; and (4) it is cost effective—a single ToneBox costs under \$250—making it a truly scalable system. While the current design of the ToneBox does not track the behavior of individual mice, the Raspberry Pi-based design allows knowledgeable users to add, for example, cameras or RFIDs (radio-frequency identification), to track individual mice. As a low-cost and high-throughput tool designed for both entry-level and expert behaviorists, we expect the ToneBox to be a transformative tool in behavioral phenotyping and drug discovery, and for understanding the neural basis of perception, cognition, and action. When combined with neuroimaging in freely behaving mice, the ToneBox has the potential to facilitate rapid new discoveries in neuroscience research.

References

- Balcombe JP, Barnard ND, Sandusky C (2004) Laboratory routines cause animal stress. *Contemp Top Lab Anim Sci* 43:42–51.
- Balzani E, Falappa M, Balci F, Tucci V (2018) An approach to monitoring home-cage behavior in mice that facilitates data sharing. *Nat Protoc* 13:1331–1347.
- Corcoran KA, Lu Y, Turner RS, Maren S (2002) Overexpression of hAPPswe impairs rewarded alternation and contextual fear conditioning in a transgenic mouse model of Alzheimer's disease. *Learn Mem* 9:243–252.
- de Hoz L, Nelken I (2014) Frequency tuning in the behaving mouse: different bandwidths for discrimination and generalization. *PLoS One* 9:e91676.
- Ewing SG, Grace AA (2013) Evidence for impaired sound intensity processing during prepulse inhibition of the startle response in a

- rodent developmental disruption model of schizophrenia. *J Psychiatr Res* 47:1630–1635.
- Francis NA, Kanold PO (2017) Automated operant conditioning in the mouse home cage. *Front Neural Circuits* 11:10.
- Francis NA, Winkowski DE, Sheikhattar A, Armengol K, Babadi B, Kanold PO (2018) Small networks encode decision-making in primary auditory cortex. *Neuron* 97:885–897.e6.
- Gess A, Schneider DM, Vyas A, Woolley SMN (2011) Automated auditory recognition training and testing. *Anim Behav* 82:285–293.
- Kuchibhotla KV, Gill JV, Lindsay GW, Papadoyannis ES, Field RE, Sten TA, Miller KD, Froemke RC (2017) Parallel processing by cortical inhibition enables context-dependent behavior. *Nat Neurosci* 20:62–71.
- Kurt S, Ehret G (2010) Auditory discrimination learning and knowledge transfer in mice depends on task difficulty. *Proc Natl Acad Sci U S A* 107:8481–8485.
- Logan GD, Van Zandt T, Verbruggen F, Wagenmakers EJ (2014) On the ability to inhibit thought and action: general and special theories of an act of control. *Psychol Rev* 121:66–95.
- Poddar R, Kawai R, Ölveczky BP (2013) A fully automated high-throughput training system for rodents. *PLoS One* 8:e83171.
- Radziwon KE, June KM, Stolzberg DJ, Xu-Friedman MA, Salvi RJ, Dent ML (2009) Behaviorally measured audiograms and gap detection thresholds in CBA/CaJ mice. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 195:961–969.
- Rose T, Jaepel J, Hübener M, Bonhoeffer T (2016) Cell-specific restoration of stimulus preference after monocular deprivation in the visual cortex. *Science* 352:1319–1322.
- Schaefer AT, Claridge-Chang A (2012) The surveillance state of behavioral automation. *Curr Opin Neurobiol* 22:170–176.
- Sofroniew NJ, Flickinger D, King J, Svoboda K (2016) A large field of view two-photon mesoscope with subcellular resolution for in vivo imaging. *Elife* 5:e14472.
- Zhou X, Lu JY, Darling RD, Simpson KL, Zhu X, Wang F, Yu L, Sun X, Merzenich MM, Lin RC (2015) Behavioral training reverses global cortical network dysfunction induced by perinatal antidepressant exposure. *Proc Natl Acad Sci U S A* 112:2233–2238.