Psychophysical measurement of contrast sensitivity in the behaving mouse

Mark H. Histed, Lauren A. Carvalho, and John H. R. Maunsell

Department of Neurobiology and Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts

Submitted 1 July 2011; accepted in final form 30 October 2011

Histed MH, Carvalho LA, Maunsell JH. Psychophysical measurement of contrast sensitivity in the behaving mouse. J Neurophysiol 107: 758–765, 2012. First published November 2, 2011; doi:10.1152/jn.00609.2011.—To understand how activity in mammalian neural circuits controls behavior, the mouse is a promising model system due to the convergence of genetic, optical, and physiological methods. The ability to control and quantify behavior precisely is also essential for these studies. We developed an operant visual detection paradigm to make visual psychophysical measurements: head-fixed mice make responses by pressing a lever. We designed this task to permit neurophysiological studies of behavior in cerebral cortex, where activity is variable from trial to trial and neurons encode many types of information simultaneously. To study neural responses in the face of this complexity, we trained mice to do a task where they perform hundreds of trials daily and perceptual thresholds can be measured. We used this task to measure both visual acuity and the minimum detectable contrast in behaving mice. We found that the mouse contrast response function is similar in shape to other species. They can detect low-contrast stimuli, with a peak contrast threshold of 2%, equivalent to ~15° eccentric in human vision. Mouse acuity is modest, with an upper limit near 0.5 cycles/°, consistent with prior data.

Acuity and contrast thresholds have been previously measured in mice physiologically and behaviorally. We focus here on photopic measurements, as prior studies have worked mainly in the photopic range, although scotopic conditions are also likely to be important to a nocturnal animal like the mouse. Most methods agree that maximum photopic acuity is approximately 0.5–0.6 cycles per degree (cpd), whereas measurements of contrast sensitivity vary.

Two established behavioral methods rely on visual reflexes: optomotor studies (Abdeljail et al. 2005; Prusky et al. 2004) measure motion of the head or body in response to full-field motion, and optokinetic studies (Sinex et al. 1979; van Alphen et al. 2009) measure motion of the eyes. Operant studies have used mazes (Gian Franceschi et al. 1999) and water mazes (Prusky et al. 2000). The optokinetic, optomotor, and operant studies have all yielded acuity thresholds at 0.4–0.6 cpd. Data from nearly all of these studies show that animals can detect at best a minimum full-field contrast of 5–15%. The exception is the recent work of van Alphen (2009), which shows that stimuli of ~1% contrast can produce optokinetic eye movements.

Physiologically, visual evoked potentials have been used to measure contrast thresholds and acuity. These are often measured from the scalp or skull above V1 (Fischer et al. 2007; Lickey et al. 2004; Ridder and Nusinowitz 2006) but can also be measured using local field potential recordings from penetrating electrodes (Porciatti et al. 1999). These studies also estimate maximum acuity of 0.5–0.6 cpd and contrast thresholds near 5%.

In this behavioral paradigm, animals report when a visual stimulus changes. We use the paradigm to measure fundamental limits of the mouse visual system by asking them to detect stimuli of varying contrast and spatial frequency.

The density of retinal ganglion cells (RGCs) and the magnification of the retinal image limit maximum acuity. Consequently, higher acuity is seen in animals with larger eyes and higher peak density of RGCs, such as primates (human and macaques: 40–60 cpd; Merigan and Katz 1990; Thibos et al. 1987) and cats (6–10 cpd). Both primates and cats have a region of their retinas where ganglion cells are concentrated: primates have a fovea where single cones map to single RGCs, and cats have an elongated area centralis/visual streak (Hughes 1977) with a density of RGCs 50-fold greater than at the periphery (Stone 1965, 1978). In contrast, ganglion cell density in mice varies only by a factor of 2–4 between the central retina and periphery (Jeon et al. 1998; Salinas-Navarro et al. 2009).

Thus mouse acuity is lower than most other mammalian species and is likely to be limited by the size of the image projected onto the retina and the density of retinal cells. However, do mouse contrast thresholds differ from other

Address for reprint requests and other correspondence: M. H. Histed, Dept. of Neurobiology, Harvard Medical School, 220 Longwood Ave., Boston, MA 02115 (e-mail: mark_histed@hms.harvard.edu).
species? If the mouse central retina is like the peripheral retina of other species, contrast thresholds might be similar. Humans and macaques have a minimum detectable contrast threshold near 0.5% (Barten 1999; van Nes et al. 1967) and cats ~1% (Blake et al. 1974). At 10–15° eccentricity in the macaque, where RGC density matches that of the mouse, contrast thresholds are still ~2% (Virsu and Rovamo 1979).

Even though RGC density and optomotor tasks predict low-contrast thresholds, operant studies in the mouse find contrast thresholds to be 5% or above, including a water maze study that showed a peak contrast threshold of 15% (Prusky and Douglas 2004) in the rat, with its similar retina, operant studies have also found contrast thresholds of ~15% (Keller et al. 2000; Meier et al. 2011). In addition, a recent study of mice (Busse et al. 2011) used an operant task with a nose-poke response to measure contrast thresholds at an optimal spatial frequency (0.13 cpd). They obtained a somewhat worse threshold than previous studies, slightly above 20%.

A key advantage of psychophysical measurements is that deviations in behavior due to attentional or motivational fluctuations can be detected and animals trained to prevent these lapses. Therefore, we developed a behavioral paradigm to make visual psychophysical measurements in the mouse. We used this paradigm to measure mouse maximum acuity and minimum contrast threshold. Our data on maximum acuity agree well with past measurements. However, we find a minimum contrast threshold of ~2%, much better than previously measured behaviorally. This is consistent with the idea that the mouse retina has photopic contrast sensitivity similar to the peripheral retina of other species.

METHODS

Animals and behavior. All animal procedures were approved by the Institutional Animal Care and Use Committee at Harvard Medical School and conform to National Institutes of Health guidelines. To begin training, a custom-made titanium or stainless steel head holder was implanted on the skull in an aseptic surgery. After recovery, animals (n = 3, 30–90 days postnatal age; C57BL/6-BALB/c hybrids, see below) were trained to perform the visual change detection task via operant conditioning with positive reinforcement. During training, mice received fluid only as reinforcement or via supplement when total reinforcement volume was below a minimum of 25 ml·kg⁻¹·day⁻¹. Animals received water rewards (1–2 µl per reward once performance was stable; larger rewards were used during early training). Animals received ≈25 ml/kg and often more on nontraining days and were typically stable at 80–85% of their free-feeding weight during training periods. We did not track eye position as we wished to estimate peak acuity and used large stimuli that filled the monitor (69 × 81° of visual angle). Animals had to respond within a reaction time interval of 600–700 ms; nearly all responses occurred within 500 ms (e.g., Figs. 2 and 3), so the exact length of the reaction time window was not important for our measurements.

The 3 animals used in this study were the 1st 3 head-fixed mice we trained to perform this task. The training began by teaching animals to hold the lever until a high-contrast visual stimulus appeared. There was no explicit habituation to handling; we implanted a titanium or steel plate for head fixation, and after ~2 wk of recovery, head-fixed shaping began. This took a mean of 21 sessions (range 18–25). Next, the visual stimulus onset time was increased to the maximum (2 s) over a mean of 9 sessions (range 7–11). For unrelated reasons, these animals then stopped training for 8 wk. On restart, it took a mean of 32 sessions (25–45) for animals to perform the task stably near threshold with a range of stimulus contrasts. Total training days: mean 61, range 54–74. This time is long for typical mouse behavioral experiments but similar to the amount of time required to train a nonhuman primate on a near-threshold lever press detection task.

We have since trained more animals on similar tasks to perform neurophysiological recordings. Based on all of the mice we have trained on a visual detection task (n = 21), approximately half (n = 12) learned the task fully and performed at a high level. Of the rest, n = 3 failed to behave consistently, and we stopped training them within 10 sessions, n = 2 were withdrawn for unrelated reasons, and n = 4 learned the task but performed at a slightly lower level so we stopped training them to focus on other animals. The example animal shown in Supplemental Video S1 (available in the data supplement online at the Journal of Neurophysiology web site) is from a later set of animals trained to do the 100% contrast detection task (as in Fig. 2) only because the video was made several months after the threshold data here were collected. The video captures the essential features of the contrast threshold detection task.

The animals are 1st-generation offspring of female BALB/c and male C57BL/6 mice (Charles River Laboratories). We used this cross to facilitate breeding and secondarily to work behaviorally with vigorous hybrid animals. However, our data combined with past work (e.g., Prusky et al. 2004; van Alphen et al. 2009) show that these offspring have visual acuity similar to their C57BL/6 parents. Although some mouse strains carry alleles resulting in retinal degeneration in adulthood, neither parental strain does (Chang et al. 2002), and they have normal adult retinas (Hawes et al. 1999). The offspring animals are agouti-colored with pigmented retinas. Visual behavioral performance was stable for several months in the adult hybrid animals.

Visual stimulation. Visual stimuli were displayed on a γ-corrected liquid crystal display (LCD) display (Dell P190S). Background brightness was 60 cd/m². We measured the time to transition between background and the 100% contrast visual stimulus as <10 ms. At this brightness, pixel intensity oscillates at 200 Hz as the backlight is pulse-width modulated at that frequency. This is far above mouse flicker-fusion frequency as measured with ERG (Tanimoto et al. 2009). Regardless, for physiological experiments, this oscillation can be eliminated by setting the display to maximum brightness. Tests with a cathode ray tube (CRT) video display were made using a Viewsonic P95f+ running at 120 Hz. Monitors were linearized by measuring intensity for each gray level (Eye-One Display 2; X-Rite). The monitors were placed 18–30 cm from the eye; most measurements were made at 26-cm distance.

Threshold calculation. To calculate threshold contrast, we first removed false alarm trials. Responses that occurred within 100 ms after stimulus onset were not counted as correct as they were too early to have been responses to the stimulus. Spurious correct responses due to false alarm guesses occurring during the response period were subtracted. We estimated the false alarm rate over a time window 100 ms before to 100 ms after the stimulus and removed a proportional number of correct trials based on the length of the reaction time interval. The range of trials removed over the six example sessions in Fig. 3 was 7.5–18.9%. Response rate data (Figs. 2, 4A, and 6) are data before this correction was applied; y-axes of psychometric curves (Fig. 3, A and B, and Fig. 4C) are data after correction. We then fit a Weibull function (Quick 1974). Threshold was estimated as the midpoint between the upper and lower saturation limits of the Weibull.

RESULTS

Behavioral task: visual change detection. We trained mice to make operant responses by manipulating a lever while their heads are fixed. Animals sit in a tube that supports their weight. A surgically implanted titanium headpost is held by a clamp above the tube. A lightweight lever attached to a microswitch
Animals' performance is reliable from session to session. We set out to develop a behavioral paradigm in which psychophysical measurements could be made, and animals perform hundreds of trials in a given experimental session. This visual detection task meets these requirements. Mice perform reliably across sessions (Fig. 4). Correct and false alarm rates are stable, as shown by example data from a representative animal in Fig. 4A. This animal performed hundreds of correct trials each day over a 3-wk period (mean 583, SD 166, 5 training days/wk). Psychophysical estimates of contrast threshold are also stable (Fig. 4C).

Contrast thresholds are reliable across animals. We used the visual detection task to measure the contrast response function (Campbell and Robson 1968). We obtained contrast thresholds across a range of spatial frequencies (Fig. 5; n = 3 animals). The resulting contrast response functions are consistent across animals. The minimum detectable contrast level was ~2% (sensitivity = 50), at a spatial frequency of ~0.1 cpd.

We used an LCD monitor to collect these data, but CRT monitors are better characterized for visual measurements and have artifacts that are better understood. Even though we linearized the luminance steps of the LCD, we wished to completely rule out any artifacts that might result from potential temporal “overshoot” or spatial dithering by internal processing in the LCD. Therefore, we made measurements with a CRT at a single spatial frequency (0.1 cpd), near the contrast threshold peak. At that spatial frequency, animals' contrast thresholds are measured...
sensitivity with an LCD was 54 (mean; SD 4.6; n = 3 animals; data same as in Fig. 5). Their contrast sensitivity with the CRT was very similar: 44, SD 6.4 (n = 7 measurements from 2 animals), confirming that the LCD monitor accurately replicated the desired contrast steps.

Animals showed low lapse rates (Fig. 6). A nonzero lapse rate would appear in Fig. 3, A and C, as an upper asymptote below 100%, but in this task animals typically made correct responses for the highest contrast stimuli. Across all trials, the median proportion of correct responses was ~60%. Animals showed moderate false alarm rates (i.e., lever releases before the stimulus appeared). These did not affect our threshold measurements, as false alarms are ignored when computing the psychometric function. Furthermore, we corrected for spurious correct rates due to guesses that occurred while the stimulus was displayed (see METHODS; Macmillan and Creelman 2005). False alarms increased during training when low-contrast stimuli were added to the 100% contrast stimuli that were used during initial training (data not shown). This is a rational strategy; decision criteria should decrease when less-detectable stimuli are used, resulting in more false alarms. Although it should be possible to train animals to reduce early responses (by manipulating task timing or reward schedule as a function of stimulus onset time), because it does not affect threshold measurements we decided not to train animals to suppress false alarms.

Fig. 3. Mice are reliable psychophysical observers. A: psychometric function for 1 animal for 1 behavioral session. Grating spatial frequency: 0.09 cycles per degree (cpd). Measured threshold is 1.8% contrast. Error bars: 95% confidence intervals (for a binomial using the Clopper-Pearson method). Percentage correct is calculated over correct response and missed trials only. False alarms are ignored (see METHODS); as they come before the stimulus onset, they are not responses to the visual stimulus and must be the same over all contrast conditions. In addition, the spurious correct rate due to false alarm guesses is calculated and subtracted (METHODS). B: median reaction times for the same session (error bars: SE over trials). Reaction times were not computed for the 2 lowest contrasts as too few correct responses were made. Reaction times are slower for lower contrasts, as seen in human vision (Manahilov et al. 2003). C: data for 6 different sessions of similar spatial frequency (green and blue lines: 0.12 cpd; yellow line: 0.09 cpd; remainder: 0.1 cpd). Thresholds range from 1.7 to 2.3%. One curve has a slightly shallower slope, and this has little effect on the threshold; the contrast levels we chose were selected to provide good estimates of threshold and not slope. Yellow: same data as in A. D: reaction times for sessions shown in C. Blue line, mean across sessions; gray bars, SE over sessions.
type for a particular experiment depends on many things, including the numbers of trials produced and animals’ speed of learning. An important difference may be brain stability, which is needed for in vivo microscopy. Licking has been shown to produce larger displacements of the brain, especially in the dorsal/ventral direction, than paw motions (Andermann et al. 2010), and thus lever tasks may be more suitable for, e.g., two-photon microscopy, although direct comparisons have not yet been made.

A final goal was to control the animals’ behavior to obtain measurements limited by perceptual factors. When animals are performing near threshold, fluctuations in motivation, attention, or arousal can be detected in increased error rates or other variations in psychometric functions. We have trained animals to perform consistently to reduce these fluctuations (Fig. 4). Also, a benefit of making psychometric measurements is that changes in internal state can be detected. Animals perform this task while they rest stably in the behavioral apparatus, facilitating optical and electrophysiological recording and manipulations.

Choice of paradigm: yes-no and forced-choice tasks. We used a “yes-no” task to measure thresholds. When the stimulus appeared, animals either released the lever to report its presence or made a “no” response by holding the lever. Another possibility is a two-alternative forced-choice (2AFC) task, where two stimuli are presented at different locations or times and animals must choose one of the two options. The most important reason we chose a detection task is because we thought it easier to train animals to perform with a single lever. An additional minor factor is that a 2AFC task typically requires more trials to measure thresholds because the chance performance level of 50% compresses the response range. On the other hand, thresholds measured with 2AFC tasks are often a factor of ~2 better than yes-no tasks. In a 2AFC task, animals compare two stimuli separated by only a short time, whereas in a
yes-no task, animals must choose a criterion level based on the signal and noise distributions when the stimulus is present and absent. The difference in thresholds may be due to criterion deviating slightly from optimal in yes-no tasks due to the longer time intervals over which subjects sample the stimuli (Macmillan and Creelman 2005). If we had used a 2AFC task, we might thus measure better contrast thresholds. Matching tasks, such as what Virsu and Rovamo (1979) used in human subjects, also require comparisons over short time intervals and may be more similar to a 2AFC task.

Our measurements could have been degraded if animals used a nonoptimal decision strategy. Deviations from optimal behavior fall into two major categories: fluctuations in internal decision criterion or fluctuations in attentional or motivational state (failing to pay attention or choosing to try hard on some trials and not others). Both deviations would produce changes in threshold estimates from day to day, but our threshold estimates were repeatable both within the same animal and across animals (Figs. 3–5). Substantial day-to-day threshold fluctuations were seen by Busse et al. (2011) in a contrast detection task and might have been caused by variance in motivation, which would explain why their threshold estimate (~20% at 0.13 cpd) was much higher than ours (~2% at 0.13 cpd). Although the 2AFC design used by Busse et al. (2011) can provide better thresholds, it does not prevent attentional or motivational changes from degrading measurements. The low lapse rates and relatively stable thresholds that we measured suggest that our animals were well-motivated and that the thresholds we measured accurately reflect behavioral abilities for these mice.

Thresholds for detecting the onset of a stimulus and for discriminating the features of a stimulus can differ. Thibos and colleagues (1996) have studied this extensively in humans. When human observers are asked to report the onset of a visual grating, in the periphery they can detect the presence of high-spatial-frequency gratings for which orientation is expected to make thresholds for large stimuli lower than thresholds with smaller stimuli (Virsu and Rovamo 1979; Watson 1992), and smaller stimuli could explain why some measurements in rats found contrast thresholds ~15% (Keller et al. 2000; Meier et al. 2011). Correspondingly, van Alphen et al. (2009) might have found lower contrast thresholds than ours because their stimuli filled the entire visual field. Second, although our minimum contrast thresholds are within a factor of two of those measured with optokinetic (eye movement) responses (van Alphen et al. 2010), it is difficult to construct contrast sensitivity functions from their data because they did not search for minimal detectable contrasts or spatial frequencies. Perceptual learning was unlikely to affect our thresholds, as the spatial frequency of the stimulus was most often changed from one session to the next (cf. Merigan and Katz 1990). It is also important to consider the method used to calculate threshold when comparing between studies. We used the 50% point of the psychometric function, i.e., the midpoint between the upper and lower saturation points. Animals perform above chance below this threshold: for example, in Fig. 3, the 50% threshold is 2.6%, but animals perform above chance for a contrast of 1.5%. Most previous work in rodents,
including the optokinetic study of van Alphen et al. (2010), has used the minimal threshold, making our measurements conservative compared with the optokinetic measurements.

The most important factor likely to differentiate our results, showing a peak contrast threshold of 2%, from previous operant studies that obtained thresholds of 15% or higher (Busse et al. 2011; Prusky and Douglas 2004; Prusky et al. 2004) is our control over animals’ motivational state. First, because we measured a psychometric function in each session, we could use the lapse rate to detect decreased motivation, and the majority of our lapse rates were near zero (Fig. 6). Also, our measurements are highly reliable, both from day to day for the same animal (Figs. 3 and 4) and across animals at similar spatial frequencies (Fig. 5). If our animals had deviated from optimal perceptual behavior, we would expect to see fluctuating performance or a higher lapse rate. Thus, in this task, mice are likely to be performing very near their perceptual limit. Because behavior is limited by perception, and not internal state changes like attention or motivation, the task is suitable for studies of neural activity in the mouse visual system.

Conclusion. These data show that mice have good photopic contrast sensitivity, similar to that at ~15° eccentricity in the human. We have also demonstrated that psychophysical methods can be applied in the mouse to measure quantities that characterize their perceptual systems. This psychophysical paradigm is a useful tool for linking structure and function of neural circuits in the mouse to behavior.

ACKNOWLEDGMENTS

We thank R. H. Masland, M. A. Andermann, D. A. Ruff, and J. E. Dowling for useful comments on the manuscript. A. Thavikulwat, S. Slevoda, and T. Wang helped with training and contributed to development of training protocols.

GRANTS

This work was supported by the Howard Hughes Medical Institute.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS

M.H. and J.H.R.M. conception and design of research; M.H. and L.A.C. performed experiments; M.H. analyzed data; M.H. and J.H.R.M. interpreted results of experiments; M.H.H. prepared figures; M.H.H. drafted manuscript; M.H.H. and J.H.R.M. edited and revised manuscript; M.H., L.A.C., and J.H.R.M. approved final version of manuscript.

REFERENCES


