***Title*:** Sound improves neuronal encoding of visual stimuli in mouse primary visual cortex

***Authors*:** Aaron M. Williams[1,2,3], Christopher F. Angeloni[1,4], Maria N. Geffen[1,2,3]

***Affiliations*:**
[1]Department of Otorhinolaryngology, University of Pennsylvania, Philadelphia, United States
[2]Department of Neuroscience, University of Pennsylvania, Philadelphia, United States
[3]Department of Neurology, University of Pennsylvania, Philadelphia, United States
[4]Department of Psychology, University of Pennsylvania, Philadelphia, United States

***Corresponding author*:** Maria Neimark Geffen
Stemmler Hall G10
3450 Hamilton Walk
Philadelphia, PA 19104
mgeffen@pennmedicine.upenn.edu

Number of pages: 29

Number of figures: 9

Abstract word count: 267
Introduction word count: 850
Discussion word count: 2053

***Keywords*:** audiovisual integration, primary visual cortex, stimulus decoding, electrophysiology

Preprint: published on BioRXiv https://doi.org/10.1101/2021.08.03.454738

Acknowledgements
The authors thank Gabrielle Samulewicz for assistance with experiments and members of the Geffen laboratory for helpful discussions, as well as Dr. Jay Gottfried and Dr. Yale Cohen at the University of Pennsylvania. This work was supported by funding by the National Institute on Deafness and Other Communication Disorders at the National Institute of Health grants 5T32DC016903 to AMW, F31DC016524 to CFA, and R01DC015527, R01DC014479, and R01NS113241 to MNG.

Abstract

In everyday life, we integrate visual and auditory information in routine tasks such as navigation and communication. Whereas concurrent sound can improve visual perception, the neuronal correlates of this audiovisual integration are not fully understood. Specifically, it remains unknown whether sound-induced improvement in detection and discriminability of visual stimuli is reflected in neuronal firing patterns in the primary visual cortex (V1). Furthermore, presentation of sound can induce movement in the subject, but little is understood about whether and how sound-induced movement affects audiovisual integration in V1. We investigated how sound and movement interact to modulate V1 visual responses in awake, head-fixed mice and whether this interaction improves neuronal encoding of the visual stimulus. We presented visual drifting gratings with and without simultaneous auditory white noise to awake male and female mice while recording mouse movement and V1 neuronal activity. Sound modulated light-evoked activity of 80% of light-responsive neurons, with 95% of neurons increasing activity when the auditory stimulus was present. Sound consistently induced movement. However, a generalized linear model revealed that sound and movement had distinct and complementary effects of the neuronal visual responses. Furthermore, decoding of the visual stimulus from the neuronal activity was improved with sound, even when controlling for movement. Thus, sound and movement modulate visual responses in complementary ways, improving neuronal representation of the visual stimulus. This study clarifies the role of movement as a potential confound in neuronal audiovisual responses and expands our knowledge of how multimodal processing is mediated in the awake brain.

Significance statement

Sound and movement are both known to modulate visual responses in the primary visual cortex, however sound-induced movement has largely remained unaccounted for as a potential confound in audiovisual studies in awake animals. Here, authors found that sound and movement both modulate visual responses in an important visual brain area, the primary visual cortex, in distinct, yet complementary ways. Furthermore, sound improved encoding of the visual stimulus even when accounting for movement. This study reconciles contrasting theories on the mechanism underlying audiovisual integration and asserts the primary visual cortex as a key brain region participating in tripartite sensory interactions.

Introduction

Our brains use incoming sensory information to generate a continuous perceptual experience across sensory modalities. The neuronal systems underlying sensory perception of different modalities interact in a way that often improves perception of the complementary modality (Gingras et al., 2009; Gleiss and Kayser, 2012; Bigelow and Poremba, 2016; Hammond-Kenny et al., 2017; Meijer et al., 2018; Stein et al., 2020). In the audiovisual realm, it is often easiest to understand what someone is saying in a crowded room by additionally relying on visual cues such as lip movement and facial expression (Maddox et al., 2015; Tye-Murray et al., 2016). The McGurk effect and flash-beep illusion are other common perceptual phenomena that demonstrate mutual interactions between the auditory and visual systems (McGurk and MacDonald, 1976; Shams et al. 2002).

The benefits of additional sensory modalities on unisensory processing do not just apply to complex vocal and auditory behavioral interactions. Concurrent sounds such as auditory white noise and pure tones improve sensitivity to and discriminability of visual contrast gradients in humans (Lippert et al., 2007; Chen et al., 2011; Tivadar et al., 2020). The use of these basic

audiovisual stimuli has demonstrated that the most robust multisensory perceptual improvements occur around threshold discrimination levels of the otherwise unisensory modality (Chen et al., 2011; Gleiss and Kayser, 2012; Breman et al., 2017). The relative timing of the sensory components is also a factor in their integration. Simultaneous onset and offset of the auditory and visual components strengthens multisensory perceptual improvements compared to asynchronous stimuli (Lippert et al., 2007). Multisensory integration is often optimal when modulations in visual intensity and auditory amplitude are temporally congruent (Atilgan et al., 2018), likely mimicking covariance of multisensory signals from natural objects. Despite this current understanding of audiovisual integration at a perceptual level, a detailed understanding of the neuronal code that mediates this improvement has proved elusive.

Previous studies of neuronal correlates of audiovisual integration found that the primary sensory cortical areas participate in this process (Wang et al., 2008; Ibrahim et al., 2016; Meijer et al., 2019; Deneux et al., 2019). The primary visual cortex (V1) contains neurons whose light-evoked firing rates are modulated by sound, as well as neurons that are responsive to sound alone (Knöpfel et al., 2019). Orientation and directional tuning of individual neurons are also affected by sound. In anesthetized mice, layer 2/3 neurons in V1 exhibited sharpened tuning in the presence of sound (Ibrahim et al., 2016), providing a potential mechanism through which sound improves visual encoding. However, another study in awake mice found heterogeneous changes across neurons in visual tuning curve bandwidth with and without sound (Meijer et al., 2017). These contrasting findings raise the question of whether previously reported sound-induced changes in V1 neuronal activity in awake animals resulted in improved visual processing, and through which coding schemes these effects are mediated. Ultimately, this hypothesized improvement in visual encoding would provide a missing link between cross-sensory neuronal responses and the field's current understanding of behavioral and perceptual effects described above.

An important factor that has thus far been unaccounted for in audiovisual studies is that awake animals are subject to brain-wide changes in neuronal activity due to stimulus-aligned, uninstructed movements (Musall et al., 2019). Sound-induced movement represents a potential confound for audiovisual studies in awake animals because whisking and locomotion modulate neuronal activity in the sensory cortical areas. In V1, movement enhances neuronal visual responses and improves neuronal encoding of the visual scene (Niell and Stryker, 2010; Dardalat and Stryker, 2017). Conversely, in the auditory cortex (AC), locomotion suppresses neuronal spontaneous and auditory responses (Nelson et al., 2013; Schneider and Mooney, 2018; Bigelow et al., 2019). Therefore, the contribution of movement to neuronal responses to multi-sensory stimuli is likely due to multiple processes and can greatly affect audiovisual integration.

Thus, audiovisual integration in V1 may not simply represent afferent information from auditory brain regions. Whereas V1 neurons are sensitive to the optogenetic stimulation (Ibrahim et al., 2016) and pharmacologic suppression (Deneux et al., 2019) of AC neurons, the modulation of V1 activity may instead be a byproduct of uninstructed sound-induced movements which themselves modulate visual responses (Bimbard et al., 2021). Here, we tested these alternative explanations of the extent to which locomotion contributes to audiovisual integration in V1 by performing extracellular recordings of neuronal activity in V1 while monitoring movement in awake mice presented with audiovisual stimuli. We used these results to build on prior studies reporting sound-induced changes in V1 visual responses (Ibrahim et al., 2016; Meijer et al., 2017; McClure and Polack, 2019), in order to determine whether and through what coding mechanism this cross-modal interaction improves visual encoding. The audiovisual stimulus consisted of auditory white noise and visual drifting gratings in order to allow comparison of

sound's effect across the visual contrast parameter. We found that the majority of neurons in V1 were responsive to visual and auditory stimuli. Sound and movement exerted distinct yet complementary effects on shaping the visual responses. Importantly, sound improved discriminability of the visual stimuli both in individual neurons and at a population level, an effect that persisted when accounting for movement.

Materials and methods

Mice

All experimental procedures were in accordance with NIH guidelines and approved by the IACUC at the University of Pennsylvania. Mice were acquired from Jackson Laboratories (5 male, 6 female, aged 10-18 weeks at time of recording; B6.Cast-*Cdh23*$^{Ahl+}$ mice [Stock No: 018399]) and were housed at 28°C in a room with a reversed light cycle and food provided ad libitum. Experiments were carried out during the dark period. Mice were housed individually after headplate implantation. Euthanasia was performed using CO_2, consistent with the recommendations of the American Veterinary Medical Association (AVMA) Guidelines on Euthanasia. All procedures were approved by the University of Pennsylvania IACUC and followed the AALAC Guide on Animal Research. We made every attempt to minimize the number of animals used and to reduce pain or discomfort.

Data availability

All data including the spike timing from the recordings will be made available on Dryad upon publication here: https://doi.org/10.5061/dryad.sxksn033q

Surgical procedures

Mice were implanted with skull-attached headplates to allow head stabilization during recording, and skull-penetrating ground pins for electrical grounding during recording. The mice were anesthetized with 2.5% isoflurane. A ~1mm craniotomy was performed over the right frontal cortex, where we inserted a ground pin. A custom-made stainless steel headplate (eMachine Shop) was then placed on the skull at midline, and both the ground pin and headplate were fixed in place using C&B Metabond dental cement (Parkell). Mice were allowed to recover for 3 days post-surgery before any additional procedures took place.

Electrophysiological recordings

All recordings were carried out inside a custom-built acoustic isolation booth. 1-2 weeks following the headplate and ground pin attachment surgery, we habituated the mice to the recording booth for increasing durations (5, 15, 30 minutes) over the course of 3 days. On the day of recording, mice were placed in the recording booth and anesthetized with 2.5% isoflurane. We then performed a small craniotomy above the left primary visual cortex (V1, 2.5mm lateral of midline, 0-0.5 mm posterior of the lambdoid suture). Mice were then allowed adequate time to recover from anesthesia. Activity of neurons were recorded using a 32-channel silicon probe (NeuroNexus A1x32-Poly2-5mm-50s-177). The electrode was lowered into the primary visual cortex via a stereotactic instrument to a depth of 775-1000μm. Following the audiovisual stimulus presentation, electrophysiological data from all 32 channels were filtered between 600 and 6000 Hz, and spikes belonging to single neurons and multi-units were identified in a semi-automated manner using KiloSort2 (Pachitariu et al., 2016).

Audiovisual stimuli

The audiovisual stimuli were generated using MATLAB (MathWorks, USA), and presented to mice on a 12" LCD monitor (Eyoyo) with a 60Hz framerate and through a magnetic speaker

(Tucker-Davis Technologies) placed to the right of the mouse. The visual stimulus was generated using the PsychToolBox package for MATLAB and consisted of square wave drifting gratings 1 s in duration, 4-Hz temporal frequency, and 0.1 cycles/°. The gratings moved in 12 directions, evenly spaced 0°-360°, and were scaled to a range of 5 different visual contrast levels (0, 0.25, 0.5, 0.75, 1), totaling 60 unique visual stimuli. The auditory stimulus was sampled at 400 kHz and consisted of a 1 s burst of 70 dB white noise. The visual grating was accompanied by the auditory noise on half of trials (120 unique trial types, 10 repeats each), with simultaneous onset and offset. A MATLAB-generated TTL pulse aligned the onset of the auditory and visual stimuli, and was verified using a ThorLabs photodetector and microphone. This TTL pulse was also used to align the electrophysiological recording data with the audiovisual stimulus trials. The auditory-only condition corresponded to the trials with a visual contrast of 0. The trial order was randomized and was different for each recording.

Data analysis and statistical procedures

Spiking data from each recorded unit was organized by trial type and aligned to the trial onset. The number of spikes during each trial's first 0-300ms was input into a generalized linear model (GLM; predictor variables: visual contrast [continuous variable 0, 0.25, 0.5, 0.75 ,1], sound [0 or 1]; response variable: number of spikes during 0-300ms; Poisson distribution, log link function), allowing the classification of each neuron's responses as having a main effect ($p<0.05$) of light, sound, and/or a light-sound interaction. Neurons that were responsive to both light and sound or had a significant light-sound interaction term were classified as "light-responsive sound-modulated." To quantify the supra- or sub-linear integration of the auditory and visual responses, we calculated the linearity ratio of neurons' audiovisual responses. This ratio was defined as FR_{AV} / (FR_V + FR_A), and the sound-only response FR_A was calculated using the trials with a visual contrast of 0.

We calculated mutual information (MI) between neuronal responses and the five different visual contrast level, as well as between neuronal responses and the 12 different drifting grating directions, in order to guide the response time window used for our subsequent analyses. We calculated mutual information according to the equations (Borst and Theunissen, 1999):

$$I(R,S) = H(R) - H(R|S)$$

$$H(R) = -\sum_i p(r_i)\log_2 p(r_i)$$

$$H(R|S) = -\sum_j p(s_j)\sum_i p(r_i|s_j)\log_2 p(r_i|s_j)$$

where I(R,S) is the MI between the neuronal response R and visual stimulus S, H(R) is the entropy of neuronal response R, and H(R|S) is the entropy of neuronal response R given the stimulus S. S_j represents the stimulus parameter either visual contrast or grating direction, and r_i represents the number of spikes in a specific time window. We used a sliding 10ms time window to serially calculate MI with the visual stimulus across the neuronal response. We then averaged the MI trace across neurons to generate a population mean trace.

We quantified changes in response timing by calculating response latency, onset slope, and onset response duration. First, mean peristimulus time histograms (PSTH) were constructed for each trial type using a 10 ms sliding window. The latency was calculated as the first time bin after stimulus onset in which the mean firing rate at full contrast exceeded 1 standard deviation

above baseline. The slope Hz/ms slope was calculated from the trial onset to the time of the peak absolute value firing rate. The response duration was calculated using the full width at half maximum of the peak firing rate at stimulus onset (limited to 0-300 ms).

Orientation selectivity and direction selectivity were determined for all light-responsive neurons. The preferred direction of each direction-selective neuron was defined as the drifting grating direction that evoked the largest mean firing rate at the highest contrast level (FR_{pref}). We calculated orientation and direction-selective indices (Zhao et al., 2013) for each neuron according to:

$$\mathrm{OSI} = \frac{FR_{pref} - FR_{ortho}}{FR_{pref} + FR_{ortho}} \qquad \mathrm{DSI} = \frac{FR_{pref} - FR_{antipref}}{FR_{pref} + FR_{antipref}}$$

where FR_{ortho} and $FR_{antipref}$ are the mean firing rates in the orthogonal (90°) and anti-preferred (180°) directions, respectively. One-tailed permutation testing was performed by comparing these OSI and DSI values to pseudo OSI and DSI values obtained by 200 random shuffles of the firing rates from the pooled preferred and orthogonal or anti-preferred trials. If a neuron's actual OSI or DSI value was >95% of shuffled OSI or DSI values, the neuron was classified as "orientation-" or "direction-selective," respectively. To determine whether there were statistically significant changes in the preferred direction from the visual to audiovisual conditions, we applied a bootstrapping procedure, subsampling the visual trials for each neuron 1000 times and creating a confidence interval of the mean shift in preferred direction (degrees) for each population randomization.

We assessed and controlled for sound-induced movement as a potential confound for the audiovisual effects observed. During a subset of V1 recordings (9 recordings, 5 mice), mouse movement was tracked throughout stimulus presentation. Video recording was performed using a Raspberry Pi 4 Model B computer system with an 8MP infrared Raspberry Pi NoIR Camera V2 attachment. The camera was positioned to the front and left of the mice, which allowed capture of primarily the forepaw and whisking motion but with more limited hindpaw motion visualization. The video was converted to MP4 format, and motion was quantified by calculating the frame-by-frame difference, i.e. the percentage of pixels that differed from the prior video frame. This approach captured both whisking and locomotive behavior. This movement value for each recording was then aligned to the trials of the audiovisual stimulus from the recording trials for further analysis.

Similar to above, a GLM (predictor variables: visual contrast level, sound presence, average motion during each trial; response variable: trial spikes during 0-300ms; Poisson distribution, log link function) classified each neuron as having a main effect ($p<0.05$) of light, sound, or motion, as well as the pairwise interactions of these parameters. Light-responsive sound-modulated neurons, according to the above definition, that additionally displayed either a main effect of motion or significant light-motion or sound-motion interaction terms were classified as "motion-modulated" and were included for further analysis.

We visualized the overall distribution of mouse subject movement across trials by calculating a z-score for each trial. The movement during the trial was first compared to the baseline 100ms prior to the trial onset in order to obtain a normalized value. These normalized values were then pooled together, and we subtracted the group average and divided by the group standard deviation in order to obtain a z-score for each trial, which represented whether the mouse moved more or less compared to other trials.

In order to reconstruct peristimulus time histograms of light-responsive, sound-modulated, motion-modulated neurons, we used a separate GLM. Using a 10ms sliding window across all trials, we input the visual contrast level, sound presence, and motion during that window (discretized into five bins) as predictor variables, and the number of spikes during that window as response variables, into the GLM (Poisson distribution, log link function) to calculate coefficients for light, sound, motion, and their pairwise interactions. This approach allowed us to reconstruct the mean PSTH of individual neurons observed during each trial type by calculating:

$$\text{Spikes}_t = \exp\left(\sum_i p_{t,i} \cdot c_{t,i}\right)$$

where the spikes in time window *t* are determined by the values *p* and coefficients *c* of predictor variable *i*. From there, we used this same equation to estimate the shape of the PSTHs when varying sound and motion in order to determine differential effects these parameters had on the temporal trajectory of neurons' visual responses.

The *d'* sensitivity index (Stanislaw and Todorov, 1999; von Trapp et al., 2016) was used to calculate the directional discriminability of direction-selective neurons. The *d'* sensitivity index between two directions θ_1 and θ_2 is calculated as:

$$d' = \frac{\mu_{\theta_1} - \mu_{\theta_2}}{\sqrt{\frac{1}{2}(\sigma_{\theta_1}^2 + \sigma_{\theta_2}^2)}}$$

where μ_θ and σ_θ are the response mean and standard deviation, respectively, for direction θ. For each neuron, the sensitivity index was calculated in a pairwise manner for preferred direction versus all other directions and then aligned relative to the preferred direction in order to test sensitivity index as a function of angular distance from preferred direction.

We used a maximum likelihood estimate approach (Montijn et al., 2014; Meijer et al., 2017) to decode the visual stimulus direction from the neuronal responses based on Bayes rule:

$$P(\theta|A_{trial}) = \frac{P(A_{trial}|\theta)P(\theta)}{P(A_{trial})}$$

For decoding using individual neurons, the likelihood $P(A_{trial}|\theta)$ for each orientation or direction was computed based on the Poisson response distribution across all trials of that orientation or direction, with a leave-one-out cross-validation technique in which the probe trial (A_{trial}) was excluded from the training data. The prior $P(\theta)$ was uniform, and the normalization term $P(A_{trial})$ was similarly applied to all directions. Therefore, the posterior probability $P(\theta|A_{trial})$ was proportional to and based on evaluating the likelihood function at the value of the probe trial. For orientation-selective neurons, decoding was performed between the preferred and orthogonal orientations, and for direction-selective neurons, decoding was performed between the preferred and anti-preferred directions. For decoding using populations of neurons, neurons were pooled across recording sessions. A similar approach was used; however, here, the posterior probability $P(\theta|A_{pop})$ was proportional to the joint likelihood $P(A_{pop}|\theta)$ of the single-trial activity across all N neurons in the population (A_{pop}):

$$P(A_{pop}|\theta) = \prod_{neuron\ i}^{N} P(A_{trial}|\theta)_i$$

With this population-based analysis, pairwise decoding was performed between every orientation and its orthogonal orientation (1 of 2 options), as well as decoding one direction from all possible directions (1 of 12 options).

Additionally, we used a support vector machine (SVM) to corroborate the findings of the MLE-based decoder. The SVM was implemented using MATLAB's fitcsvm function with a linear kernel to predict the drifting grating direction based on single-trial population responses. Similarly, a leave-one-out cross-validation technique was used, and pairwise decoding was performed between every combination of two stimulus directions.

Statistics

Figure data are displayed as means with standard error of the mean (SEM), unless otherwise noted. Shapiro-Wilk tests were used to assess normality, and the statistical tests performed are indicated in the text, figures, and Table 1. For multi-group and multivariate analysis (e.g., ANOVA and Kruskal-Wallis tests) in which a significant ($p<0.05$) interaction was detected, we subsequently performed a post hoc Bonferroni-corrected test. P-values reported as 0 are too small to be accurately calculated by Matlab ($p<2.2e-301$), due to characteristically large data sets. See Table 1 for a detailed summary of statistical results and post hoc comparisons.

Results

Sound enhances the light-evoked firing rate of a subset of V1 neurons

Previous work identified that sound modulates visual responses in V1 (Ibrahim et al., 2016; Meijer et al., 2018; McClure and Polack, 2019), yet how that interaction affects stimulus encoding in individual neurons and as a population *in the awake brain* remains unclear. Furthermore, whether that interaction can be exclusively attributed to sound or to sound-induced motion is controversial (Bimbard et al., 2021). To elucidate the principles underlying audiovisual integration, we presented audiovisual stimuli to awake mice while performing extracellular recordings in V1 (Figure 1A). The visual stimulus consisted of drifting gratings in 12 directions presented at 5 visual contrast levels (Figure 1B). On half of the trials, we paired the visual stimulus with a 70 dB burst of white noise from a speaker positioned next to the screen (Figure 1C), affording 10 trials of each unique audiovisual stimulus condition (Figure 1C). Twelve recording sessions across six mice were spike sorted, and the responses of these sorted neurons were organized by trial type to compare across audiovisual stimulus conditions. Figure 1D-G demonstrates an example unit tuned for gratings aligned to the 30°-210° axis whose baseline and light-evoked firing rate are increased by the sound.

Sound modulated the activity of the majority of V1 neurons. We used a generalized linear model (GLM) to classify neurons as light-responsive and/or sound-responsive based on their firing rate at the onset (0-300 ms) of each trial. We chose to classify neurons based on their onset response because the first 300 ms had the highest mutual information with both the visual contrast level as well as the drifting grating orientation (Figure 2A-C). Using this classification method, we found that 86.2% (703/816) of units were responsive to increasing visual stimulus contrast levels, and of these visually responsive units, 80.1% (563/703 neurons, 12 recording sessions in 6 mice) were significantly modulated by the presence of sound (Figure 3A).

Because the depth electrode penetrated all layers of V1, we were able to estimate the depth of each unit based on the amplitude of the spike waveform recorded by local electrodes. Surprisingly, we found that the majority of units across each depth were either sound-responsive or sound-modulated light-responsive (Figure 3F-H). We then constructed an average PSTH from the response profiles of sound-modulated light-responsive neurons, which revealed that the largest change in light-evoked firing rate occurred at the onset of the stimulus (Figure 3B). Averaged across neurons, we found a robust increase in the magnitude of the visually evoked response across visual contrast levels (Figure 3C; p(vis)=1.2e-100, p(aud)=1.6e-88, p(interact)=5.7e-4, paired 2-way ANOVA; $p_{c=0}$=2.1e-51, $p_{c=0.25}$=2.6e-62, $p_{c=0.5}$=5.7e-75, $p_{c=0.75}$=1.1e-81, $p_{c=1}$=2.0e-81, post hoc Bonferroni-corrected paired t-test, Table 1). This difference was driven by the majority of neurons (95%) that increased their firing rate in the presence of sound. However, some neurons exhibited lower light-evoked and sound-evoked firing rates relative to baseline.

This change in firing rate can be potentially supra-linear, linear or sub-linear based on whether the audiovisual response is, respectively, greater, equal or less than the sum of the unimodal light-evoked and sound-evoked firing rates. At medium to high visual contrast levels, integration of the audiovisual stimulus was predominantly supra-linear (Figure 3D-E; p=1.6e-12, Kruskal-Wallis test; $p_{c=0.25}$=0.053, $p_{c=0.5}$=0.004, $p_{c=0.75}$=4.6e-8, $p_{c=1}$=2.1e-5, post hoc Bonferroni-corrected Wilcoxon signed rank test, Table 1). In summary, these results show that sound supra-linearly increases the magnitude of the light-evoked response in the majority of V1 neurons.

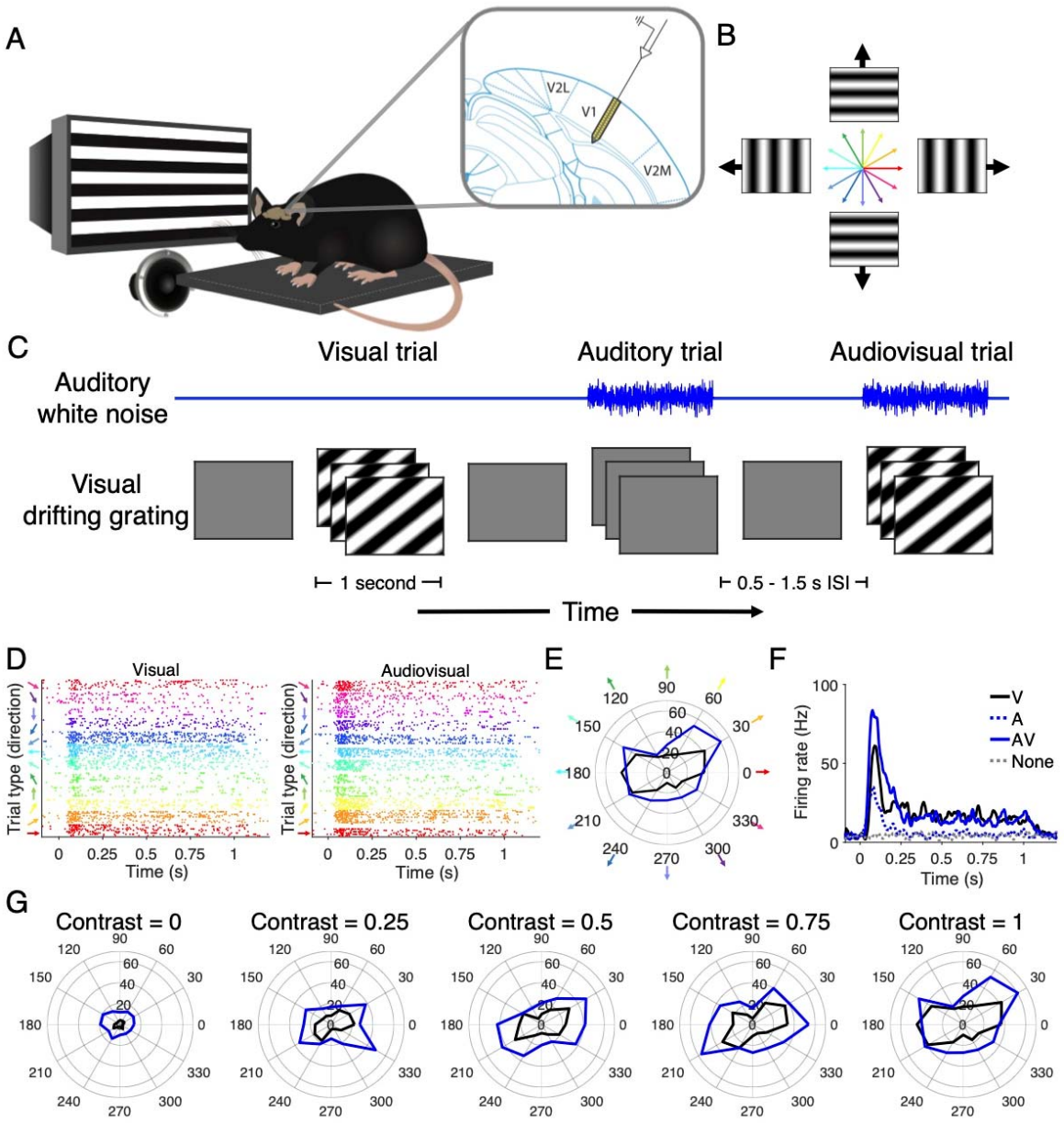


Figure 1 | **Audiovisual stimulus presentation** (A) Diagram (left) demonstrating that mice were head-fixed and presented with audiovisual stimuli from the right spatial field while electrophysiological recordings were performed in V1 (right). (B) Visual stimuli consisted of drifting gratings of 12 directions. (C) Auditory, visual, and audiovisual trials were randomly ordered and spaced with variable inter-stimulus intervals. (D) Raster plots of visual (left) and audiovisual (right) trials of an example neuron exhibiting visual orientation tuning. (E) Polar plot demonstrating the orientation tuning and magnitude of response (Hz) of the same example neuron in *E*. (F) PSTH of the same neuron in *E* demonstrating enhanced firing in response to audiovisual stimuli compared to unimodal stimuli. (G) Example neuron in *E* displays enhanced firing rate with sound across visual contrast levels.

Sound reduces the orientation- and direction-selectivity of tuned neurons

Having observed sound-induced changes in the magnitude of the visual response, we next assessed whether these changes in magnitude affected neuronal tuning profiles in the awake brain. Mouse V1 neurons typically have receptive fields tuned to a specific visual stimulus orientation and, to a lesser extent, stimulus direction (Métin et al, 1988; Rochefort et al., 2011;

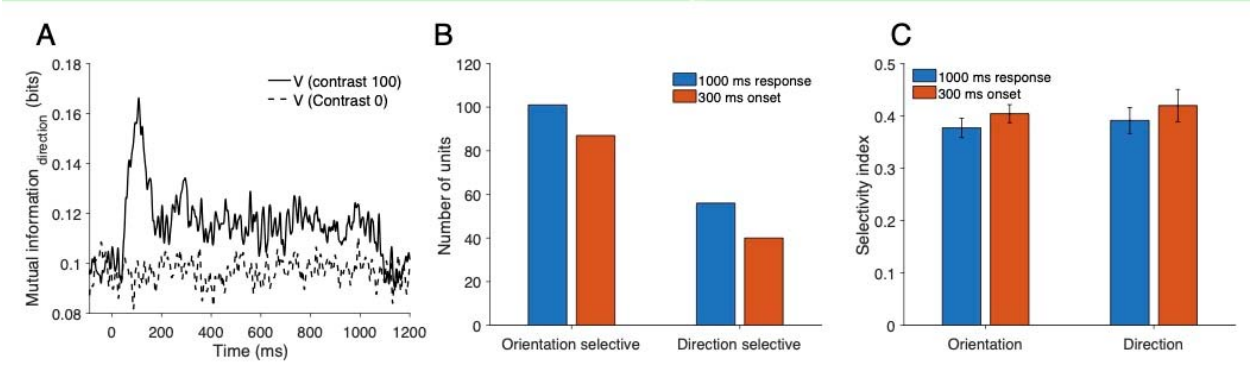


Figure 2 | Mutual information of neuronal responses, and depth distribution of responsive neurons (A) Mutual information (MI) between neuronal responses and drifting grating direction, averaged across neurons. The solid line is MI at full visual contrast, and the dotted line is MI at zero visual contrast, serving as a negative control. (B) We found a slight reduction in the number of neurons classified as orientation or direction selective when based on the initial 300 ms onset response compared to the entire 1000 ms response. (C) The OSI and DSI of classified neurons was slightly higher when calculated using the initial 300 ms onset response compared to the whole 1000 ms response.

Fahey et al., 2019). We first tested whether sound altered tuning preferences of V1 neurons. In light-responsive neurons, we calculated the orientation and direction-selective indices (OSI and DSI) as well as pseudo indices based on random permutations of the trials (see Methods), and classified neurons in which the true indices were >95% of the pseudo indices as "orientation-" or "direction-selective." Using this stringent selection criterion, we found that 13.9% (78/563) of neurons were orientation-selective, whereas 2.1% (12/563) were direction-selective. In these neurons, we determined their preferred grating orientation or direction by calculating half the complex phase of the response profile at full visual contrast (Niell and Stryker, 2008). We observed little shift in the preferred direction from the visual to audiovisual condition (Figure 3I). This shift in visual tuning preference may be due to auditory input, or it may reflect noise in the neuronal responses. To test this, we performed an additional permutation test by repeatedly sampling the visual responses. We found that the resulting distribution of preferred direction shifts resembled the observed distribution under the audiovisual condition and the observed mean shift in degrees was within the limits of the sampled distribution (Figure 3I inset). Furthermore, the observed mean shift was below the 5th percentile of the sampled distribution (Figure 3I inset). Therefore, the preferred orientation and direction of selective neurons was more reliable between the visual and audiovisual conditions than what would be predicted by neuronal noise alone.

In addition to testing a shift in preferred direction, we investigated whether sound altered the neurons' tuning selectivity. Tuning selectivity captures how strongly an individual neuron responds to stimuli of a certain condition as compared to others, e.g. grating orientation and drift direction (Zhao et al., 2013). We found a small reduction in the OSI from the visual to audiovisual conditions (Figure 3J; p=0.0018, paired Student's t-test), which may reflect disproportionate changes in firing rate at the preferred versus orthogonal directions. We also found a reduction in the DSI in the presence of sound (Figure 3K; p=0.021, paired Student's t-test). Combined, these results suggest that sound's enhancement of the magnitude of light-evoked responses has minimal or potentially diminishing effects on the tuning selectivity of neurons.

Changes in neuronal response latency, onset duration, and variability in audiovisual compared to visual conditions

Behaviorally, certain cross-modal stimuli elicit shorter reaction times than their unimodal counterparts (Diederich and Colonius, 2004; Colonius and Diederich, 2017; Meijer et al., 2018).

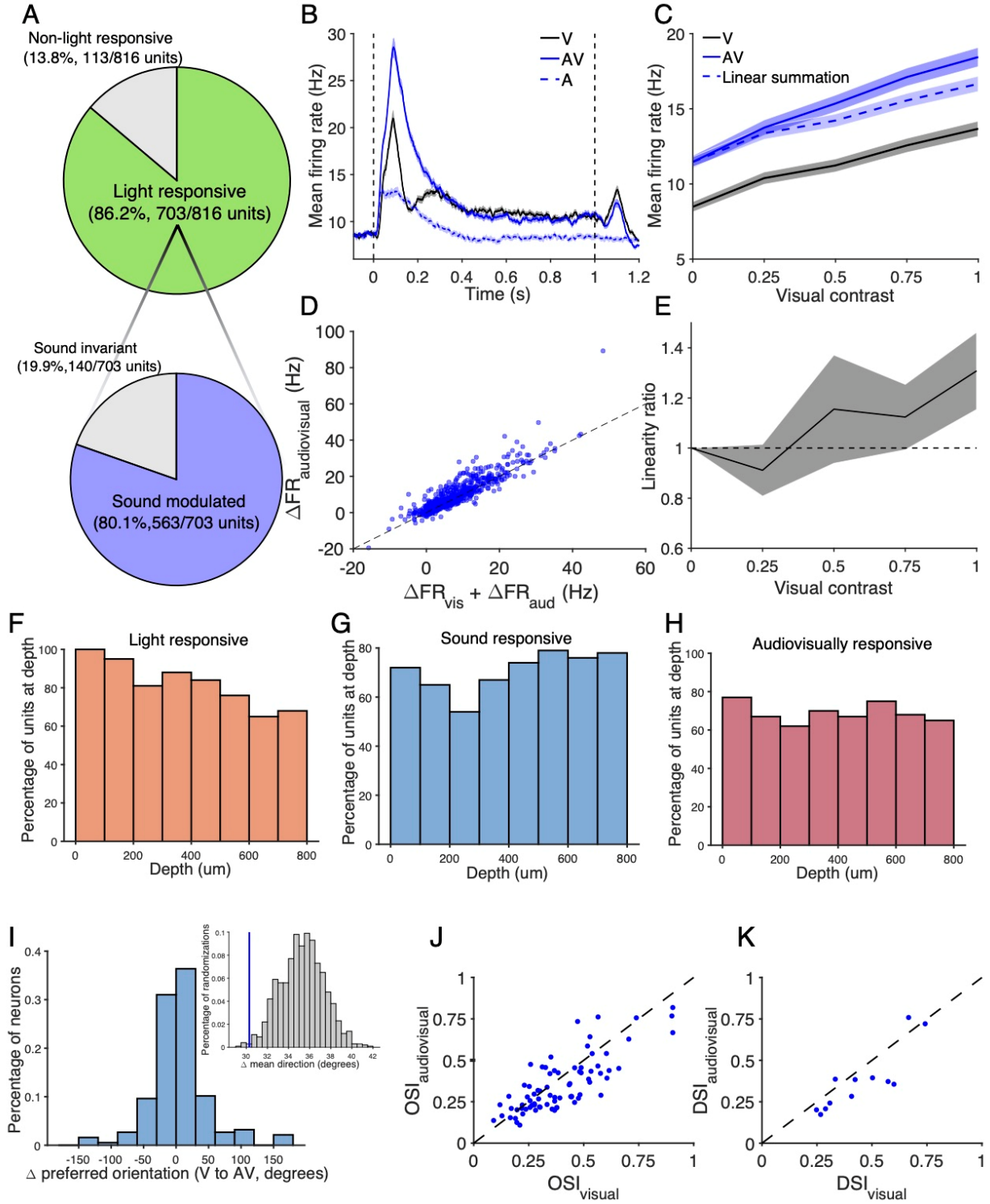


Figure 3 | Sound enhances visual responses in a supra-linear manner (A) Sound modulates visually evoked activity in 80.1% of light-responsive neurons in V1. (B) Comparison of visual, auditory, and audiovisual PSTHs averaged across all light-responsive sound-modulated neurons. Visual and audiovisual PSTHs correspond to the highest visual contrast level. (C) The magnitude of audiovisual onset responses (0-300ms) is greater than that of the visual

response in light-responsive sound-modulated neurons (n=563, p(vis)=1.2e-100, p(aud)=1.6e-88, p(interact)=5.7e-4, 2-way repeated measures ANOVA; post hoc Bonferroni-corrected paired t-test). The expected linear sum of the unimodal auditory and visual responses is included. (D) At full visual contrast, the observed audiovisual response in the majority of neurons is greater than the linear sum of the unimodal auditory and visual responses. (E) A linearity ratio above 1 demonstrates audiovisual responses in V1 represent supra-linear integration of the unimodal signals (n=563, p=1.6e-12, Kruskal-Wallis test, post hoc Bonferroni-corrected Wilcoxon signed rank test). (F-H) Histograms demonstrating the percentage of neurons at each 100 um depth bin that were classified as light, sound, and audiovisually responsive, based on the recording electrode with the largest spike waveform amplitude. (I) Histogram depicting changes in preferred drifting grating directions, calculated using half of the complex phase, with sound in orientation-selective neuron. In the inset, the observed mean change in preferred direction (blue) is within the expected distribution (gray) based on shuffled permutations using the visual response variability. (J) A slight reduction in the orientation selectivity index was observed in orientation-selective neurons (n=78, p=0.0018, paired t-test). (K) A slight reduction in the direction selectivity index was also observed in direction-selective neurons (n=12, p=0.021, paired t-test).

Therefore, we hypothesized that sound reduces the latency of the light-evoked response at a neuronal level as well. For each neuron, we calculated the response latency as the first time bin after stimulus onset at which the firing rate exceeded 1 standard deviation above baseline (Figure 4A), and found that sound reduced the response latency across contrast levels (Figure 4B; p(vis)=6.9e-4, p(aud)=6.8e-15, p(interact)=0.045, paired 2-way ANOVA; $p_{c=0.25}$=2.3e-4, $p_{c=0.5}$=7.1e-12, $p_{c=0.75}$=4.6e-5, $p_{c=1}$=9.9e-4, post hoc Bonferroni-corrected paired t-test, Table 1). We additionally calculated the slope of the onset response of light-responsive sound-modulated neurons, measured from trial onset until the time at which each neuron achieved its peak firing rate (Figure 4C). We found that sound increased the slope of the onset response (Figure 4D; p(vis)=3.5e-121, p(aud)=2.7e-15, p(interact)=0.038, paired 2-way ANOVA; $p_{c=0.25}$=1.4e-4, $p_{c=0.5}$=8.9e-13, $p_{c=0.75}$=3.6e-12, $p_{c=1}$=5.5e-8, post hoc Bonferroni-corrected paired t-test, Table 1), both indicating that the response latency was reduced in the audiovisual condition compared to the visual condition. Additionally, the duration of the light-evoked response, defined as the full width at half maximum of the peak onset firing rate, increased in the presence of sound (Figure 4E,F; p(vis)=1.3e-10, p(aud)=8.7e-98, p(interact)=0.23, paired 2-way ANOVA). Both of these timing effects were preserved across contrast levels. Therefore, the latency and onset duration of audiovisual responses of V1 neurons is enhanced compared to visual responses.

Having observed changes in response magnitude and timing, we next investigated the effect of sound on the variability of light-evoked responses. If individual neurons encode the visual stimulus using changes in their firing rate, a more consistent response would entail less spread in the response magnitude relative to the mean response across trials of a single stimulus type. We quantified this relationship using the coefficient of variation (CV) defined as the ratio of the standard deviation to the response mean (Gur et al., 1997). We hypothesized that sound reduces the CV of light-evoked responses, corresponding to reduced response variability and higher signal-to-noise ratio. Figure 4G depicts the relationship between response magnitude and CV in an example sound-modulated light-responsive neuron, demonstrating that increased response magnitude correlates with reduced CV. Consistent with sound increasing the visual response magnitude in the majority of sound-modulated light-responsive neurons (Figure 4), we observed a reduction of CV in the audiovisual condition relative to the visual condition when averaged across these neurons (Figure 4H; p(vis)=0.28, p(aud)=4.2e-103, p(interact)=0.38, paired 2-way ANOVA). Taken together, these results indicate that sound not only modulates the magnitude of the visual response (Figure 4), but also improves the timing and consistency of individual neurons' responses (Figure 4).

Sound-induced movement does not account for sound's effect on visual responses

It is known that whisking and locomotive behaviors modulate neuronal activity in mouse visual cortex (Niell and Stryker, 2010) and auditory cortex (Nelson et al., 2013; Schneider and Mooney,

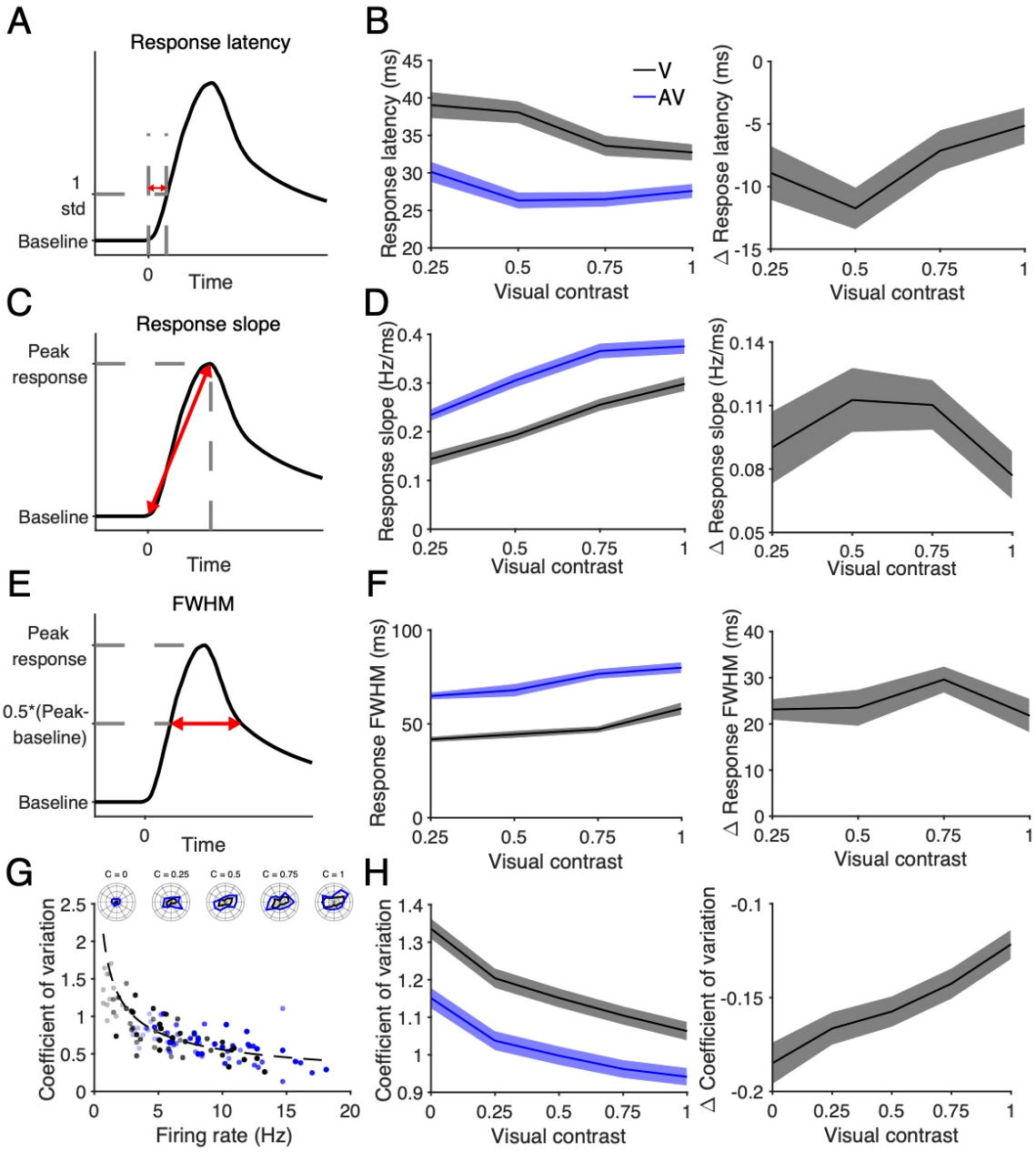


Figure 4 | Changes in neuronal response latency, onset duration, and variability in audiovisual compared to visual conditions (A) Diagram of the calculation of response latency, the first time bin in which the FR exceeds 1 std above baseline. (B) Audiovisual response latency is less than that of the visual response (left: absolute, right: difference; p(vis)=6.9e-4, p(aud)=6.8e-15, p(interact)=0.045, paired 2-way ANOVA, post hoc Bonferroni-corrected paired t-test, Table 1). (C) Diagram of the calculation of response onset slope, the peak change in FR over the latency to peak response. (D) The slope of the audiovisual response is greater than that of the visual response (left: absolute, right: difference; n=563, p(vis)=3.5e-121, p(aud)=2.7e-15, p(interact)=0.038, paired 2-way ANOVA, post hoc Bonferroni-corrected paired t-test). (E) Diagram of the calculation of FWHM, the width of the onset response at half maximum FR. (F) The FWHM of the audiovisual response is greater than that of the visual response (left: absolute, right: difference; n=367, p(vis)=1.3e-10, p(aud)=8.7e-98, p(interact)=0.23 paired 2-way

ANOVA). (G) An example neuron demonstrating that increased response magnitude corresponds to lower CV according to an inverse square root relationship. The black and blue dots represent visual and audiovisual responses, respectively, and the dot transparency corresponds to visual contrast level. The dotted lines are fitted y=c/sqrt(x) curves, where c is a constant. The above inset is the polar plots corresponding to the example neuron. (H) Lower coefficient of variation indicates reduced response variability in audiovisual compared to visual responses (left: absolute, right: difference; n=563, p(vis)=0.28, p(aud)=4.2e-103, p(interact)=0.38, paired 2-way ANOVA).

2018; Bigelow et al., 2019). Therefore, having established that sound robustly modulates visual responses (Figure 3), we tested whether and to what extent these observed changes were more accurately attributable to sound-induced movement. In an additional cohort of mice, we performed V1 extracellular recordings with the same audiovisual stimuli described above while recording movement activity of the mice throughout stimulus presentation. Movement was calculated as frame-by-frame difference in video pixels, an approach that captured both locomotive and subtle whisking behavior. Despite being head-fixed to afford stable electrophysiological recordings, the mice were positioned on a smooth stage that freely allowed volitional locomotion. We found that both visual and auditory stimuli did evoke whisking and locomotive behavior in mice (Figure 5A). We first compared this movement during the stimulus trial to the 100ms baseline prior to trial onset and found that movement was higher during audiovisual trials compared to visual trials (Figure 5B; p=9.1e-5, paired t-test). However, there were many visual trials in which substantial movement occurred, as well as audiovisual trials in which little movement was detected (Figure 5C). Because of this large variability in sound-induced movement, we were able to control for movement when comparing visual and audiovisual activity in the recorded neurons.

We used a GLM to classify each neuron as light-, sound-, and/or motion-responsive based on the neuron's firing rate and mouse's movement activity during the onset (0-300ms) of the trial. The vast majority of light-responsive neurons, 71.1% (249/350), displayed both sound- and motion-modulated visual responses (Figure 5D). 11.1% (39/350) and 5.2% (18/350) of light-responsive neurons were purely sound- or motion-modulated, respectively. An additional 12.6% (44/350) were invariant to sound or motion. We then compared the visually and audiovisually evoked firing rates of neurons when controlling for movement. Among sound- and motion-modulated light-responsive neurons, the firing rate was higher on audiovisual trials than visual trials when movement was held constant (Figure 5E), especially when mice showed limited movement. On trials in which the mice were largely stationary (z-score<-0.5, 43% of visual trials, 32% of audiovisual trials) or displayed moderate levels of movement (-0.5<z-score<1.5, 51% of visual trials, 57% of audiovisual trials), the mean firing rate of neurons was 54-62% higher when sound was presented than when sound was absent. The firing rates under the two stimulus conditions converged on trials in which the mice displayed high movement activity (z-score>1.5, 4.8% of visual trials, 11% of audiovisual trials; Figure 5E,F; p(move)=0.010, p(aud)=1.4e=13, p(interact)=1.8e-8, unbalanced 2-way ANOVA; $p_{stationary}$=1.5e-14, $p_{low\ motion}$=7.1e-10, $p_{high\ motion}$=0.6, post hoc Bonferroni-corrected two-sample t-test, Table 1). Notably, increasing movement activity was correlated with increased firing rates on visual trials, but was correlated with decreasing firing rates among audiovisual trials (Figure 5F). These results indicate that sound modulated visually evoked neuronal activity even when accounting for sound-induced movement in awake mice, with the exception of when mice showed the highest amount of movement, during which there was little effect of sound on firing rates.

Sound and movement have distinct and complementary effects on visual responses

To further parse out the role of sound and movement on audiovisual responses, we used a separate GLM to capture the time course of these parameters' effects on visually evoked

activity. For each neuron, we used a GLM with a sliding 10ms window to reconstruct the PSTH based on the visual contrast level, sound presence, and movement during that time window (Figure 6A). Figure 6B shows an example neuron in which the GLM accurately captures the light-evoked, sound-evoked, and audiovisually evoked PSTHs using the average movement for each trial type. Across neurons, the GLM-estimated PSTHs accurately reconstructed observed PSTHs, with the

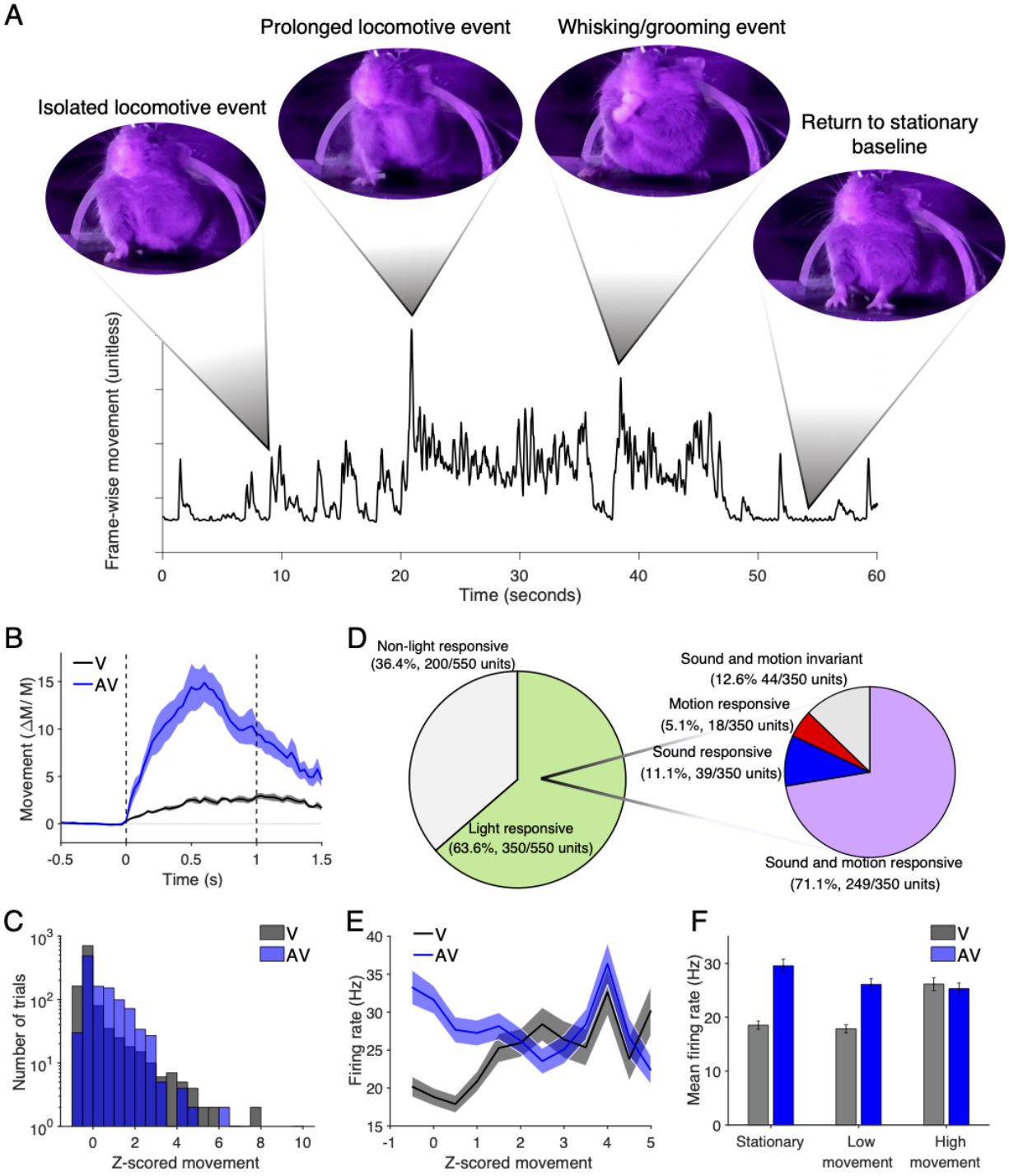


Figure 5 | Sound modulates visual activity when controlling for stimulus-induced movement (A) Sample trace from example video recording demonstrating the detection and quantification of locomotive and whisking events

during electrophysiological recordings. (B) Mice displayed more movement response to audiovisual trials than in visual trials (n=9 recording sessions; p=9.1e-5, paired t-test). (C) Histogram of trials' z-scored movements show a range of levels of movement during both visual and audiovisual trials. (D) Venn diagram demonstrating that 87% of light-responsive neurons exhibited some combination of sound- and movement-responsiveness. (E) Comparison of firing rate of sound- and motion-modulated light-responsive neurons across trials with a range of z-scored movement. (F) Responses to audiovisual stimuli evoke larger magnitude responses than visual stimuli when mice were stationary (z-score<-0.5) or displayed low to moderate movement (-0.5<z-score<1.5), but responses were not significantly different when mice displayed the highest amount of movement (z-score>1.5; p(motion)=0.001, p(aud)=1.4e-13, p(interact)=1.8e-8, 2-way ANOVA, post hoc Bonferroni-corrected two-sample t-test)

highest correlation when all parameters were included in the estimate (Figure 6C-E). We leveraged the coefficients fit to each neuron (Figure 6A) to estimate the unique contribution of each predictor to the firing rates as a function of time (see Materials and Methods). In the absence of movement, sound predominantly enhanced neuronal activity at the onset of the visual response and suppressed activity during the response's sustained period (Figure 6F; n=295 fitted neurons, paired t-test at each time window [1391], α=3.6e-5). Conversely, movement had little effect on the onset activity in the absence of sound, but rather enhanced firing rates during the response's sustained period (Figure 6G; n=295 fitted neurons, paired t-test at each time window [1391], α=3.6e-5). Together, sound and movement had complementary effects in which both the onset and sustained portions of the visual response were enhanced (Figure 6H; n=295 fitted neurons, paired t-test at each time window [1391], α=3.6e-5). Again notably, the peak onset response under the audiovisual condition was lower when movement was included in the estimate (Figure 6H). These findings indicate not only that movement is unable to account for the changes in onset response reported above, but also that sound and motion have distinct and complementary effects on the time course of visually evoked activity in V1.

Decoding of the visual stimulus from individual neurons is improved with sound

Behaviorally, sound can improve the detection and discriminability of visual responses, however whether that improved visual acuity is reflected in V1 audiovisual responses is unknown. Despite many studies reporting how sound affects visual responses in V1, whether these changes result in improved neuronal encoding of the visual stimulus, especially in the awake brain, has yet to be directly demonstrated. The increase in response magnitude and decrease in CV suggest that sound may improve visual stimulus discriminability in individual V1 neurons. Consistent with these changes in response magnitude and variability, we observed sound-induced improvements in the *d'* sensitivity index between responses to low contrast drifting grating directions among orientation- and direction-selective neurons (Figure 7A,B), further indicating improved orientation and directional discriminability in individual neurons. To directly test this hypothesis, we used the neuronal responses of individual neurons to estimate the visual stimulus drifting grating orientation and direction. We trained a maximum likelihood estimate (MLE)-based decoder (Montijn et al., 2014; Meijer et al., 2017) on trials from the preferred and orthogonal orientations in orientation-selective neurons and on trials from the preferred and anti-preferred directions in direction-selective neurons. We used leave-one-out cross-validation and cycled the probe trial through the repeated trials of the stimulus condition in order to calculate the mean decoding performance. The MLE decoder's output was the orientation or direction with the maximum posterior likelihood based on the training data (Figure 7C). This decoding technique achieves high decoding accuracy (Figure 7D). When averaged across sound-modulated orientation-selective neurons, decoding performance was improved on audiovisual trials compared to visual trials (Figure 7E; p(vis)=4.8e-112, p(aud)=7.8e-4, p(interact)=0.71, paired 2-way ANOVA), with the greatest improvements at low to intermediate contrast levels (Figure 7F). We applied this approach to sound-modulated direction-selective

643 units and found similar trends towards improvements at low contrast levels (Figure 7F,H;
644 p(vis)=2.1e-4, p(aud)=0.18, p(interact)=0.78, paired 2-way ANOVA), limited by fewer and
645 weaker direction-selective neurons in V1. These results demonstrate that sound-induced
646 changes in response magnitude and consistency interact in order to improve neuronal
647 representation of the visual stimulus in individual neurons.
648
649
650

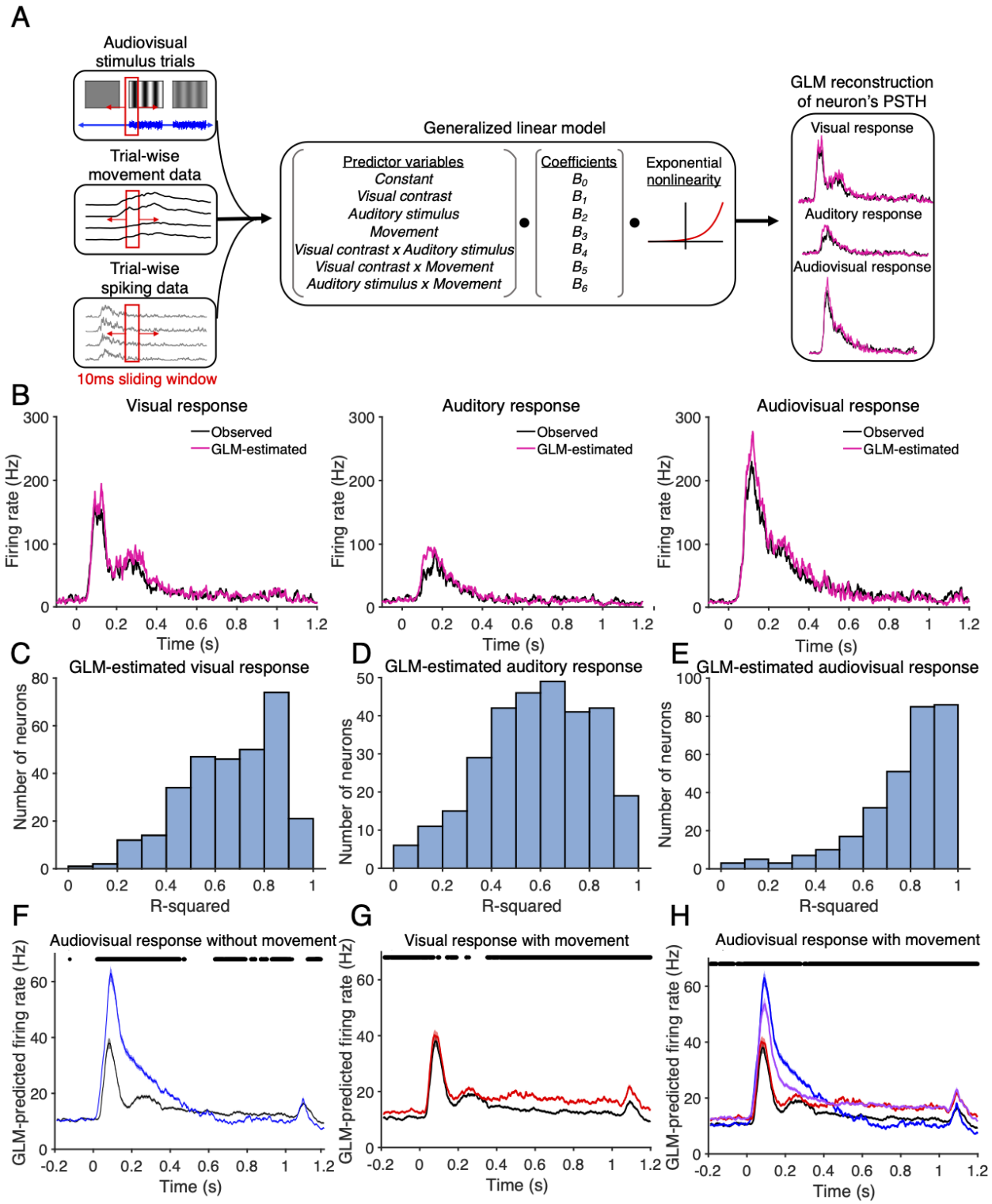


Figure 6 | Sound and movement modulate visual responses in distinct but complementary ways (A) Diagram illustrating the use of a GLM to reconstruct individual neurons' PSTHs based on neuronal responses and mouse movement during stimulus presentation. The GLM was then used to predict the time course of neuronal responses audiovisual stimuli with and without movement. (B) Observed trial-averaged PSTHs for visual-only (left), auditory-only (middle), and audiovisual (right) trials overlaid with GLM estimates based on the selected stimulus features. (C-E) Histograms demonstrating R^2 values of the GLM-estimated PSTHs, averaged across sound- and motion-

modulated light-responsive neurons. Moderate to high R^2 values across the population indicate a good ability for the GLM to estimate neuronal firing rates. (F-H) GLM-predicted visually evoked PSTHs with and without sound and motion. Asterisks indicate time windows in which there was a significant difference between the *light* prediction and the *light+sound, light+motion,* and *light+sound+motion* predictions, respectively. (F) Excluding motion highlights that sound primarily enhances the onset response. Asterisks indicate time windows in which there was a significant difference (n=295 fitted neurons; paired t-test, α=3.6e-5). (G) Excluding sound highlights that motion primarily enhances the sustained portion of the response. Asterisks indicate time windows in which there was a significant difference (n=295 fitted neurons; paired t-test, α=3.6e-5). (H) Sound and motion together enhance both the onset and sustained periods of the visually evoked response. (n=295 fitted neurons; paired t-test, α=3.6e-5).

Population-based decoding of the visual stimulus improves with sound

V1 uses population coding to relay information about the various stimulus dimensions to downstream visual areas (Montijn et al., 2014, Berens et al., 2012), so we next tested whether these improvements in visual stimulus encoding in individual neurons extended to the population level. We began by training a support vector machine (SVM) to perform pairwise classification of visual drifting grating directions based on neuronal population activity. We again used a leave-one-out cross-validation approach when training and testing the SVM (Figure 8A). Unsurprisingly, decoding accuracy improved as more neurons were included in the population (Figure 8B), achieving an accuracy of ~90% when averaged across all pairwise orientation comparisons. At full visual contrast, there was little difference between the performance on visual and audiovisual trials.

However, at low to intermediate visual contrast levels, classification performance robustly increased on audiovisual trials as compared to visual trials (Figure 8D). This improvement in performance was greatest when comparing orthogonal drifting grating orientations (Figure 8E; p(vis)=1.8e-61, p(aud)=1.9e-8, p(interact) = 2.4e-4, 2-way ANOVA; $p_{c=0}$=0.12, $p_{c=0.25}$=0.0016, $p_{c=0.5}$,=0.0014, $p_{c=0.75}$=0.0023; $p_{c=1}$=1, post hoc Bonferroni-corrected paired t-test, Table 1). A similar improvement was also observed in decoding opposite drifting grating directions (Figure 8F, p(vis)=1.1e-21, p(aud)=9.0e-9, p(interact)=0.0019, 2-way ANOVA; $p_{c=0}$=0.55, $p_{c=0.25}$=5.3e-5, $p_{c=0.5}$=0.0036, $p_{c=0.75}$=0.17, $p_{c=1}$=0.0036, post hoc Bonferroni-corrected paired t-test, Table 1). These results indicate that sound improves neuronal population encoding of grating orientation and drift direction.

Similar performance levels were also observed when decoding drifting grating orientation and direction using an MLE-based population decoder, indicating that the results were not specific to the decoding algorithm. Again, performance improved with increasing population sizes (Figure 8C), and accuracy was higher on audiovisual trials than visual trials (Figure 8G-I; orientation: p(vis)=2.3e-66, p(aud)=0.61, p(interact)=9.6e-11, 2-way ANOVA; $p_{c=0}$-5.8e-4, $p_{c=0.25}$=1.8e-4, $p_{c=0.5}$=0.3, $p_{c=0.75}$=0.53, $p_{c=1}$=0.15, post hoc Bonferroni-corrected paired t-test, Table 1; direction: p(vis)=4.6e-26, p(aud)=0.51, p(interact)=4.1e-6, 2-way ANOVA; $p_{c=0}$=0.037, $p_{c=0.25}$=6.4e-6, $p_{c=0.5}$=0.036, $p_{c=0.75}$-0.16, $p_{c=1}$=0.14, post hoc Bonferroni-corrected paired t-test, Table 1).

Expanding on the SVM approach, the MLE-based decoder allowed us to perform not only pairwise classification, but also classification of 1 out of all 12 drifting grating directions. When trained and tested in this fashion, MLE decoding performance again improved at low to intermediate contrast levels on audiovisual trials (Figure 8J-L), before reaching asymptotic performance of ~45% at full visual contrast (Figure 8L; p(vis)=2.2e-92, p(aud)=1.9e-5, p(interact)=2.7e-11, 2-way ANOVA; $p_{c=0}$=0.012, $p_{c=0.25}$=1.4e=10, $p_{c=0.5}$=0.48, $p_{c=0.75}$=0.0013, $p_{c=1}$=0.5, post hoc Bonferroni-corrected paired t-test, Table 1). Similar results were found when organizing the neurons by recording session instead of pooling all neurons together (data not shown). Taken together, these results

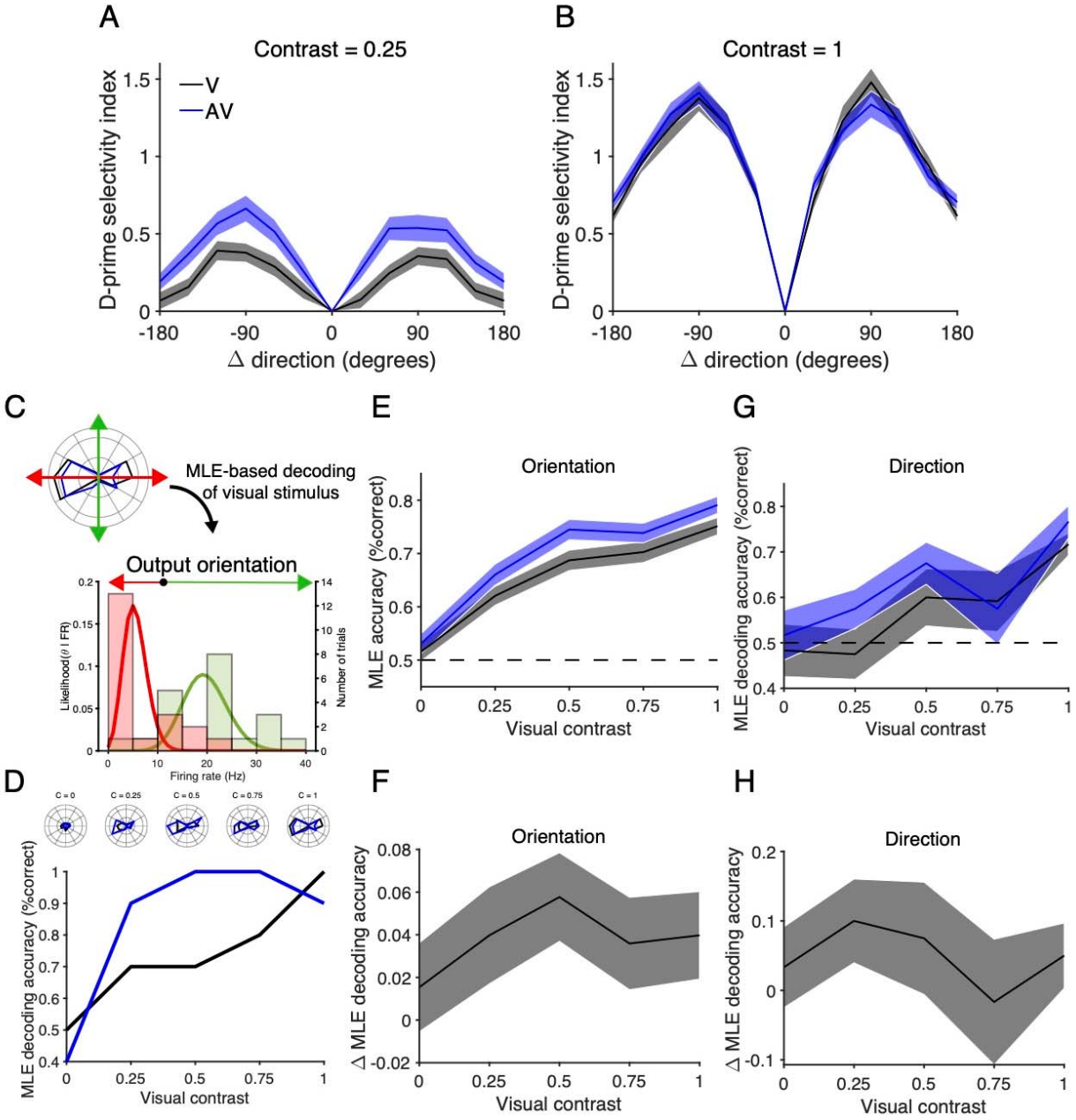
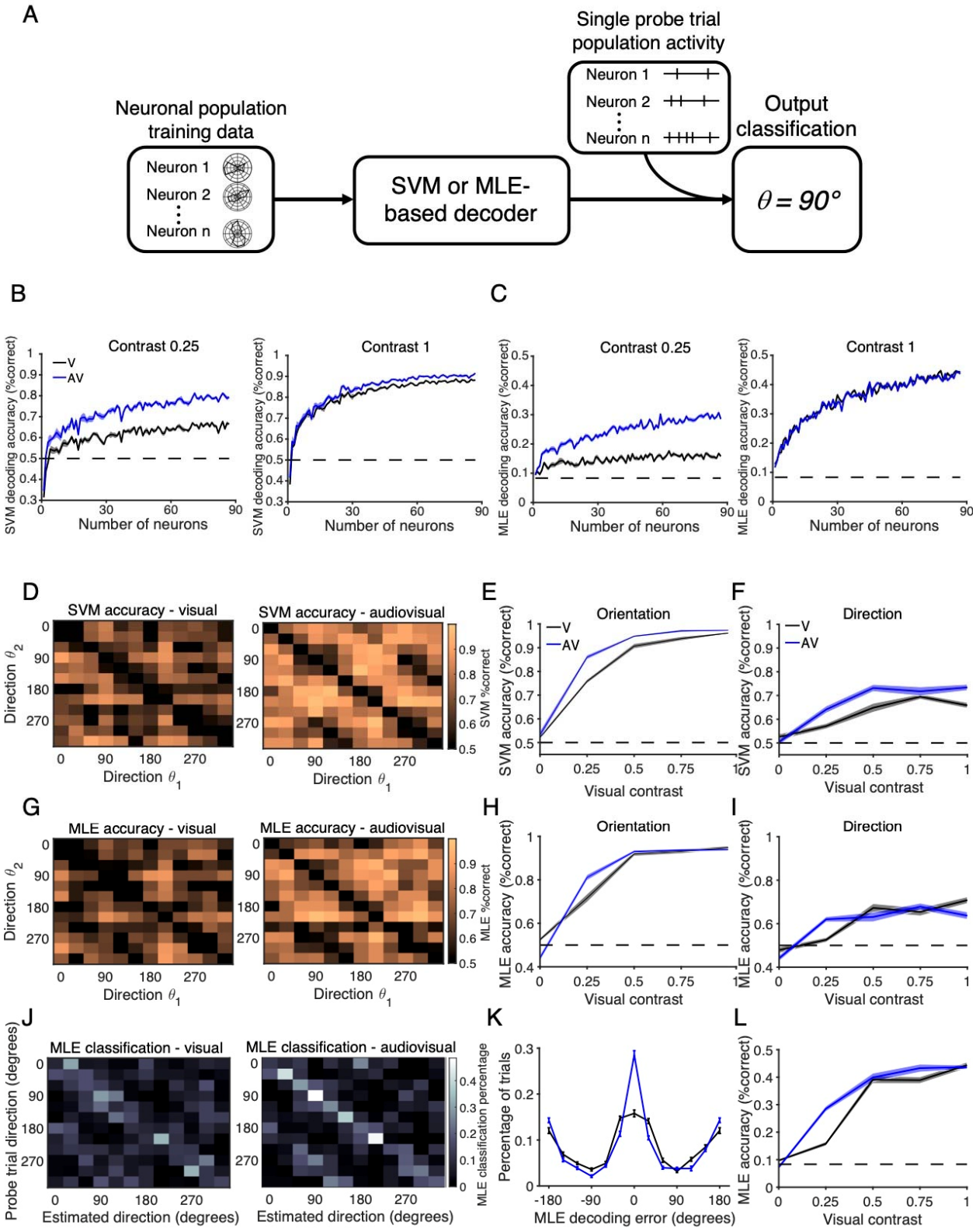


Figure 7 | Sound improves decoding of drifting grating direction and orientation in individual neurons (A-B) The *d'* sensitivity index between neuronal responses to drifting grating directions, averaged across orientation- and direction-selective neurons. Enhancements are observed at low visual contrast (A), whereas minimal changes are present at full contrast (B). (C) Diagram illustrating MLE-based decoding of an individual neuron's preferred versus orthogonal orientations. (D) Performance of the MLE decoder, trained on an example orientation-selective neuron, in decoding the neuron's preferred versus orthogonal orientations. The neuron's polar plots are shows in the above inset. (E-F) Absolute (E) and difference (F) in decoding accuracy of preferred versus orthogonal orientations, averaged across sound-modulated orientation-selective neurons, demonstrating higher performance in the audiovisual condition (n=78, p(vis)=4.8e-112, p(aud)=7.8e-4, p(interact)=0.71, paired 2-way ANOVA). (G-H) Absolute (G) and difference (H) in decoding accuracy of preferred versus anti-preferred directions, averaged across sound-modulated direction-selective neurons. No significant effect of sound on decoding accuracy was observed (n=12, p(vis)=2.1e-4, p(aud)=0.18, p(interact)=0.78, paired 2-way ANOVA).

723



724
725 **Figure 8 | Sound improves accuracy of population-based visual stimulus decoding** (A) Schematic illustrating the
726 decoding of the drifting grating direction using either an SVM or MLE decoder trained on neuronal population

activity. (B) Accuracy of SVM pairwise classification, average across all direction pairs, as the neuronal population size included in the decoder increases. Visual contrast 0.25 is on the left, and full visual contrast is on the right. (C) Accuracy of MLE decoding 1 of 12 drifting grating options, as the neuronal population size increases. Again, visual contrast 0.25 is on the left, and full visual contrast is on the right. (D) Accuracy of SVM pairwise classification of drifting grating directions on visual (left) and audiovisual (right) trials, contrast 0.25. (E) SVM decoding accuracy improved with sound when classifying orthogonal drifting grating orientations (n=10 randomizations, p(vis)=1.8e-61, p(aud)=1.9e-8, p(interact)=2.4e-4, 2-way ANOVA, post hoc Bonferroni-corrected paired t-test). (F) SVM decoding accuracy when classifying opposite drifting grating directions, demonstrating improved performance with sound (n=10 randomizations, p(vis)=1.1e-21, p(aud)=9.0e-9, p(interact)=0.0019, 2-way ANOVA, post hoc Bonferroni-corrected paired t-test). (G) Accuracy of MLE pairwise classification of drifting gratings on visual (left) and audiovisual (right) trials, contrast 0.25. (H) MLE decoding accuracy when classifying orthogonal drifting grating orientations improved with sound (n=10 randomizations, p(vis)=2.3e-66, p(aud)=0.61, p(interact)=9.6e-11, 2-way ANOVA, post hoc Bonferroni-corrected paired t-test). (I) MLE decoding accuracy when classifying opposite drifting grating directions, demonstrating less effect of sound on performance (n=10 randomizations, p(vis)=4.6e-26, p(aud)=0.51, p(interact)=4.1e-6, 2-way ANOVA, post hoc Bonferroni-corrected paired t-test). (J) Heat map of actual vs MLE-output directions under visual (left) and audiovisual (right) trials, contrast 0.25. MLE decoder could choose between all 12 drifting grating directions. (K) MLE decoder classification percentage, comparing estimated direction to actual direction. (L) Overall decoding accuracy of MLE decoder when choosing between all 12 drifting grating directions improved with sound (n=20 randomizations, p(vis)=2.2e-92, p(aud)=1.9e-5, p(interact)=2.7e-11, 2-way ANOVA, post hoc Bonferroni-corrected paired t-test).

indicate that sound improves neuronal encoding of the visual stimulus both in individual neurons and at a population level, especially at intermediate visual contrast levels.

Sound improves stimulus decoding when controlling for sound-induced movements

It is known that locomotion improves visual processing in V1 (Dardalat and Stryker, 2017). We next tested whether the sound-induced improvement in visual stimulus representation (Figure 8) was attributable to sound's effect on visual responses or indirectly via sound-induced movement. As we previously observed, sound and movement enhanced the onset and sustained portion of the visual response, respectively (Figure 6). We therefore hypothesized that the improvement on MLE decoding performance, based on the visual response onset, would be present even when accounting for sound-induced uninstructed movements. We tested this hypothesis by expanding on the GLM-based classification of neurons described in Figure 6. Using the same GLM generated for each neuron, we modified the movement variable and its corresponding pairwise predictors to the lowest observed value, and then used the GLM coefficients and the exponential nonlinearity to estimate each neuron's audiovisual response magnitude when regressing out the effect of motion (Figure 9A, Materials and Methods). We then input these estimated trial-wise neuronal responses into the same MLE-based decoder described above. Using this approach, we found that in individual orientation-selective neurons, controlling for the effect of motion on audiovisual trials had little effect on decoding accuracy across contrast levels (Figure 9B-C; p(vis)=7.7e-93, p(aud)=0.055, p(interact)=0.058, paired 2-way ANOVA, Table 1). However, regressing out both sound and motion from the audiovisual responses resulted in decoding accuracy that resembled that on visual trials (Figure 9B-C; p(vis)=8.1e-95, p(aud) = 0.55, p(interact)=0.24, paired 2-way ANOVA, Table 1). These results in individual neurons indicate that sound and not movement primarily drives the improvements in decoding accuracy in audiovisual trials. We found similar results when implementing this approach in the MLE-based population decoder. We again found that that decoding performance on audiovisual trials when regressing out motion was still significantly improved compared to that on visual trials (Figure 9D-E; p(vis)=1.4e-38, p(aud)=6.0e-8, p(interact)=0.0015, 2-way ANOVA; $p_{c=0}$=0.30, $p_{c=0.25}$=0.0012, $p_{c=0.5}$=0.0022, $p_{c=0.75}$=0.0044, $p_{c=1}$=0.35, Bonferroni-corrected paired t-test).

778 Furthermore, regression of both sound and movement from audiovisual trials resulted in
779 population decoding performance similar to that on visual trials (Figure 9D-E; p(vis)=2.5e-39,
780 p(aud)=0.48,

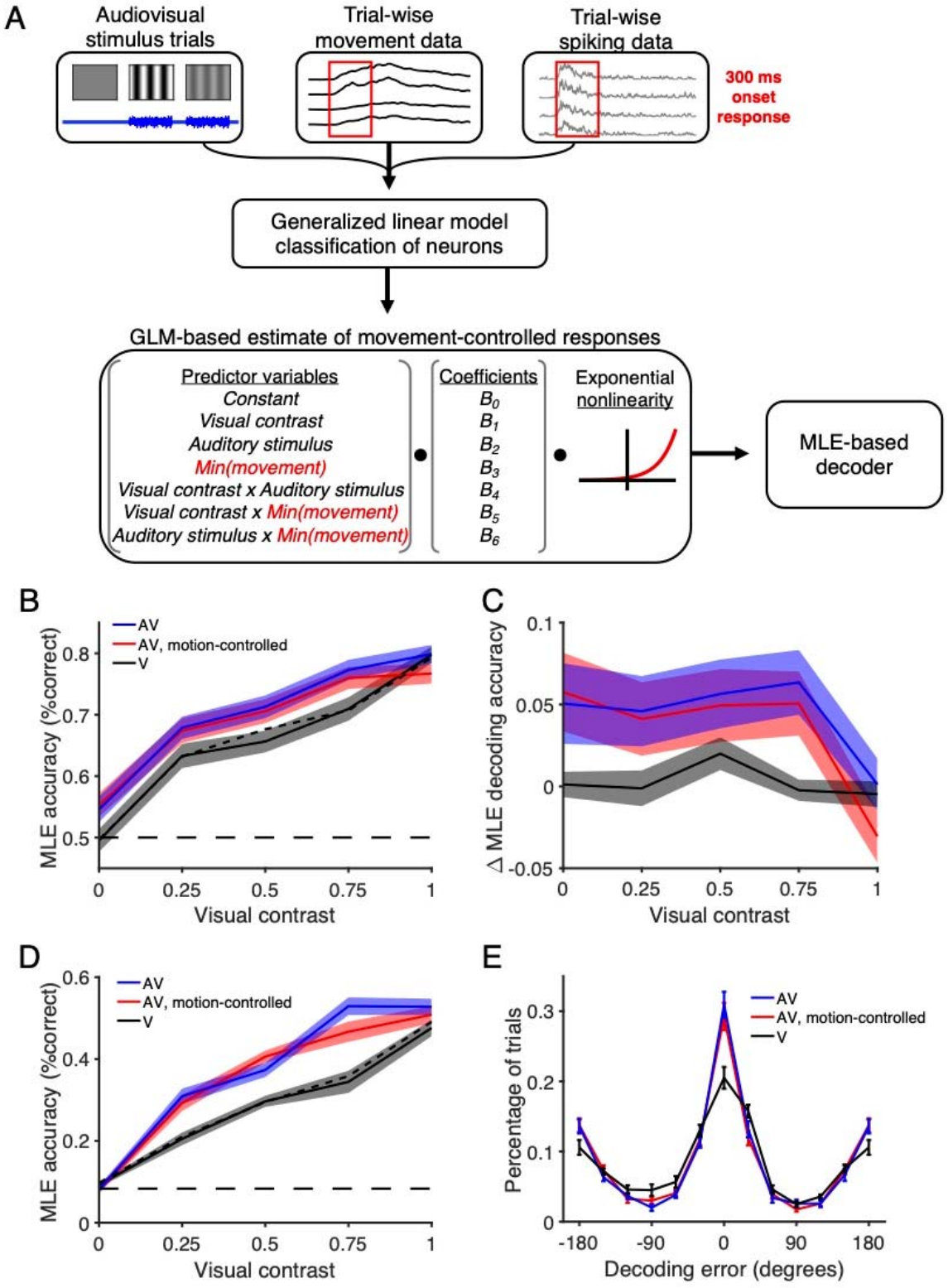


Figure 9 | Sound improved decoding performance when controlling for motion. (A) Diagram illustrating the use of a GLM to calculate each predictor variable's coefficient. These are then used when varying the predictor variables to estimate trial-wise neuronal responses, which are then into the MLE-based decoder. (B) Absolute accuracy of

decoding orientation among orientation-selective, sound/motion-modulated light-responsive neurons, comparing visual responses (black, solid) to audiovisual responses (blue) and audiovisual responses when regressing out motion (red). The finely dotted line represents audiovisual responses when controlling for the effects of both motion and sound. (C) Relative decoding accuracy compared to decoding on visual trials. Regressing out motion did not reduce performance compared to audiovisual trials (n=85 neurons, p(vis)=7.7e-93, p(aud)=0.055, p(interact)=0.058, paired 2-way ANOVA), whereas regressing out both motion and sound resulted in comparable performance to visual trials (n=85 neurons, p(vis)=8.1e-95, p(aud)=0.55, p(interact)=0.24, paired 2-way ANOVA). (D) Population decoding accuracy of population-based decoder on audiovisual trials (blue) is preserved even when controlling for motion (red) compared to decoding of visual trials (black; n=10 randomizations, p(vis) = 1.4e-38, p(aud)=6.0e-8, p(interact)=0.0015, 2-way ANOVA; $p_{c=0}$=0.30, $p_{c=0.25}$0.0012, $p_{c=0.5}$=0.0022, $p_{c=0.75}$=0.0044, $p_{c=1}$=0.35, Bonferroni-corrected paired t-test). The finely black dotted line represents decoding accuracy when regressing out both sound and motion. (E) MLE decoder classification percentage, comparing estimated direction to actual direction, contrast 0.25. Little difference is observed between audiovisual trials and audiovisual trials when controlling for motion, whereas both are more accurate than visual trials.

p(interact)=0.99, 2-way ANOVA). These results demonstrate that sound improves visual stimulus decoding on audiovisual trials at both a single neuron and population level. Moreover, this enhancement persists when controlling for sound-induced motion.

Discussion

Audiovisual integration is an essential aspect of sensory processing (Stein et al., 2020). In humans, audiovisual integration is used in everyday behaviors such as speech perception and object recognition (Fujisaki et al., 2014). In animal models, audiovisual integration improves the detection and discriminability of unisensory auditory and visual stimuli (Gleiss and Kayser, 2012; Meijer et al., 2018). However, the neuronal mechanisms underlying these behavioral improvements are still being investigated. Specifically, it remains unclear how sound-induced changes in spiking activity affect neuronal encoding of the visual stimulus. Furthermore, whether the reported audiovisual integration can more accurately be attributed to sound-induced movement is still being clarified.

The goal of the present study was to test whether sound drives not and improvement in encoding and decoding of sounds in awake subjects, and to test the hypothesis that sound improves neuronal encoding of visual stimuli in V1 independent of sound-induced movement. We performed extracellular recordings in V1 while presenting combinations of visual drifting gratings and auditory white noise and recording movement of awake mice. The drifting gratings were presented at a range of visual contrast levels to determine the threshold levels at which sound is most effective. As in previous studies, we found neurons in V1 whose spontaneous and visually evoked firing rates are modulated by sound (Figure 3). Notably, the effects we observed were stronger and more response-enhancing than in previous studies (80.1% of neurons were modulated by sound, with ~95% exhibiting sound-induced increases in firing rate). When accounting for movement in awake animal subjects, we found that the neurons' audiovisual responses actually represented a mixed effect of both sound- and movement-sensitivity (Figure 5), an effect in which sound primarily enhances the onset response whereas movement complementarily enhances the sustained response (Figure 6). We also found that sound-induced changes in response magnitude and consistency combined to improve the discriminability of drifting grating orientation and direction in individual neurons and at a population level (Figure 7). The improvements in neuronal encoding were most pronounced at low to intermediate visual contrast levels, a finding that supports the current understanding that

audiovisual integration is most beneficial for behavioral performance under ambiguous unisensory conditions (Gleiss and Kayser, 2012; Meijer et al., 2018; Stein et al., 2020), as found in human psychophysics (Lippert et al., 2007; Chen et al., 2011). Importantly, the improvement in neuronal encoding was based on firing at the onset of the visual response, indicating that the auditory signal itself is responsible for improvements in visual encoding and not attributable to uninstructed movements. This was directly demonstrated by the persistence of sound-induced improvements in stimulus decoding, even when controlling for the effect of motion (Figure 9).

Auditory and locomotive inputs distinctly shape visual responses

We present the novel finding that sound and movement have distinct and complementary effects on visual responses. Specifically, we found that sound primarily enhances the firing rate at the onset of the visual response, whereas motion enhances the firing rate during the sustained period of the visual response (Figure 6F-H). Prior audiovisual studies in the awake brain of mice used calcium imaging (Meijer et al, 2017; McClure and Polack 2019), a recording modality limited to the supragranular layers of V1. Additionally, the temporal resolution of calcium imaging limits the ability to detect the temporal differences in how sound and movement independently affect V1 responses in the awake brain. Therefore, our use of a depth electrode that spanned the cortical layers for electrophysiology in the awake brain enabled robust characterization of distinct temporal effects of sound and movement on visual responses.

Our initial classification of sound-modulated neurons and the subsequent decoding analyses were based on firing rates during the onset period. Therefore, despite robust differences in movement during visual and audiovisual trials, motion was unable to account for the sound-induced changes in neuronal responses that resulted in improved neuronal encoding (Figure 9). The distinct effects that sound and locomotion have on visual responses also adds nuance to our understanding of how motion affects visual processing, as other groups have predominantly used responses averaged across the duration of the stimulus presentation in categorizing motion responsive neurons in V1 (Neil and Stryker, 2010; Dardalat and Stryker, 2017). Our findings indicate that the timing of cross-sensory interactions is an important factor in the classification and quantification of multisensory effects.

We also observed that motion decreases the magnitude of the enhancing effect that sound has on the onset of the visual response (Figure 5E, 6H). This finding suggests a degree of suppressive effect that motion has on this audiovisual interaction. A potential mechanism for this result may relate to the circuits underlying audiovisual integration in V1. Other groups have shown using retrograde tracing, optogenetics and pharmacology that the AC projects directly to V1 and is responsible for the auditory signal in this region (Falchier et al., 2002; Ibrahim et al., 2016; Deneux et al., 2019). It is currently understood that unlike in V1, in other primary sensory cortical areas including the AC movement suppresses sensory evoked activity (Nelson et al., 2013; Schneider and Mooney, 2018; Bigelow et al., 2019). Therefore, one explanation for this observation is that despite motion enhancing the visual response magnitude in the absence of sound, the suppressive effect that motion has on sound-evoked responses in the AC leads to weaker AC enhancement of visual activity on trials in which the mice move. A detailed experimental approach using optogenetics or pharmacology would be required to test this hypothesis of a tripartite interaction and would also reveal the potential contribution of other auditory regions.

Enhanced response magnitude and consistency combine to improve neuronal encoding

Signal detection theory indicates that improved encoding can be mediated both by enhanced signal magnitude as well as reduced levels of noise (von Trapp et al., 2016). When using purely magnitude-based metrics of discriminability, OSI and DSI, we found a small reduction from the

visual to audiovisual conditions (Figure 3J,K). However, we also observed that sound reduced the CV of visual responses (Figure 4), a measure of the trial-to-trial variability in response. When we measured the *d'* sensitivity index of neuronal responses, a measure that factors in both the mean response magnitude and trial-to-trial variability, we found that sound improved the discriminability of drifting grating orientation and direction (Figure 7A,B). These findings indicate that the improved discriminability of visual responses in individual neurons was mediated not only by changes in response magnitude but also by the associated improvement in response consistency between trials. Prior studies using patch-clamp approaches showed that V1 neurons in anesthetized animals improve visual encoding by sharpening their tuning profiles (Ibrahim et al., 2016), a magnitude-based coding scheme. The difference between these findings and those reported in the current study potentially represent different coding schemes present in anesthetized and awake brains. It is therefore important to consider response variability in awake brains in addition to magnitude-based metrics when quantifying tuning and discriminability in neurons (Churchland et al., 2011; Mazurek et al., 2014).

Prior studies using calcium imaging found equivocal results when investigating whether sound-induced changes in visual responses led to improved population encoding of the visual stimulus (Meijer et al., 2017). The improved discriminability of grating orientation and direction by individual neurons supports our finding that the presence of sound enhances population encoding of the visual stimulus. Again, one explanation for this difference may be the recording modality and analysis parameters. We performed electrophysiological recordings of spiking activity and limited our quantification to the onset of the stimulus (0-300 ms), the time window in which there was the greatest change in firing rate across neurons. Our focus on the onset response was based on our initial finding that mutual information between the neuronal responses and visual stimuli was highest during this onset period, a finding supported by previous studies (Figure 2; Dardalat and Stryker, 2017). Calcium imaging, however, may lack the temporal resolution required to detect the trial-by-trial differences in spiking activity associated with improved neuronal discriminability during this timeframe. Extracellular electrophysiology also allowed us to take advantage of large numbers of neurons from a range of cortical depths in awake animals to include in the population analysis, as opposed to patch-clamp approaches with a limited number of neurons (Ibrahim et al., 2016). Finally, presenting a wide range of visual contrast levels allowed us to demonstrate that sound improves neuronal encoding at low to intermediate contrasts, above which further improvement is difficult to demonstrate due to already reliable encoding in the absence of sound. Altogether, these differences in methodology allowed us to more directly demonstrate that V1 neuronal encoding of visual stimuli is improved by sound in the awake brain, as well as disentangle the contributions of sound and sound-induced motion in that process.

Stimulus parameters relevant to audiovisual integration

Sensory neurons are often tuned to specific features of unisensory auditory and visual stimuli, and these features are relevant to cross-sensory integration of the signals. In the current study we paired the visual drifting gratings with a static burst of auditory white noise as a basic well-controlled stimulus. Previous studies found that temporally congruent audiovisual stimuli, e.g. amplitude-modulated sounds accompanying visual drifting gratings, evoke larger changes in response than temporally incongruent stimuli in the mouse visual cortex (Meijer et al., 2017; Atilgan et al., 2018), and therefore using such stimuli would potentially result in even stronger effects than we observed. Auditory pure tones can also induce changes in V1 visual responses (McClure and Polack, 2019). However, in other brain regions such as the inferior colliculus, audiovisual integration is highly dependent on spatial congruency between the unimodal inputs (Bergan and Knudsen, 2009). Additional studies are needed to explore the full range of auditory stimulus parameters relevant to visual responses in V1.

Our results show that spatially congruent, static white noise is sufficient to improve V1 neuronal response magnitude and latency to light-evoked responses. These results likely extend to natural and ethologically relevant stimuli as well. Indeed, rhesus macaque monkeys demonstrate psychometric and neurometric improvements in tasks such as conspecific vocalization detection and object recall (Hwang and Romanski, 2015; Bigelow and Poremba, 2016; Breman et al., 2017). Humans are also capable of perceptually integrating audiovisual stimuli ranging from paired visual drifting gratings and auditory white noise (Lippert et al., 2007; Chen et al., 2007), to the McGurk effect and virtual reality simulated driving (McGurk and MacDonald, 1976; Marucci et al., 2021). We therefore posit that the audiovisual integration of basic sensory stimuli in early sensory areas may form the foundation for functional integration by higher cortical areas and ultimately behavioral improvements.

Neuronal correlates of multisensory behavior

Our findings of multisensory improvements in neuronal performance are supported by numerous published behavioral studies in humans and various model organisms (Gleiss and Kayser, 2012; Meijer et al., 2018; Stein et al., 2020). Training mice to detect or discriminate audiovisual stimuli allows the generation of psychometric performance curves in the presence and absence of sound. We would hypothesize that the intermediate visual contrast levels in which we see improvements in neural encoding would align with behavioral detection threshold levels. One could also correlate the trial-by-trial neural decoding of the visual stimulus with the behavioral response on a stimulus discriminability task, an analysis that could provide information about the proximity of the V1 responses to the behavioral perception and decision. Additionally, a behavioral task could allow the comparison of neural responses between passive and active observing, helping to reveal the role of attention on how informative or distracting one stimulus is about the other.

Multisensory integration in other systems

It is useful to contextualize audiovisual integration by considering multisensory integration that occurs in other primary sensory cortical areas. The auditory cortex contains visually responsive neurons and is capable of binding temporally congruent auditory and visual stimulus features in order to improve deviance detection within the auditory stimulus (Atilgan et al., 2018; Morrill and Hasenstaub, 2018). Additionally, in female mice, pup odors reshape AC neuronal responses to various auditory stimuli and drive pup retrieval behavior (Cohen et al., 2011; Marlin et al., 2015), demonstrating integration of auditory and olfactory signals. However, whether these forms of multisensory integration rest on similar coding principles of improved SNR observed in the current V1 study is unknown. Investigation into this relationship between the sensory cortical areas will help clarify the neuronal codes that support multisensory integration, and the similarities and differences across sensory domains.

Acknowledgements

The authors thank Gabrielle Samulewicz for assistance with experiments and members of the Geffen laboratory for helpful discussions, as well as Dr. Jay Gottfried and Dr. Yale Cohen at the University of Pennsylvania. This work was supported by funding by the National Institute on Deafness and Other Communication Disorders at the National Institute of Health grants 5T32DC016903 to AMW, F31DC016524 to CFA, and R01DC015527, R01DC014479, and R01NS113241 to MNG.

References

Atilgan, H., Town, S. M., Wood, K. C., Jones, G. P., Maddox, R. K., Lee, A., & Bizley, J. K. (2018). Integration of Visual Information in Auditory Cortex Promotes Auditory Scene Analysis through Multisensory Binding. *Neuron*, *97*(3), 640–655.e4. https://doi.org/10.1016/j.neuron.2017.12.034

Bergan, J.F., Knudsen, E.I. (2009). Visual modulation of auditory responses in the owl inferior colliculus. *J Neurophysiol*, 101(6):2924-33. Doi: 10.1152/jn.91313.2008

Berens, P., Ecker, A.S., Cotton, R.J., Ma, W.J., Bethge, M., Tolias, A.S. (2012) A fast and simple population code for orientation in primate V1. *J Neurosci*, 32(31): 10618-26. doi: 10.1523/JNEUROSCI.1335-12.2012

Bigelow, J., Poremba, A. (2016) Audiovisual integration improves monkeys' short-term memory. *Anim Cogn*, 19(4): 799-811. Doi: 10.1007/s10071-016-0979-0

Bigelow, J., Morrill, R.J., Dekloe, J., Hasenstaub, A.R. (2019) Movement and VIP interneuron activation differentially modulate encoding in mouse auditory cortex. *eNeuro*, 6(5); ENEURO.0164-19.2019. doi: 10.1523/ENEURO.0164-19.2019

Bimbard, C., Sit, T.P.H., Lebedeva, A., Harris, K.D., Carandini, M. (2021) Behavioral origin of sound-evoked activity in visual cortex. BioRXiv. Doi: https://doi.org/10.1101/2021.07.01.450721

Borst, A., Theunissen, F.E. (1999) Information theory and neural coding. *Nat Neurosci*, 2(11):947-57. Doi: 10.1038/14731

Breman, P., Massoudi, R., Wanrooij, M.M.V., Van Opstal, A.J. (2017) Audio-visual integration in a redundant target paradigm: A comparison between rhesus macaque and man. *Front Syst Neurosci*, 11:89. Doi: 10.3389/fnsys.2017.00089

Churchland, A.K., Kiani, R., Chaudhuri, R., Wang, X., Pouget, A., Shadlen, M.N. (2011) Variance as a signature of neural computations during decision making. *Neuron*, 69(4):818-31. Doi: 10.1016/j.neuron.2010.12.037

Cohen, L., Rothschild, G., Mizrahi, A. (2011). Multisensory integration of natural odors and sounds in the auditory cortex. *Neuron*, 72(2): 357-69. Doi: 10.1016/j.neuron.2011.08.019

Colonius, H., & Diederich, A. (2017). Measuring multisensory integration: from reaction times to spike counts. *Scientific reports*, *7*(1), 3023. https://doi.org/10.1038/s41598-017-03219-5

Crosse, M. J., Butler, J. S., & Lalor, E. C. (2015). Congruent Visual Speech Enhances Cortical Entrainment to Continuous Auditory Speech in Noise-Free Conditions. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *35*(42), 14195–14204. https://doi.org/10.1523/JNEUROSCI.1829-15.2015

Dardalat M.C., Stryker M.P. (2017) Locomotion enhances neural encoding of visual stimuli in mouse V1. *J Neurosci*, 37(14):3764-75. Doi: 10.1523/JNEUROSCI.2728-16.2017

1026 Deneux T., Harrell E.R., Kempf A., Ceballo S., Filipchuk A., Bathellier B. (2019) Context-
1027 dependent signaling of coincident auditory and visual events in primary visual cortex. eLife
1028 8: e44006. doi: 10.7554/eLife.44006

1029 Denison, R. N., Driver, J., & Ruff, C. C. (2013). Temporal structure and complexity affect
1030 audio-visual correspondence detection. *Frontiers in psychology*, *3*, 619.
1031 https://doi.org/10.3389/fpsyg.2012.00619

1032 Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement:
1033 effects of stimulus onset and intensity on reaction time. *Perception & psychophysics*, *66*(8),
1034 1388–1404. https://doi.org/10.3758/bf03195006

1035 Fahey, P.G., Muhammad, T., Smoth, C., Froudarakis, E., Cobos, E., Fu, J., Walker, E.Y.,
1036 Yatsenko, D., Sinz, F.H., Reimer, J., Tolias, A.S. (2019). A global map of orientation tuning
1037 in mouse visual cortex. BioRXiv.

1038 Falchier, A., Clavagnier, S., Barone, P., Kennedy, H. (2002) Anatomical evidence of
1039 multimodal integration in primate striate cortex. *J Neurosci*, 22(13): 5749-59. Doi:
1040 10.1523/JNEUROSCI.22-13-05749.2002.

1041 Fujisaki, W., Goda, N., Motoyoshi, I., Komatsu, H., Nishida, S. (2014) Audiovisual integration
1042 in the human perception of materials. *J Vis*, 14(4):12. doi: 10.1167/14.4.12

1043 Gingras, G., Rowland, B. A., & Stein, B. E. (2009). The differing impact of multisensory and
1044 unisensory integration on behavior. *The Journal of neuroscience : the official journal of the*
1045 *Society for Neuroscience*, *29*(15), 4897–4902. https://doi.org/10.1523/JNEUROSCI.4120-
1046 08.2009

1047 Gleiss, S., & Kayser, C. (2012). Audio-visual detection benefits in the rat. *PloS one*, *7*(9),
1048 e45677. https://doi.org/10.1371/journal.pone.0045677

1049 Gur, M., Beylin, A., Snodderly, D.M. (1997) Response variability of neurons in primary visual
1050 cortex (V1) of alert monkeys. *J Neurosci*, 17(8):2914-20. doi: 10.1523/JNEUROSCI.17-08-
1051 02914.1997

1052 Hammond-Kenny, A., Bajo, V. M., King, A. J., & Nodal, F. R. (2017). Behavioural benefits of
1053 multisensory processing in ferrets. *The European journal of neuroscience*, *45*(2), 278–289.
1054 https://doi.org/10.1111/ejn.13440

1055 Hirokawa, J., Sadakane, O., Sakata, S., Bosch, M., Sakurai, Y., & Yamamori, T. (2011).
1056 Multisensory information facilitates reaction speed by enlarging activity difference between
1057 superior colliculus hemispheres in rats. *PloS one*, *6*(9), e25283.
1058 https://doi.org/10.1371/journal.pone.0025283

1059 Hwang, J., Romanski, L. (2015) Prefrontal neuronal responses during audiovisual
1060 mnemonic processing. *J Neurosci*, 35(3): 960-71. Doi: 10.1523/JNEUROSCI.1328-14.2015

1061 Ibrahim, L. A., Mesik, L., Ji, X. Y., Fang, Q., Li, H. F., Li, Y. T., Zingg, B., Zhang, L. I., & Tao,
1062 H. W. (2016). Cross-Modality Sharpening of Visual Cortical Processing through Layer-1-
1063 Mediated Inhibition and Disinhibition. *Neuron*, *89*(5), 1031–1045.
1064 https://doi.org/10.1016/j.neuron.2016.01.027

Knöpfel, T., Sweeney, Y., Radulescu, C. I., Zabouri, N., Doostdar, N., Clopath, C., & Barnes, S. J. (2019). Audio-visual experience strengthens multisensory assemblies in adult mouse visual cortex. *Nature communications*, *10*(1), 5684. https://doi.org/10.1038/s41467-019-13607-2

Maddox, R. K., Atilgan, H., Bizley, J. K., & Lee, A. K. (2015). Auditory selective attention is enhanced by a task-irrelevant temporally coherent visual stimulus in human listeners. *eLife*, *4*, e04995. https://doi.org/10.7554/eLife.04995

Marlin, B.J., Mitre, M., D'amour, J.A., Chao, M.V., Froemke, R.C. (2015). Oxytocin enables maternal behavior by balancing cortical inhibition. *Nature*, 520(7528): 499-504. doi: 10.1038/nature14402

Marucci, M., Flumeri, G.D., Borghini, G., Sciaraffa, N., Scandola, M., Pavone, E.F., Babiloni, F., Betti, V., Arco, P. (2021) The impact of multisensory integration and perceptual load in virtual reality settings on performance, workload and presence. *Sci Rep*, 11(1);4831. Doi: 10.1038/s41598-021-84196-8.

Mazurek, M., Kager, M., Van Hooser, SD. (2014) Robust quantification of orientation selectivity and direction selectivity. *Front in Neural Circuits*, 8: 92. Doi: https://doi.org/10.3389/fncir.2014.00092

McClure, J. P., Jr, & Polack, P. O. (2019). Pure tones modulate the representation of orientation and direction in the primary visual cortex. *Journal of neurophysiology*, *121*(6), 2202–2214. https://doi.org/10.1152/jn.00069.2019

Mcgurk, H., and Macdonald, J. (1976). Hearing lips and seeing voices. Nature 264, 746–748. doi: 10.1038/264746a0

Meijer, G. T., Montijn, J. S., Pennartz, C., & Lansink, C. S. (2017). Audiovisual Modulation in Mouse Primary Visual Cortex Depends on Cross-Modal Stimulus Configuration and Congruency. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *37*(36), 8783–8796. https://doi.org/10.1523/JNEUROSCI.0468-17.2017

Meijer, G. T., Pie, J. L., Dolman, T. L., Pennartz, C., & Lansink, C. S. (2018). Audiovisual Integration Enhances Stimulus Detection Performance in Mice. *Frontiers in behavioral neuroscience*, *12*, 231. https://doi.org/10.3389/fnbeh.2018.00231

Meijer, G. T., Mertens, P., Pennartz, C., Olcese, U., & Lansink, C. S. (2019). The circuit architecture of cortical multisensory processing: Distinct functions jointly operating within a common anatomical network. *Progress in neurobiology*, *174*, 1–15. https://doi.org/10.1016/j.pneurobio.2019.01.004

Métin, C., Godement, P., & Imbert, M. (1988). The primary visual cortex in the mouse: receptive field properties and functional organization. *Experimental brain research*, *69*(3), 594–612. https://doi.org/10.1007/BF00247312

Montijn J.S., Vinck M., Pennartz C.M.A. (2014) Population coding in mouse visual cortex: response reliability and dissociability of stimulus tuning and noise correlation. *Front Comput Neurosci* 8:58. Doi: 10.3389/fncom.2014.00058

Morrill, R. J., & Hasenstaub, A. R. (2018). Visual Information Present in Infragranular Layers of Mouse Auditory Cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *38*(11), 2854–2862. https://doi.org/10.1523/JNEUROSCI.3102-17.2018

Musall, S., Kaufman, M.T., Juavinett, A.L., Gluf, S., Churchland, A.K. (2019) Single-trial neural dynamics are dominated by richly varied movements. *Nat Neurosci*, 22(10):1677-1686. Doi: 10.1038/s41593-093-019-0502-4.

Nelson A., Schneider D.M., Takatoh J., Sakurai K., Wang F., Mooney R. (2013) A circuit for motor cortical modulation of auditory cortical activity. *J Neurosci*, 33(36):14342-53. Doi: 10.1523/JNEUROSCI.2275-13.2013.

Niell, C.M., Stryker M.P. (2008) Highly receptive fields in mouse visual cortex. *J Neurosci*, 28(30): 7520-36. Doi: https://doi.org/10.1523/JNEUROSCI.0623-08.2008

Niell C.M., Stryker M.P. (2010) Modulation of visual responses by behavioral state in mouse visual cortex. *Neuron*, 65(4):472-9. Doi: 10.1016/j.neuron.2010.01.033.

Quak, M., London, R. E., & Talsma, D. (2015). A multisensory perspective of working memory. *Frontiers in human neuroscience*, *9*, 197. https://doi.org/10.3389/fnhum.2015.00197

Pachitariu M., Steinmetz N., Kadir S., Carandini M., Harris K.D. (2016) Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels. *BioRχiv*. Doi: https://doi.org/10.1101/061481

Rochefort, N. L., Narushima, M., Grienberger, C., Marandi, N., Hill, D. N., & Konnerth, A. (2011). Development of direction selectivity in mouse cortical neurons. *Neuron*, *71*(3), 425–432. https://doi.org/10.1016/j.neuron.2011.06.013

Schneider D.M., Mooney R. (2018) How movement modulates hearing. *Annu Rev Neurosci*, 41:553-72. Doi: 10.1146/annurev-neuro-072116-031215.

Shams L, Kamitani Y, Shimojo S (2002) Visual illusion induced by sound. *Brain Res Cogn Brain Res* 14(1):147-52. Doi: 10.1016/s0926-6410(02)00069-1

Stanislaw, H., Todorov, N. (1999) Calculation of signal detection theory measures. *Behav Res Methods Instrum Comput*. 31(1):137-49. Doi: 10.3758/bf03207704

Stein, B. E., Stanford, T. R., & Rowland, B. A. (2020). Multisensory Integration and the Society for Neuroscience: Then and Now. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *40*(1), 3–11. https://doi.org/10.1523/JNEUROSCI.0737-19.2019

Tye-Murray, N., Spehar, B., Myerson, J., Hale, S., & Sommers, M. (2016). Lipreading and audiovisual speech recognition across the adult lifespan: Implications for audiovisual integration. *Psychology and aging*, *31*(4), 380–389. https://doi.org/10.1037/pag0000094

von Trapp, G., Buran, B. N., Sen, K., Semple, M. N., & Sanes, D. H. (2016). A Decline in Response Variability Improves Neural Signal Detection during Auditory Task

Performance. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *36*(43), 11097–11106. https://doi.org/10.1523/JNEUROSCI.1302-16.2016

Wang, Y., Celebrini, S., Trotter, Y., Barone, P. (2008) Visuo-auditory interactions in the primary visual cortex of the behaving monkey: electrophysiological evidence. *BMC Neurosci*, 9:79. Doi: 10.1186/1471-2202-9-79

Zhao X., Chen H., Liu X., Cang J. (2013) Orientation-selective responses in the mouse lateral geniculate nucleus. *J Neurosci* 33(31):12751-763. Doi: 10.1523/JNEUROSCI.0095-13.2013

Table 1: Statistical comparisons

Comparison	Fig	Test	Test statistic	N	df	p-value	Post hoc test	Post hoc α	Post hoc comparison	Post hoc p-value
Mean firing rate, V vs AV	3C	Paired 2-way ANOVA	F(vis)=340 F(aud)=506 F(interact)=75	565 neurons	vis=4 aud=1 interact = 4	p(vis) = 1.2e-100 p(aud) = 1.6e-88 p(interact) = 5.7e-4	Bonferroni-corrected paired t-test	0.01	Contrast 0, V vs AV	2.1e-50
									Contrast 0.25, V vs AV	2.6e-62
									Contrast 0.5, V vs AV	5.7e-75
									Contrast 0.75, V vs AV	1.1e-81
									Contrast 1, V vs AV	2.0e-81
Linearity ratio, V vs AV	3E	Kruskal-Wallis test	Chi-sq = 61	555 neurons	4	p = 1.6e-12	Bonferroni-corrected Wilcoxon signed rank test	0.0125	Contrast 0 vs 0.25	0.053
									Contrast 0 vs 0.5	0.0040
									Contrast 0 vs 0.75	4.6e-8
									Contrast 0 vs 1	2.1e-5
Orientation selectivity index, V vs AV	3J	Paired t-test	t-stat = 3.2	78 neurons	77	p = 0.0018				
Direction selectivity index, V vs AV	3K	Paired t-test	t-stat = 2.7	12 neurons	11	p = 0.0206				
Onset response latency, V vs AV	4B	Paired 2-way ANOVA	F(vis)=5.7 F(aud)=64 F(interact)=2.7	517 neurons	vis=3 aud=1 interact=3	p(vis)=6.9e-4 p(aud)=6.8e-18 p(interact)=0.045	Bonferroni-corrected paired t-test	0.01	Contrast 0.25, V vs AV	2.3e-4
									Contrast 0.5, V vs AV	7.1e-12
									Contrast 0.75, V vs AV	4.6e-5
									Contrast 1, V vs AV	9.9e-4
Onset response slope, V vs AV	4D	Paired 2-way ANOVA	F(vis)=70 F(aud)=66 F(interact)=2.8	563 neurons	vis=3 aud=1 interact=3	p(vis)=3.5e-121 p(aud) = 2.7e-15 p(interact) = 0.038	Bonferroni-corrected paired t-test	0.01	Contrast 0.25, V vs AV	1.4e-4
									Contrast 0.5, V vs AV	8.9e-13
									Contrast 0.75, V vs AV	3.6e-12
									Contrast 1, V vs AV	5.5e-8
Onset response duration, V vs AV	4F	Paired 2-way ANOVA	F(vis)=17 F(aud)=129 F(interact)=1.4	367 neurons	vis=3 aud=1 Interact=3	p(vis)=1.3e-10 p(aud) = 8.7e-98 p(interact) = 0.23				
Response coefficient of variation, V vs AV	4H	Paired 2-way ANOVA	F(vis)=1.3 F(aud)=834 F(interact)=1.0	564 neurons	vis=4 aud=1 Interact=4	p(vis) = 0.28 p(aud) = 4.2e-103 p(interact) = 0.38				
Sound induced movement	5B	Paired t-test	t-stat = -7.2	9 recording sessions	8	p = 9.1e-5				
Firing rate across movement range, V vs AV	5F	Unbal-anced 2-way ANOVA	F(motion)=6.9 F(sound)=55 F(interact)=18	Vari-able trial count	mot=2 aud=1 Intera	p(motion) = 0.001 p(sound) = 1.4e-13 p(interact) = 1.8e-8	Bonferroni corrected two-sample t-test	0.016	Stationary, V vs AV	1.5e-14
									Low motion, V vs AV	7.1e-10
									High motion, V vs AV	0.60

					ct=2					
PSTH, light vs light/sound	6F	Paired t-test	1391 unique t-stats	295 neurons	294	1391 unique p-values, α= 0.05/1391= 3.6e-5				
PSTH, light vs light/motion	6G	Paired t-test	1391 unique t-stats	295 neurons	294	1391 unique p-values, α= 0.05/1391= 3.6e-5				
PSTH, light/sound vs light/sound/motion	6H	Paired t-test	1391 unique t-stats	295 neurons	294	1391 unique p-values, α= 0.05/1391= 3.6e-5				
Orientation decoding accuracy, individual neurons, V vs AV	7E	Paired 2-way ANOVA	F(vis)=67 F(aud)=12 F(interact)=0.54	78 neurons	vis=4 aud=1 interact=4	p(vis)=4.8e-112 p(aud)=7.8e-4 p(interact) = 0.71				
Direction decoding accuracy, individual neurons, V vs AV	7G	Paired 2-way ANOVA	F(vis)=6.9 F(aud)=2.0 F(interact)=0.43	12 neurons	vis=4 aud=1 interact=4	p(vis)=2.1e-4 p(aud)=0.18 p(interact)=0.78				
Orientation decoding accuracy, SVM, population, V vs AV	8E	2-way ANOVA	F(vis)=526 F(aud)=38 F(interact)=6	10 repeats	vis=4 aud=1 interact=4	p(vis) = 1.8e-61 p(aud) = 1.9e-8 p(interact) = 2.4e-4	Bonferroni-corrected paired t-test	0.01	Contrast 0, V vs AV Contrast 0.25, V vs AV Contrast 0.5, V vs AV Contrast 0.75, V vs AV Contrast 1, V vs AV	0.12 0.0016 0.0014 0.0023 1
Direction decoding accuracy, SVM, population, V vs AV	8F	2-way ANOVA	F(vis)=48 F(aud)=40 F(interact)=4.6	10 repeats	vis=4 aud=1 interact=4	p(vis) = 1.1e-21 p(aud) = 9.0e-9 p(interact) = 0.0019	Bonferroni-corrected paired t-test	0.01	Contrast 0, V vs AV Contrast 0.25, V vs AV Contrast 0.5, V vs AV Contrast 0.75, V vs AV Contrast 1, V vs AV	0.55 5.3e-5 0.0036 0.17 0.0036
Orientation decoding accuracy, MLE, population, V vs AV	8H	2-way ANOVA	F(vis)=682 F(aud)=0.27 F(interact)=18	10 repeats	vis=4 aud=1 interact=4	p(vis)=2.3e-66 p(aud)=0.61 p(interact) =9.6e-11	Bonferroni-corrected paired t-test	0.01	Contrast 0, V vs AV Contrast 0.25, V vs AV Contrast 0.5, V vs AV Contrast 0.75, V vs AV Contrast 1, V vs AV	5.8e-4 1.8e-4 0.30 0.53 0.15
Direction decoding accuracy, MLE, population, V vs AV	8I	2-way ANOVA	F(vis)=67 F(aud)=0.43 F(interact)=8.9	10 repeats	vis=4 aud=1 interact=4	p(vis)=4.6e-26 p(aud)=0.51 p(interact) =4.1e-6	Bonferroni-corrected paired t-test	0.01	Contrast 0, V vs AV Contrast 0.25, V vs AV Contrast 0.5, V vs AV Contrast 0.75, V vs AV Contrast 1, V vs AV	0.037 6.4e-6 0.036 0.16 0.014
Overall decoding accuracy, MLE, population, V vs AV	8L	2-way ANOVA	F(vis)=411 F(aud)=19 F(interact)=16	20 repeats	vis=4 aud=1 interact=4	p(vis)=2.2e-92 p(aud)=1.9e-5 p(interact)=2.7e-11	Bonferroni - corrected paired t-test	0.01	Contrast 0, V vs AV Contrast 0.25, V vs AV Contrast 0.5, V vs AV Contrast 0.75, V vs AV Contrast 1, V vs AV	0.012 1.4e-10 0.48 0.0013 0.50
Orientation decoding accuracy,	9B	Paired 2-way	F(vis) = 74 F(aud) = 19	85 neurons	vis=4 aud=	p(vis) =0 p(aud)=3.5e-5				

individual neurons, V vs AV		ANOVA	F(interact) = 1.5		1 interact=4	p(interact)=0.21				
Orientation decoding accuracy, individual neurons, V vs motion-corrected AV	9B	Paired 2-way ANOVA	F(vis) = 64 F(aud) = 13 F(interact) = 3	85 neurons	vis=4 aud=1 interact=4	p(vis) =0 p(aud)=5.9e-4 p(interact)=0.019	Bonferroni-corrected paired t-test	0.01	Contrast 0, V vs AV Contrast 0.25, V vs AV Contrast 0.5, V vs AV Contrast 0.75, V vs AV Contrast 1, V vs AV	0.019 0.071 0.029 0.011 0.0602
Orientation decoding accuracy, individual neurons, AV vs motion-corrected AV	9B	Paired 2-way ANOVA	F(vis) = 34 F(aud) = 3.8 F(interact) = 2.4	85 neurons	vis=4 aud=1 interact=4	p(vis) = 7.7e-93 p(aud) = 0.055 p(interact) = 0.058				
Orientation decoding accuracy, individual neurons, V vs motion/sound-corrected AV	9B	Paired 2-way ANOVA	F(vis) = 56 F(aud) = 0.36 F(interact) = 1.4	85 neurons	vis=4 aud=1 interact=4	p(vis)=8.1e-95 p(aud)=0.55 p(interact)=0.24				
Population decoding accuracy, V vs AV	9D	2-way ANOVA	F(vis) = 166 F(aud) = 52 F(interact) = 8.2	10 repeats	vis=4 aud=1 interact=4	p(vis)=1.1e-40 p(aud)=1.6e-10 p(interact)=1.1e-5	Bonferroni-corrected paired t-test	0.01	Contrast 0, V vs AV Contrast 0.25, V vs AV Contrast 0.5, V vs AV Contrast 0.75, V vs AV Contrast 1, V vs AV	0.34 2.2e-5 0.0019 8.7e-6 0.013
Population decoding accuracy, V vs motion-corrected AV	9D	2-way ANOVA	F(vis) = 147 F(aud) = 35 F(interact) = 4.8	10 repeats	vis=4 aud=1 interact=4	p(vis)=1.4e-38 p(aud)=6.0e-8 p(interact)=0.0015	Bonferroni-corrected paired t-test	0.01	Contrast 0, V vs AV Contrast 0.25, V vs AV Contrast 0.5, V vs AV Contrast 0.75, V vs AV Contrast 1, V vs AV	0.30 0.0012 0.0022 0.0044 0.35
Population decoding accuracy, V vs motion/sound-corrected AV	9D	2-way ANOVA	F(vis) = 154 F(aud) = 0.50 F(interact) = 0.088	10 repeats	vis=4 aud=1 interact=4	p(vis)=2.5e-39 p(aud) = 0.48 p(interact) = 0.99				