Systems/Circuits

Sound Improves Neuronal Encoding of Visual Stimuli in Mouse Primary Visual Cortex

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

[1]Department of Otorhinolaryngology, University of Pennsylvania, Philadelphia, Pennsylvania, 19104, [2]Department of Neuroscience, University of Pennsylvania, Philadelphia, Pennsylvania, 19104, [3]Department of Neurology, University of Pennsylvania, Philadelphia, Pennsylvania, 19104, and [4]Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania 19104

In everyday life, we integrate visual and auditory information in routine tasks such as navigation and communication. While concurrent sound can improve visual perception, the neuronal correlates of audiovisual integration are not fully understood. Specifically, it remains unclear whether neuronal firing patters in the primary visual cortex (V1) of awake animals demonstrate similar sound-induced improvement in visual discriminability. Furthermore, presentation of sound is associated with movement in the subjects, but little is understood about whether and how sound-associated movement affects audiovisual integration in V1. Here, 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 mice while recording mouse movement and V1 neuronal activity. Sound modulated activity of 80% of light-responsive neurons, with 95% of neurons increasing activity when the auditory stimulus was present. A generalized linear model (GLM) 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, an effect that persisted even when controlling for movement. These results demonstrate that 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 at a neuronal level in the awake brain.

Key words: **auditory cortex; electrophysiology; multimodal; visual cortex; visual processing**

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.

Received Dec. 2, 2021; revised Feb. 14, 2023; accepted Feb. 23, 2023.

Author contributions: A.M.W. and M.N.G. designed research; A.M.W., C.F.A., and M.N.G. performed research; A.M.W. and C.F.A. contributed unpublished reagents/analytic tools; A.M.W. and C.F.A. analyzed data; A.M.W. and M.N.G. wrote the first draft of the paper; A.M.W., C.F.A., and M.N.G. edited the paper; A.M.W. and M.N.G. wrote the paper.

This work was supported by National Institute of Health National Institute on Deafness and Other Communication Disorders Grants 5T32DC016903 (to A.M.W. and C.F.A.), F31DC016524 (to C.F.A.), and R01DC015527, R01DC014479, and R01NS113241 (to M.N.G.). We 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.

The authors declare no competing financial interests.

Correspondence should be addressed to Maria N. Geffen at mgeffen@pennmedicine.upenn.edu.

https://doi.org/10.1523/JNEUROSCI.2444-21.2023

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; Bremen 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 with asynchronous stimuli (Lippert et al., 2007). Additionally, multisensory integration is often optimal when modulations in visual intensity and auditory amplitude are temporally congruent (Denison et al., 2013; 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 is still unfolding.

Previous studies of neuronal correlates of audiovisual integration found that the primary sensory cortical areas participate in this process (Wang et al., 2008; Iurilli et al., 2012; Ibrahim et al., 2016; Deneux et al., 2019; McClure and Polack, 2019; Meijer 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), via auditory signals shown to originate in the primary auditory cortex (Deneux 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. Whereas initial studies found a suppressive signal provided by the primary auditory cortex to V1 (Iurilli et al., 2012; Ibrahim et al., 2016), later studies found heterogeneous changes across neurons in visual tuning curve bandwidth with and without sound (Meijer et al., 2017; McClure and Polack, 2019). These contrasting findings raise the question of whether previously reported sound-induced changes in V1 neuronal activity in awake animals results 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 largely unaccounted for in audiovisual studies is that awake animals are subject to brain-wide changes in neuronal activity because of 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; Dadarlat and Stryker, 2017). Conversely, in the auditory cortex (AC), locomotion suppresses spontaneous and auditory stimulus-related responses (Nelson et al., 2013; Schneider and Mooney, 2018; Bigelow et al., 2019). Therefore, the contribution of movement to neuronal responses to multisensory stimuli is likely because of 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 by-product of uninstructed sound-induced movements which themselves modulate visual responses (Bimbard et al., 2023). Here, we tested these alternative explanations of the extent to which movement 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), to determine whether and through what coding mechanism this cross-modal interaction improves visual encoding in awake mouse subjects. The audiovisual stimulus consisted of auditory white noise and visual drifting gratings 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 National Institutes of Health guidelines and approved by the IACUC at the University of Pennsylvania. Mice were acquired from The Jackson Laboratory [five male, six female, aged 10–18 weeks at time of recording; *B6N(g).Cdh23* 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 conducted 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 is available on Dryad: https://doi.org/10.5061/dryad.sxksn033q (Williams et al., 2023). Software is available on zenodo: https://doi.org/10.5281/zenodo.6603398.

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 ~1-mm 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 d post-surgery before any additional procedures took place.

Electrophysiological recordings

All recordings were conducted inside a custom-built acoustic isolation booth. One to two weeks following the headplate and ground pin attachment surgery, we habituated the mice to the recording booth for increasing durations (5, 15, 30 min) over the course of 3 d. 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, 2.5 mm 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-Poly2-5 mm-50s-177). The electrode was lowered into the primary visual cortex via a stereotactic instrument to a depth of 775–1000 μm. Following the audiovisual stimulus presentation, electrophysiological data from all 32 channels were filtered between 600 and 6000 Hz and re-referenced to the median across all channels. Spikes were identified, sorted, and initially assigned to single units or multiunits using Kilosort2 (Pachitariu et al., 2016), which we then visualized using the publicly available phy2 graphic interface. Putative single units which displayed a clear refractory period and a single cluster of spatiotemporal features in principal component space were maintained as single units. Putative multiunits were manually split into single units if the principle component features revealed two distinct clusters. Otherwise, putative multiunits were maintained as multiunits.

Audiovisual stimuli

The audiovisual stimuli were generated using MATLAB (MathWorks), and presented to mice on a 12" LCD monitor (Eyoyo) with a 60-Hz 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 five 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 0–300 ms 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 0–300 ms; 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 supralinear or sublinear integration of the auditory and visual responses, we calculated the linearity ratio (LR) 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, 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 10-ms 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 SD 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 found by calculating half the complex phase (Niell and Stryker, 2008) of the value

$$S = \frac{\sum F(\theta)e^{2i\theta}}{\sum F(\theta)}.$$

We calculated orientation and direction-selective indices (Zhao et al., 2013) for each neuron according to:

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

$$\mathrm{DSI} = \frac{FR_{pref} - FR_{antipref}}{FR_{pref} + FR_{antipref}},$$

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 >75% of shuffled OSI or DSI values, the neuron was classified as "orientation-selective" 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 (nine recordings, five mice), mouse movement was tracked throughout stimulus presentation. Video recording was performed using a Raspberry Pi four 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 then quantified using the freely available Facemap software (Stringer et al., 2019). Within the Facemap GUI, we identified

regions of interest (ROIs) on the whiskers and face to capture whisking behavior, ROIs on the extremities to capture locomotive behavior, and ROIs distributed across the mouse subject to capture general, nonspecific movement, which captured both whisking and locomotion. The separate motion energy output from each regions 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 (using the general nonspecific movement trace); response variable: trial spikes during 0–300 ms; 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 also visualized the overall distribution of mouse subject movement across trials by calculating a z score for each trial. Using the nonspecific mouse movement value from each trial, we grouped together trials from each recording session, subtracted the group average, and divided by the group SD to obtain a z score for each trial. This z score represented whether the mouse moved more or less compared with other trials from that recording session.

In order to reconstruct peristimulus time histograms of light-responsive, sound-modulated, motion-modulated neurons, we used a separate GLM. Using a 10-ms sliding window across all trials, we input the visual contrast level, sound presence, and general nonspecific movement 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 to determine differential effects these parameters had on the temporal trajectory of neurons' visual responses. In a separate analysis, we used a similarly structured GLM, but replaced the "general nonspecific movement" predictor variable with independent locomotion and whisking variables, using the Facemap output from the locomotion-related and whisking-related ROIs. This allowed us to additionally report how locomotion and whisking individually modulate visual responses, as opposed to grouped into a single nonspecific movement variable.

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:

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

where μ_θ and σ_θ are the response mean and SD, 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 to test sensitivity index as a function of angular distance from preferred direction.

We used a maximum likelihood estimate (MLE) 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 (one of two options), as well as decoding one direction from all possible directions (one 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 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 multigroup 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 ($p < 2.2e\text{-}301$), because of characteristically large datasets. 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 is still being revealed. Furthermore, whether that interaction can be exclusively attributed to sound or to sound-induced motion is controversial (Bimbard et al., 2023). To elucidate the principles underlying audiovisual integration, we presented audiovisual stimuli to awake mice while performing extracellular recordings in V1 (Fig. 1*A*). The visual stimulus consisted of drifting gratings in 12 directions presented at five visual contrast levels (Fig. 1*B*), ranging from 0% to 100%, with a static gray screen between trials. 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 (Fig. 1*C*), affording 10 trials of each unique audiovisual stimulus condition (Fig. 1*C*). Twelve recording sessions across

Table 1. Statistical comparisons

Comparison	Figure	Test	Test statistic	*N*	Df	*p*-value	*Post hoc* test	*Post hoc* α	*Post hoc* comparison	*Post hoc* *p*-value
Orientation selective index	2*C*	*t* test	*t* stat = −1.0	n_{1000ms} = 303 n_{300ms} = 269	565	*p* = 0.30				
Direction selective index	2*C*	*t* test	*t* stat = −1.6	n_{1000ms} = 143 n_{300ms} = 144	281	*p* = 0.10				
Mean firing rate, V vs AV	3*C*	Paired two-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 Contrast 0.25, V vs AV Contrast 0.5, V vs AV Contrast 0.75, V vs AV Contrast 1, V vs AV	2.1e-50 2.6e-62 5.7e-75 1.1e-81 2.0e-81
Linearity ratio, V vs AV	3*E*	Kruskal–Wallis test	χ^2 = 61	555 neurons	4	*p* = 1.6e-12	Bonferroni-corrected Wilcoxon signed-rank test	0.0125	Contrast 0 vs 0.25 Contrast 0 vs 0.5 Contrast 0 vs 0.75 Contrast 0 vs 1	0.053 0.0040 4.6e-8 2.1e-5
Orientation selectivity index, V vs AV	4*E*	Paired *t* test	*t* stat = 4.8	269 neurons	268	*p* = 2.4e-6				
Direction selectivity index, V vs AV	4*F*	Paired *t* test	*t* stat = 3.5	144 neurons	143	*p* = 6.4e-4				
Onset response latency, V vs AV	5*B*	Paired two-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 Contrast 0.5, V vs AV Contrast 0.75, V vs AV Contrast 1, V vs AV	2.3e-4 7.1e-12 4.6e-5 9.9e-4
Onset response slope, V vs AV	5*D*	Paired two-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 Contrast 0.5, V vs AV Contrast 0.75, V vs AV Contrast 1, V vs AV	1.4e-4 8.9e-13 3.6e-12 5.5e-8
Onset response duration, V vs AV	5*F*	Paired two-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	5*H*	Paired two-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	6*D*	Paired *t* test	*t* stat = −14.9	9 recording sessions	8	*p* = 4.0e-7				
Firing rate across movement range, V vs AV	6*G*	Unbalanced two-way ANOVA	*F*(motion) = 18 *F*(sound) = 32 *F*(interact) = 17	Variable trial count	mot = 2 aud = 1 Interact = 2	*p*(motion) = 1.6e-8 *p*(sound) = 1.6e-8 *p*(interact) = 3.1e-8	Bonferroni corrected two-sample *t* test	0.016	Stationary, V vs AV Low motion, V vs AV High motion, V vs AV	1.0e-15 3.1e-10 0.59
GLM PSTH, light vs light/sound	7*F*	Paired *t* test	1391 unique *t* stats	343 neurons	342	1391 unique *p*-values, α = 0.05/1391 = 3.6e-5				
GLM PSTH, light vs light/motion	7*G*	Paired *t* test	1391 unique *t* stats	343 neurons	342	1391 unique *p*-values, α = 0.05/1391 = 3.6e-5				
GLM PSTH, light vs light/sound/motion	7*H*	Paired *t* test	1391 unique *t* stats	343 neurons	342	1391 unique *p*-values, α = 0.05/1391 = 3.6e-5				
Orientation decoding accuracy, individual neurons, V vs AV	8*E*	Paired two-way ANOVA	*F*(vis) = 73 *F*(aud) = 50 *F*(interact) = 7.2	264 neurons	vis = 4 aud = 1 interact = 4	*p*(vis) = 4.8e-112 *p*(aud) = 1.7e-11 *p*(interact) = 1.0e-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	0.78 1.5e-4 2.2e-11 0.21 1.4e-6
Direction decoding accuracy, individual neurons, V vs AV	8*G*	Paired two-way ANOVA	*F*(vis) = 20 *F*(aud) = 12 *F*(interact) = 0.39	140 neurons	vis = 4 aud = 1 interact = 4	*p*(vis) = 1.2e-14 *p*(aud) = 6.9e-4 *p*(interact) = 0.82				
Orientation decoding accuracy, MLE, population, V vs AV	9*D*	Two-way ANOVA	*F*(vis) = 720 *F*(aud) = 2.8 *F*(interact) = 26	50 repeats	vis = 4 aud = 1 interact = 4	*p*(vis) = 2.6e-98 *p*(aud) = 0.098 *p*(interact) = 1.7e-82	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	2.1e-7 1.5e-12 4.1e-9 1.4e-4 9.4e-4
Direction decoding accuracy, MLE, population, V vs AV	9*E*	Two-way ANOVA	*F*(vis) = 99 *F*(aud) = 0.03 *F*(interact) = 7.8	50 repeats	vis = 4 aud = 1 interact = 4	*p*(vis) = 4.2e-90 *p*(aud) = 0.87 *p*(interact) = 8.1e-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	0.21 5.8e-4 2.5e-4 0.48 0.97

(*Table continues*.)

Table 1. Continued

Comparison	Figure	Test	Test statistic	*N*	Df	*p*-value	*Post hoc* test	*Post hoc* α	*Post hoc* comparison	*Post hoc* *p*-value
Overall decoding accuracy, MLE, population, V vs AV	9*H*	Two-way ANOVA	F(vis) = 137 F(aud) = 13 F(interact) = 5.5	20 repeats	vis = 4 aud = 1 interact = 4	p(vis) = 8.7e-55 p(aud) = 4.2e-4 p(interact) = 3.3e-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	0.011 2.9e-4 0.090 0.0054 0.57
Orientation decoding accuracy, individual neurons, V vs AV	10*B*	Paired two-way ANOVA	F(vis) = 80 F(aud) = 18 F(interact) = 1.2	90 neurons	vis = 4 aud = 1 interact = 4	p(vis) = 1.2e-80 p(aud) = 5.4e-5 p(interact) = 0.32				
Orientation decoding accuracy, individual neurons, V vs motion-corrected AV	10*B*	Paired two-way ANOVA	F(vis) = 82 F(aud) = 8.1 F(interact) = 1.7	90 neurons	vis = 4 aud = 1 interact = 4	p(vis) = 6.2e-81 p(aud) = 5.3e-3 p(interact) = 0.15				
Orientation decoding accuracy, individual neurons, V vs sound-corrected AV	10*B*	Paired two-way ANOVA	F(vis) = 70 F(aud) = 1.5 F(interact) = 1.9	90 neurons	vis = 4 aud = 1 interact = 4	p(vis) = 7.7e-72 p(aud) = 0.23 p(interact) = 0.11				
Population decoding accuracy, V vs AV	10*D*	Two-way ANOVA	F(vis) = 337 F(aud) = 133 F(interact) = 4.2e-10	10 repeats	vis = 4 aud = 1 interact = 4	p(vis) = 3.1e-53 p(aud) = 2.0e-19 p(interact) = 4.2e-10	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	0.041 5.6e-6 3.5e-6 7.0e-7 1.4e-4
Population decoding accuracy, V vs motion-corrected AV	10*D*	Two-way ANOVA	F(vis) = 230 F(aud) = 74 F(interact) = 6.7	10 repeats	vis = 4 aud = 1 interact = 4	p(vis) = 2.5e-46 p(aud) = 2.1e-13 p(interact) = 9.3e-5	Bonferroni-corrected paired *t* test	0.01	Contrast 0, V vs motion-corrected AV Contrast 0.25, V vs motion-corrected AV Contrast 0.5, V vs motion-corrected AV Contrast 0.75, V vs motion-corrected AV Contrast 1, V vs motion-corrected AV	0.94 1.1e-4 0.01 0.0023 0.021
Population decoding accuracy, V vs sound-corrected AV	10*D*	Two-way ANOVA	F(vis) = 192 F(aud) = 0.02 F(interact) = 9.3	10 repeats	vis = 4 aud = 1 interact = 4	p(vis) = 3.3e-43 p(aud) = 0.88 p(interact) = 3.4e-6	Bonferroni-corrected paired *t* test	0.01	Contrast 0, V vs motion-corrected AV Contrast 0.25, V vs motion-corrected AV Contrast 0.5, V vs motion-corrected AV Contrast 0.75, V vs motion-corrected AV Contrast 1, V vs motion-corrected AV	0.87 0.039 0.19 0.080 0.0025
Population decoding accuracy, V vs locomotion-corrected AV	10*F*	Two-way ANOVA	F(vis) = 387 F(aud) = 68 F(interact) = 8.1	10 repeats	vis = 4 aud = 1 interact = 4	p(vis) = 7.8e-56 p(aud) = 1.3e-12 p(interact) = 1.3e-5	Bonferroni-corrected paired *t* test	0.01	Contrast 0, V vs locomotion-corrected AV Contrast 0.25, V vs locomotion-corrected AV Contrast 0.5, V vs locomotion-corrected AV Contrast 0.75, V vs locomotion-corrected AV Contrast 1, V vs locomotion-corrected AV	0.010 3.5e-4 0.019 1.5e-3 2.3e-3
Population decoding accuracy, V vs whisking-corrected AV	10*F*	Two-way ANOVA	F(vis) = 387 F(aud) = 68 F(interact) = 8.1	10 repeats	vis = 4 aud = 1 interact = 4	p(vis) = 1.1e-53 p(aud) = 1.3e-14 p(interact) = 2.3e-8	Bonferroni-corrected paired *t* test	0.01	Contrast 0, V vs whisking-corrected AV Contrast 0.25, V vs whisking-corrected AV Contrast 0.5, V vs whisking-corrected AV Contrast 0.75, V vs whisking-corrected AV Contrast 1, V vs whisking-corrected AV	4.1e-3 3.8e-4 7.3e-3 1.4e-4 2.1e-3

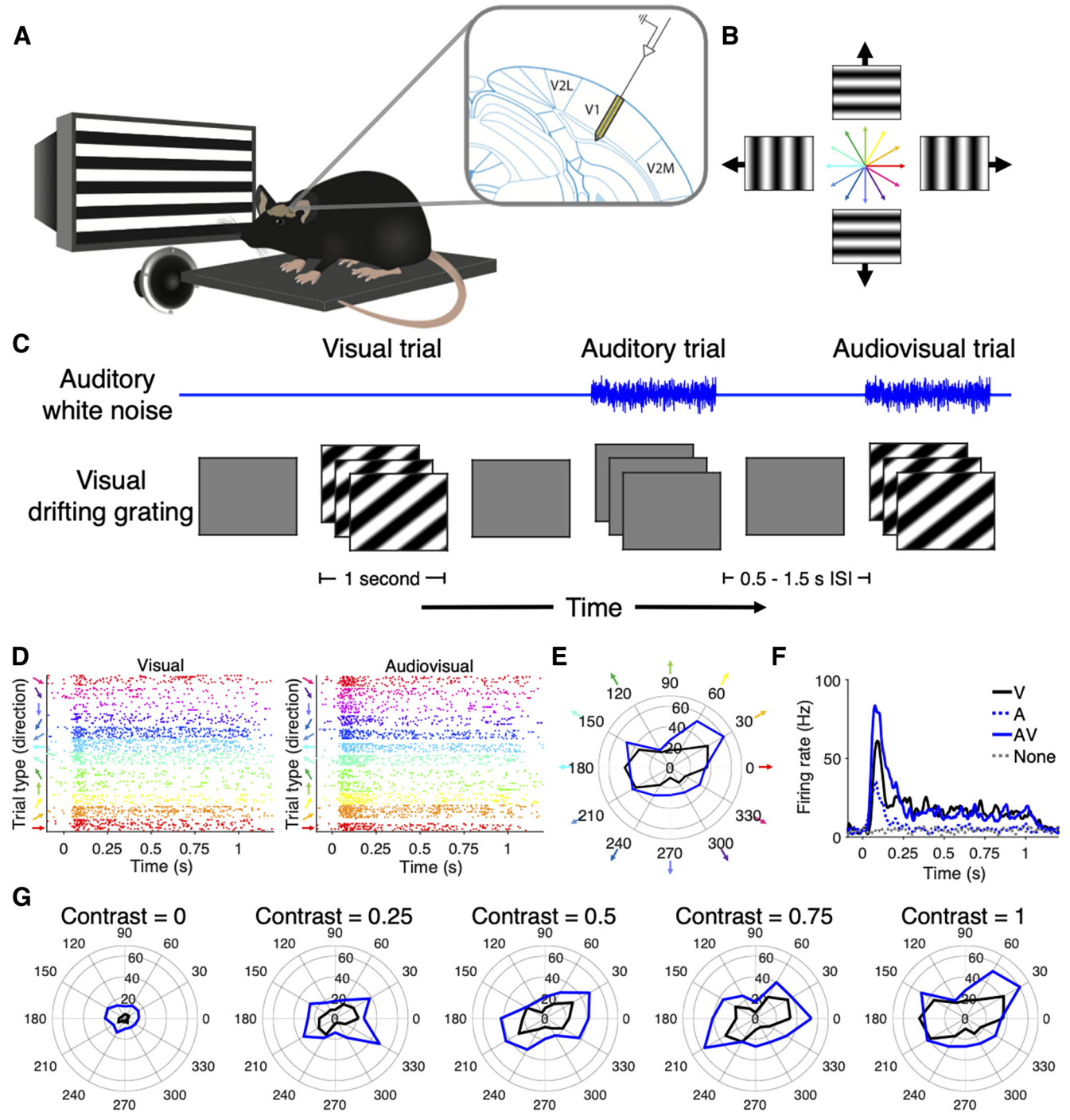

Figure 1. Audiovisual stimulus presentation. ***A***, Diagram (left) demonstrating that mice were head-fixed and presented with audiovisual stimuli from the right spatial field while electrophysiological recordings were performed in V1 (right). ***B***, Visual stimuli consisted of drifting gratings of 12 directions. ***C***, Auditory, visual, and audiovisual trials were randomly ordered and spaced with variable interstimulus 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 with unimodal stimuli. ***G***, Example neuron in ***E*** displays enhanced firing rate with sound across visual contrast levels.

six mice were spike sorted, and the responses of these sorted neurons were organized by trial type to compare across audiovisual stimulus conditions. We identified a total of 816 units across recordings, 161 (19.7%) of which were single units. Figure 1*D–G* demonstrates an example unit tuned for gratings aligned to the 30–210° axis whose baseline and light-evoked firing rate are both 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 (Fig. 2*A–C*; Table 1). 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 six mice) were significantly

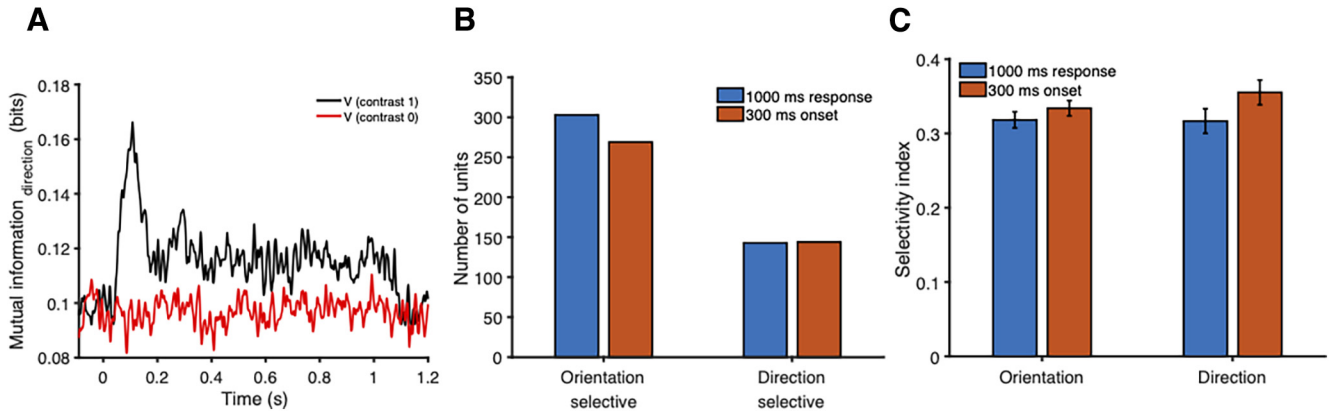

Figure 2. Classification based on sustained and onset responses. ***A***, Mutual information (MI) between neuronal responses and drifting grating direction, averaged across neurons. The black line is MI at full visual contrast, and the red line is MI at zero visual contrast. ***B***, We found a slight reduction in the number of neurons classified as orientation (269 vs 303 units) or direction selective (144 vs 143 units) when based on the initial 300-ms onset response compared with the entire 1000-ms response. ***C***, The OSI was similar (0.32 vs 0.33, $n_{1000} = 303$, $n_{300} = 269$, $p = 0.30$), as was the DSI (0.32 vs 0.36, $n_{1000} = 143$, $n_{300} = 144$, $p = 0.10$), of classified neurons when calculated using the initial 300-ms onset response compared the whole 1000-ms response.

modulated by the presence of sound (Fig. 3*A*). 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 (Fig. 3*F–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 (Fig. 3*B*). Averaged across neurons, we found a robust increase in the magnitude of the visually evoked response across visual contrast levels (Fig. 3*C*; p(vis) = 1.2e-100, p(aud) = 1.6e-88, p(interact) = 5.7e-4, paired two-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 did exhibit lower light-evoked and sound-evoked firing rates relative to baseline.

This change in firing rate can be potentially supralinear, linear or sublinear 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. We found that integration of the audiovisual stimulus was predominantly supralinear (Fig. 3*D*,*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), with 58.4% (329/563) of units with linearity ratio (LR) above 1, and 234/563 (41.6%) below 1. We also found that 6.9% (39/563) of units had LR below 0, associated with units in the top left and bottom right quadrants of Figure 3*D*. For clarity of the visualized data, Figure 3*E* excludes units with LR above 4 (18/563, 3.2%) and below −1 (17/563, 3.0%), corresponding to units that were exclusively activated under the audiovisual condition, or had contrasting enhancing and suppressive effects of sound and light. In summary, these results show that at a population level, sound supralinearly increases the magnitude of light-evoked responses; however, there is substantial variation between individual neurons.

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 how 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; Fahey et al., 2019). To characterize these tuning profiles, we calculated orientation and direction-selective indices (OSI and DSI) in audiovisually responsive neurons. In addition to this magnitude-based metric, we also calculated pseudo-indices based on randomly shuffled permutations of neurons' responses on each trial of orthogonal or opposite directions. Comparison of each neuron's true OSI or DSI to its respective distribution of pseudo-indices allowed us to incorporate trial-wise variability into the selectivity criteria. We classified each neuron as orientation or direction selective whose true OSI or DSI, respectively, were greater than the 75th percentile of its pseudo-indices. Figure 4*A*,*C* demonstrates the distribution of audiovisually responsive neurons' selectivity indices, with additional shading indicating multi and single units. Figure 4*B*,*D* shows example units along with the relationship between their true OSI or DSI and the distribution of pseudo-indices. Using this stringent selection criterion, we found that 47.8% (269/563) of neurons were orientation selective, whereas 25.6% (144/563) were direction selective. Surprisingly, we found a small reduction in the OSI from the visual to audiovisual conditions ($p = 2.4e-6$, paired Student's t test; Fig. 4*E*), which may reflect disproportionate changes in firing rate at the preferred versus orthogonal directions. We also found a slight reduction in DSI in the presence of sound ($p = 6.4e-4$, paired Student's t test; Fig. 4*F*). These sound-induced reductions in OSI and DSI were not as strong within single units ($p_{OSI,single}$ = 0.055, $p_{DSI,single}$ = 0.033; Fig. 4*E*,*F*), and were relatively uniform across unit depth (Fig. 4*G*). We also observed little shift in the preferred direction from the visual to audiovisual condition (Fig. 4*H*), as calculated as half the complex phase of the response profile at full visual contrast (Niell and Stryker, 2008). In order to determine how sound affected the shape of the tuning profile, we aligned neurons' tuning curves and normalized by each neuron's response to the full contrast visual stimulus. Surprisingly, we found that sound enhanced responses across tuning bandwidth, an effect that was present across visual contrast levels (Fig. 4*I*). Taken together, we observed that sound's enhancing effect on visual response magnitude resulted in a mild reduction in tuning selectivity in orientation and direction-selective neurons.

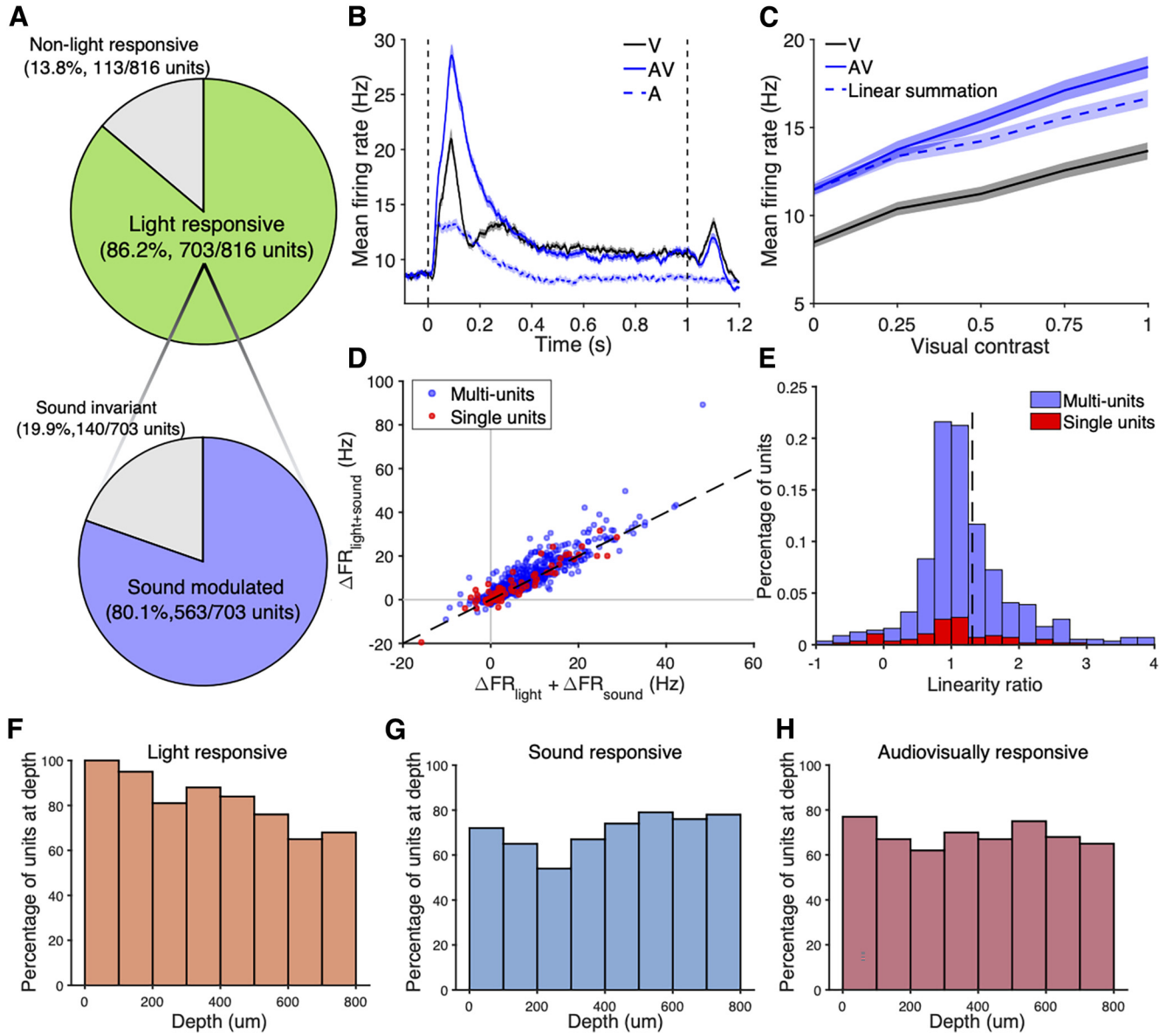

Figure 3. Sound enhances visual responses in a supralinear manner. ***A***, Sound modulates visually evoked activity in 80.1% of light-responsive neurons in V1. ***B***, Comparison of visual, auditory, and audiovisual PSTHs averaged across all light-responsive sound-modulated neurons. Visual and audiovisual PSTHs correspond to the highest visual contrast level. ***C***, The magnitude of audiovisual onset responses (0–300 ms) 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, two-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. In red are single units, and in blue are multiunits. ***E***, Histogram of linearity ratio among sound-modulated light-responsive neurons at full visual contrast. A linearity ratio above 1 demonstrates audiovisual responses in V1 represent supralinear integration of the unimodal signals ($n = 563$, $p = 1.6e-12$, Kruskal–Wallis test, *post hoc* Bonferroni-corrected Wilcoxon signed-rank test). Again red represents single units, and blue represents multiunits, and the dotted line is the population average. ***F–H***, Histograms demonstrating the percentage of neurons at each 100-μm depth bin that were classified as light, sound, and audiovisually responsive, based on the recording electrode with the largest spike waveform amplitude.

Changes in neuronal response latency, onset duration, and variability in audiovisual compared with 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). 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 SD above baseline (Fig. 5*A*), and found that sound reduced the response latency across contrast levels (Fig. 5*B*; p(vis) = 6.9e-4, p(aud) = 6.8e-15, p(interact) = 0.045, paired two-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 (Fig. 5*C*). We found that sound increased the slope of the onset response (Fig. 5*D*; p(vis) = 3.5e-121, p(aud) = 2.7e-15, p(interact) = 0.038, paired two-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 with the visual condition. Additionally, the

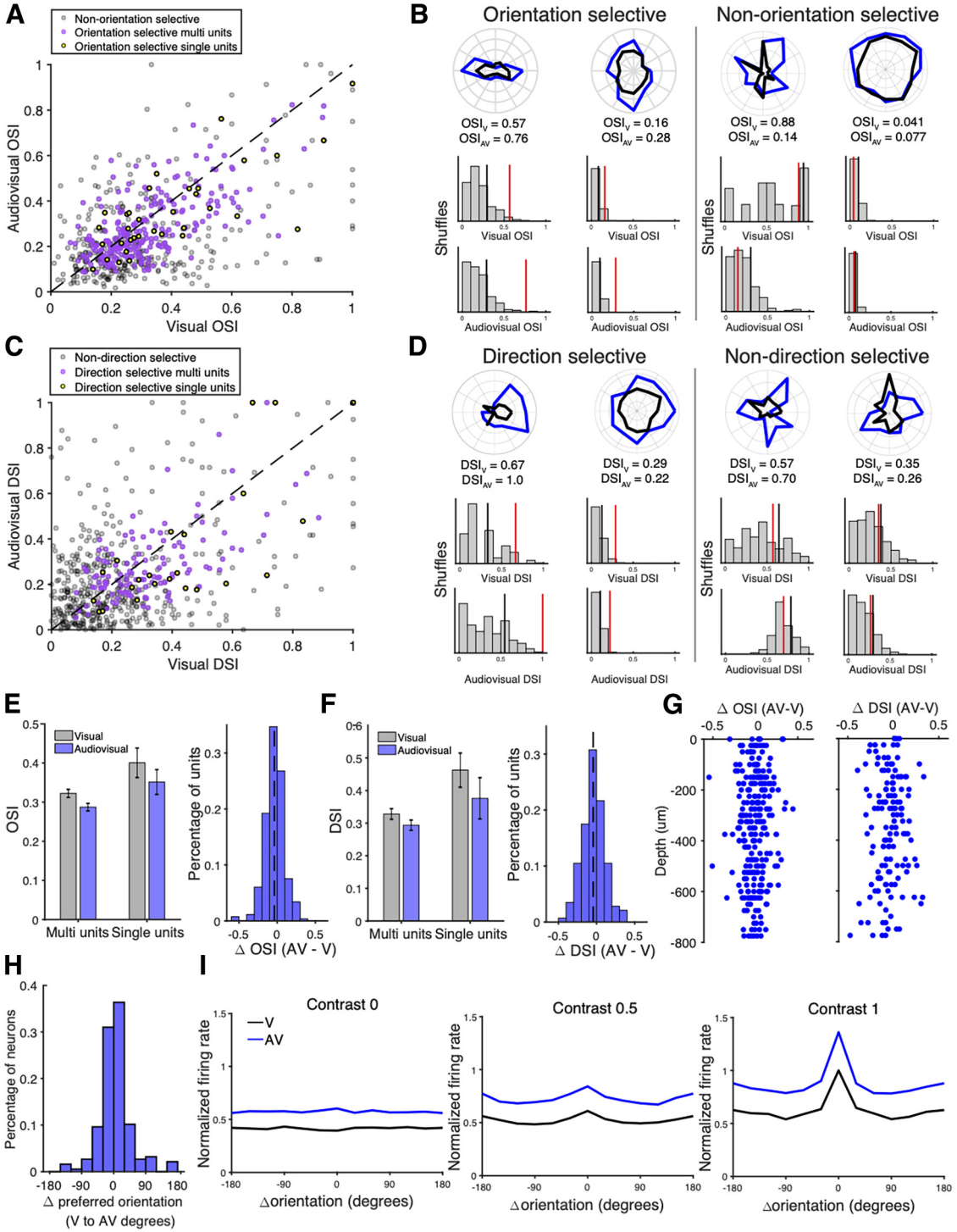

Figure 4. Sound reduces orientation and direction selectivity in tuned neurons. ***A***, Distribution of orientation selectivity index across audiovisually responsive units. Neurons with true OSI values above 75% of pseudo-indices are shaded in purple (multiunits) or yellow (single units). ***B***, Example units from ***A***, demonstrating each unit's range of pseudo-indices (gray histogram), the true OSI value (red bar), and the 75% threshold (black bar). ***C***, Distribution of direction selectivity index across audiovisually responsive units. Again, red neurons have true DSI above 75% of pseudo-indices. ***D***, Example units from ***C***, again with DSI pseudo-index distributions (gray histogram), true DSI value (red bar), and 75% threshold (black bar). ***E***, A mild reduction in OSI was observed in multiunits ($n = 234$, $p = 1.8e\text{-}5$, paired t test), with a trend in single units ($n = 34$, $p = 0.055$, paired t test). The distribution in ΔOSI shown on the right. ***F***, A mild reduction in DSI was observed in multiunits ($n = 120$, $p = 7.4e\text{-}3$, paired t test) and in single units ($n = 24$, $p = 0.033$, paired t test). The distribution in ΔDSI shown on the right. ***G***, Change in OSI (left) and DSI (right) was relatively uniform across cortical depth. ***H***, Histogram depicting changes in preferred drifting grating directions, calculated using half of the complex phase, with sound in orientation-selective neuron. ***I***, Tuning curves under the visual and audiovisual conditions across visual contrast levels, averaged across neurons, show nonspecific sound-induced enhancement.

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 (p(vis) = 1.3e-10, p(aud) = 8.7e-98, p(interact) = 0.23, paired two-way ANOVA; Fig. 5*E*,*F*). Both of these timing effects were preserved across contrast levels. Therefore, the latency and onset duration of neuronal audiovisual responses of V1 neurons is enhanced compared with visual responses.

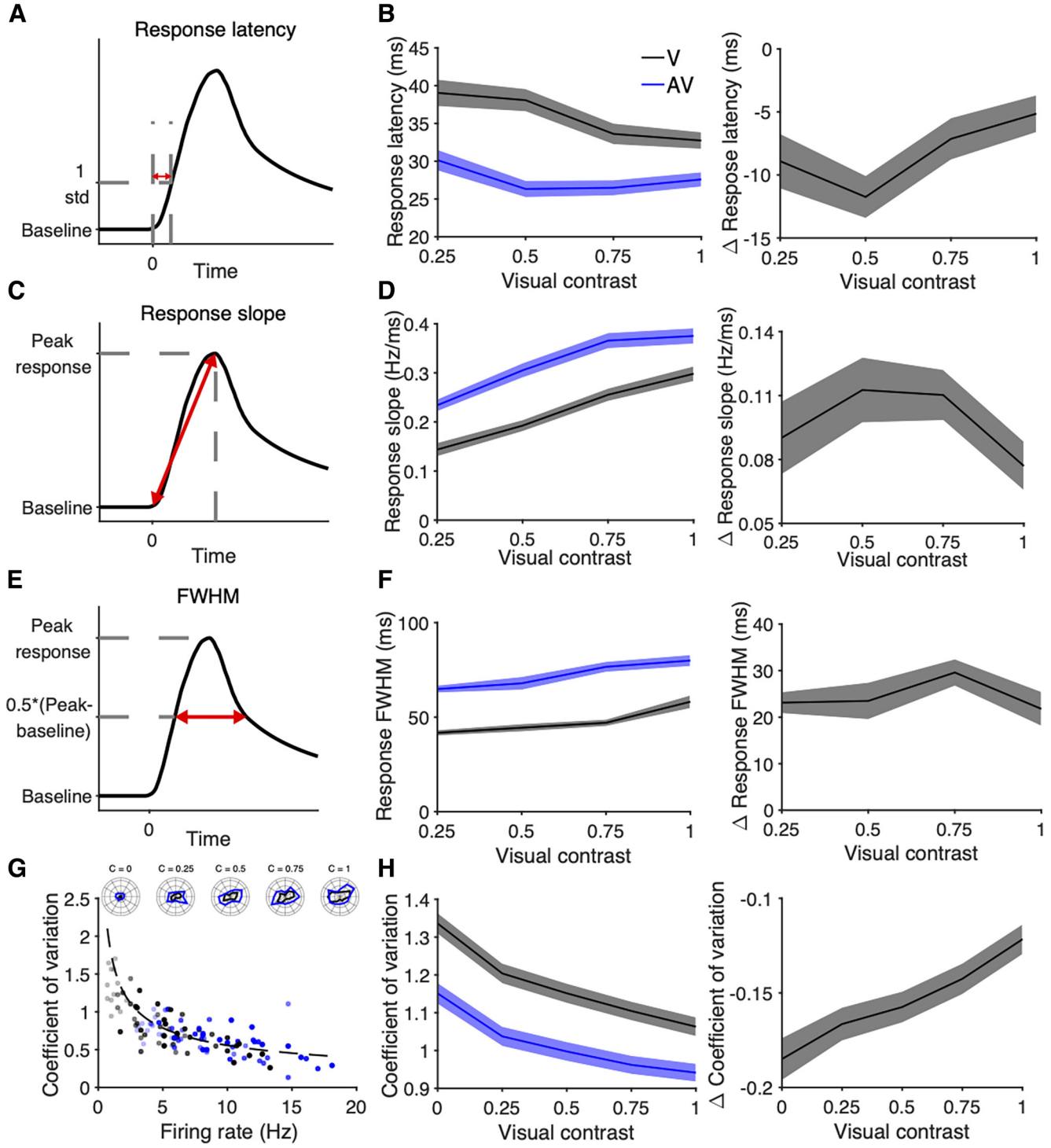

Figure 5. Changes in neuronal response latency, onset duration, and variability in audiovisual compared with visual conditions. ***A***, Diagram of the calculation of response latency, the first time bin in which the FR exceeds 1 SD above baseline. ***B***, Audiovisual response latency is less than that of the visual response (left: absolute, right: difference; p(vis) = 6.9e-4, p(aud) = 6.8e-15, p(interact) = 0.045, paired two-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 two-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 two-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 $y = c/\mathrm{sqrt}(x)$ 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 with visual responses (left: absolute, right: difference; n = 563, p(vis) = 0.28, p(aud) = 4.2e-103, p(interact) = 0.38, paired two-way ANOVA).

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 SD 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 5*G* 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 (Fig. 3), we observed a reduction of CV in the audiovisual condition relative to the visual condition when averaged across these neurons (p(vis) = 0.28, p(aud) = 4.2e-103, p(interact) = 0.38, paired two-way ANOVA; Fig. 5*H*). Taken together, these results indicate that sound not only modulates the magnitude of the visual response (Fig. 3), but also improves the timing and consistency of individual neurons' responses (Fig. 5).

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; Mesik et al., 2019) and auditory cortex (Nelson et al., 2013; Schneider and Mooney, 2018; Bigelow et al., 2019). Therefore, having established that sound robustly modulates visual responses (Fig. 3), we tested whether and to what extent these observed changes were more accurately attributable to sound-associated uninstructed movement in the mouse subjects. 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 (Fig. 6*A*). Despite being head-fixed to afford stable electrophysiological recordings, the mice were positioned on a smooth stage that freely allowed volitional movements. We used the publicly available Facemap software to process the video data (Stringer et al., 2019). Using its preprogrammed GUI, regions of interest (ROIs) were placed around the whiskers and face of mouse subjects to identify and quantify whisking and facial behavior (Fig. 6*B*, bottom). ROIs were also placed on the limbs to identify locomotion (Fig. 6*B*, middle), and additional ROIs were distributed across the entire mouse subject, including the face and limbs, to capture general nonspecific movements (Fig. 6*B*, top).

The energy output of each of these ROI regions throughout the video recording was then aligned with the audiovisual stimulus to process, identify, and quantify stimulus-correlated movements. For each recording session, we calculated averages for each trial type to compare visual and audiovisual movement responses for each mouse subject. We found that both visual and auditory stimuli did evoke whisking and locomotive behavior in mice, with combined audiovisual stimuli evoking a larger degree of both behaviors than isolated visual stimuli (Fig. 6*C*). Using the general movement trace, which included both locomotive and whisking behavior, we subtracted the 100-ms baseline before trial onset from the movement trace throughout the trial, and then divided by that baseline value to calculate fold increase over baseline. Using this method, we found that movement was higher during audiovisual trials compared with visual trials (p = 4.0e-7, paired t test; Fig. 6*D*). However, there were many visual trials in which substantial movement occurred, as well as audiovisual trials in which little movement was detected (Fig. 6*E*). Because of this 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 general movement activity during the onset (0–300 ms) of the trial. The vast majority of light-responsive neurons, 83.3% (400/480), displayed both sound-modulated and motion-modulated visual responses (Fig. 6*F*). 11.0% (53/480) and 1.7% (8/480) of light-responsive neurons were purely sound or motion modulated, respectively. An additional 4.0% (19/480) were invariant to sound or motion. We then compared the visually and audiovisually evoked firing rates of neurons when accounting for movement. Among sound-modulated and motion-modulated light-responsive neurons, the firing rate was higher on audiovisual trials than visual trials when movement was held constant (Fig. 6*G*), especially when mice showed limited movement. On trials in which the mice were largely stationary (z score < -0.5, 49% of visual trials, 33% of audiovisual trials) or displayed moderate levels of movement ($-0.5 < z$ score < 1.5, 45% of visual trials, 55% 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, 4.9% of visual trials, 12% of audiovisual trials; Fig. 6*G*,*H*; p(move) = 1.6e-8, p(aud) = 1.6e-8, p(interact) = 3.1e-8, unbalanced two-way ANOVA; $p_{\text{stationary}}$ = 1.0e-15, $p_{\text{low motion}}$ = 3.1e-10, $p_{\text{high motion}}$ = 0.59, *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 (Fig. 6*H*). Similar effects were observed when we alternatively organized trials by degrees of locomotion or whisking, with these uninstructed movements reducing the magnitude-enhancing effect that sound had on the visual responses. 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 display high 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 10-ms window to reconstruct the PSTH based on the visual contrast level, sound presence, and general movement, which included both locomotion and whisking behavior, during that time window (Fig. 7*A*). Figure 7*B* shows two example neurons in which the GLM estimated 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 the observed PSTHs (Fig. 7*C*–*E*). We leveraged the coefficients fit to each neuron (Fig. 7*A*) to estimate the unique contribution of each predictor to the firing rates as a function of time (see Materials and Methods). When the movement parameter was

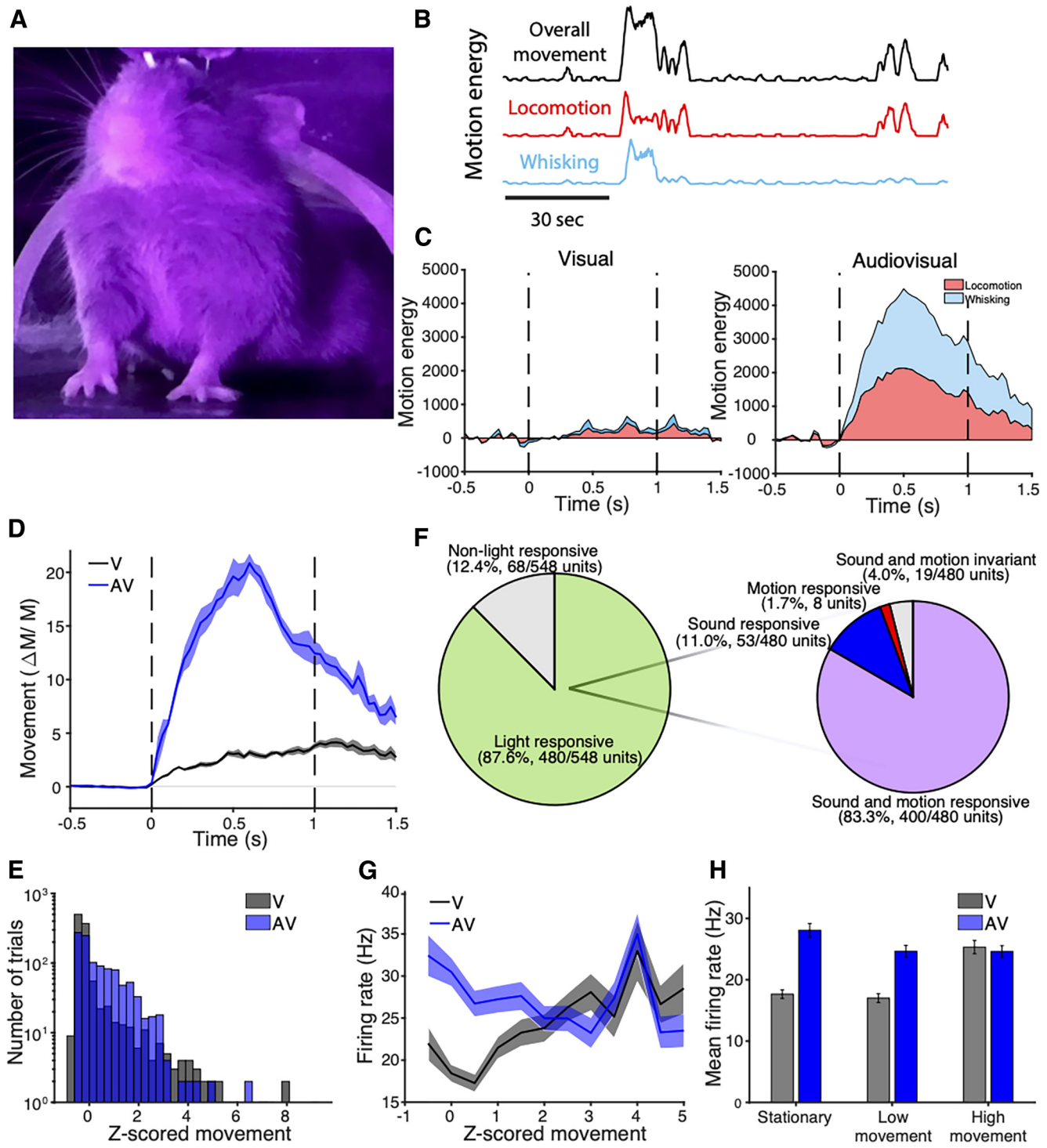

Figure 6. Sound modulates visual activity when controlling for stimulus-induced movement. ***A***, A still image demonstrating the video capture of mouse subjects during recording sessions. ***B***, Sample whisking, locomotion, and overall movement trace outputs using Facemap video analysis. ***C***, Average stimulus-aligned locomotion and whisking behavior on visual (left) and audiovisual (right) trials, relative to baseline 100 ms before stimulus onset. ***D***, Mice displayed more general movement, including both locomotion and whisking, response to audiovisual trials than in visual trials ($n = 9$ recording sessions; $p = 4.0\text{e-}7$, paired t test). ***E***, Histogram of trials' z-scored movements show a range of levels of movement during both visual and audiovisual trials. ***F***, Venn diagram demonstrating that 96% of light-responsive neurons exhibited some combination of sound and movement responsiveness. ***G***, Comparison of firing rate of sound-modulated and motion-modulated light-responsive neurons across trials with a range of z-scored movement. ***H***, 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) = 1.6e-8, p(aud) = 1.6e-8, p(interact) = 3.1e-8, two-way ANOVA, *post hoc* Bonferroni-corrected two-sample t test).

minimized, sound predominantly enhanced neuronal activity at the onset of the visual response and suppressed activity during the response's sustained period [$n = 295$ fitted neurons, paired t test at each time window (1391), $\alpha = 3.6\text{e-}5$; Fig. 7*F*]. Conversely, movement had limited effect on the onset activity in the absence of sound, but rather primarily enhanced firing rates during the response's sustained period [$n = 295$ fitted neurons, paired t test at each time window (1391), $\alpha = 3.6\text{e-}5$; Fig. 7*G*, red trace]. In a

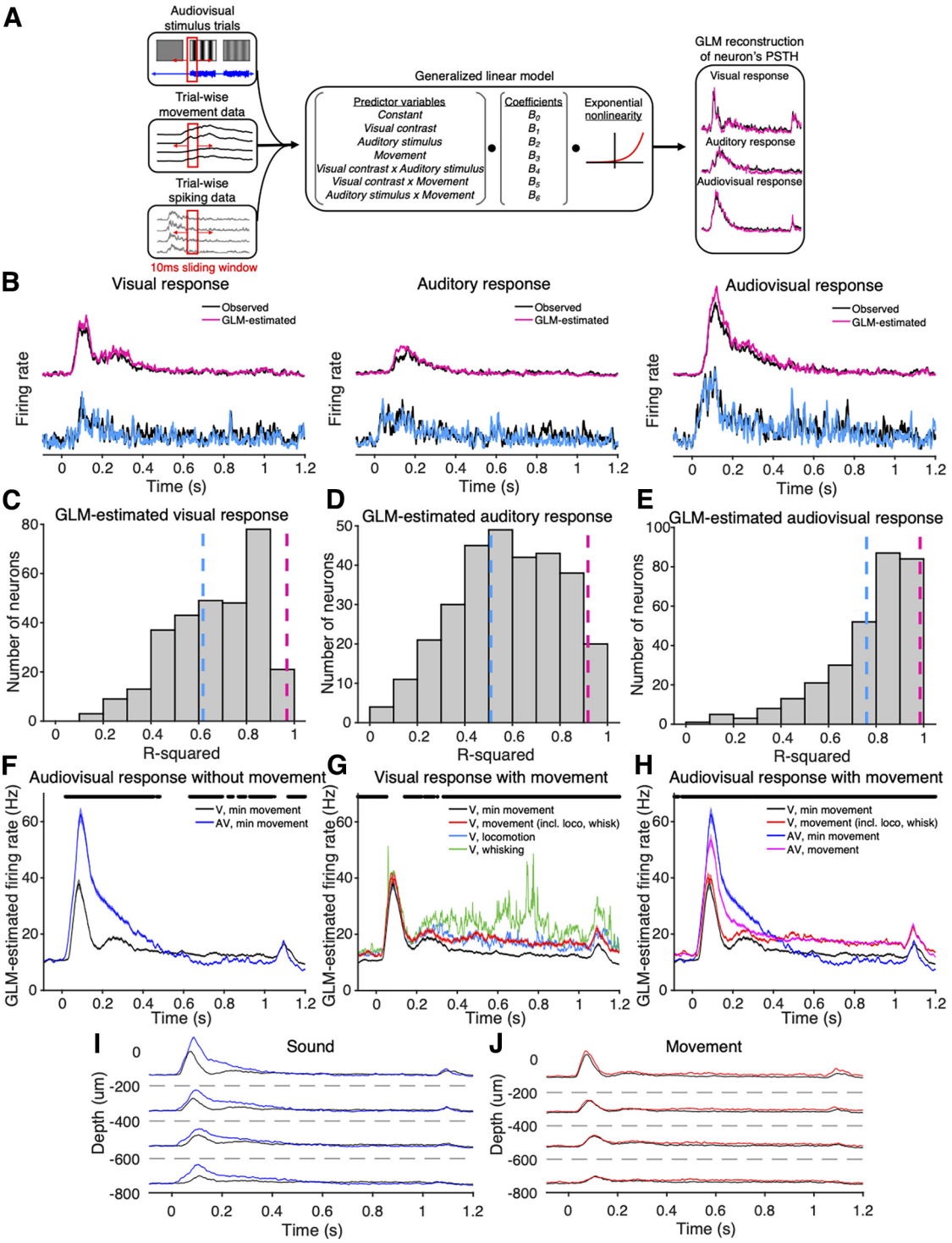

Figure 7. Sound and movement modulate visual responses in distinct but complementary ways. ***A***, Diagram illustrating the use of a GLM to reconstruct individual neurons' PSTHs based on neuronal responses and mouse movement during stimulus presentation. The GLM was then used to predict the time course 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 for two example units (blue and pink). ***C–E***, Histograms demonstrating R^2 values of the GLM-estimated PSTHs, averaged across sound-modulated 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. The dashed pink and blue line shows the R^2 value associated with the two example units in ***B***. ***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 ($n = 343$ fitted neurons; paired t test, $\alpha = 3.6e\text{-}5$). ***G***, Excluding sound highlights that nonspecific motion (red), as well as locomotion (blue) and whisking (green), primarily enhance the sustained portion of the response. Asterisks indicate time windows in which there was a significant difference ($n = 343$ fitted neurons; paired t test, $\alpha = 3.6e\text{-}5$). ***H***, Sound and motion together enhance both the onset and sustained periods of the visually evoked response compared with the isolated visual response ($n = 343$ fitted neurons; paired t test, $\alpha = 3.6e\text{-}5$). ***I***, Sound's enhancing effect on the visual response onset was slightly stronger at superficial cortical depths compared with deeper layers, when averaged across neurons in 200-μm depth bins. ***J***, Similarly, movement's enhancing effect on the visual response sustained portion was slightly stronger at superficial cortical depths, when averaged across neurons in 200-μm depth bins.

separate analysis, the nonspecific movement variable was replaced by two independent variables representing locomotion and whisking, and a similar GLM including coefficients for each movement subtype was fit to each neuron's PSTH. The estimated visual response PSTH with average locomotion and minimal whisking, as well as with average whisking and minimal locomotion, are also on display in Figure 7*G* (teal and green traces, respectively). Together, sound and movement had complementary effects in which both the onset and sustained portions of the visual response were enhanced compared with the isolated visual response [$n = 295$ fitted neurons, paired t test at each time window (1391), α = 3.6e-5; Fig. 7*H*, pink trace]. Again notably, the peak onset response under the audiovisual condition was lower when movement was included in the estimate (Fig. 7*H*, pink vs blue traces). We additionally grouped neurons by their cortical depth, and found that the distinct effects that sound and movement had on visual responses were largely preserved across layers (Fig. 7*I*,*J*), although were slightly larger in magnitude at superficial depths. 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. Many studies have reported how sound affects visual responses in V1, but whether these changes improve neuronal encoding of the visual stimulus in the awake brain has not been robustly 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-selective and direction-selective neurons (Fig. 8*A*,*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 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 and probe trial (Fig. 8*C*). This decoding technique achieves high decoding accuracy (Fig. 8*D*). When averaged across sound-modulated orientation-selective neurons, decoding performance was improved on audiovisual trials compared with visual trials (p(vis) = 4.8e-112, p(aud) = 1.7e-11, p(interact) = 1.0e-5, paired two-way ANOVA; $p_{c=0}$ = 0.78, $p_{c=0.25}$ = 1.5e-4, $p_{c=0.5}$ = 2.2e-11, $p_{c=0.75}$ = 0.21, $p_{c=1}$ = 1.4e-6; Fig. 8*E*), with the greatest improvements at low to intermediate contrast levels. We applied this approach to sound-modulated direction-selective units and found similar sound-induced improvements in decoding accuracy (p(vis) = 1.2e-15, p(aud) = 6.9e-4, p(interact) = 0.82, paired two-way ANOVA; Fig. 8*G*). Furthermore, similar effects were observed in both single unites and multiunits (Fig. 8*F*,*H*). These results demonstrate that sound-induced changes in response magnitude and consistency interact to improve neuronal representation of the visual stimulus in individual neurons.

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 again used a leave-one-out cross-validation approach when training and testing the decoder (Fig. 9*A*). Unsurprisingly, decoding accuracy improved as more neurons were included in the population (Fig. 9*B*). We began by using the MLE-based decoder to perform pairwise classification of visual drifting grating directions based on neuronal population activity. 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 increased on audiovisual trials as compared with visual trials (Fig. 9*C*). This improvement in performance was greatest when comparing orthogonal drifting grating orientations (Fig. 9*D*; p(vis) = 2.6e-98, p(aud) = 0.098, p(interact) = 1.7e-82, two-way ANOVA; $p_{c=0}$ = 2.1e-7, $p_{c=0.25}$ = 1.5e-12, $p_{c=0.5}$ = 4.1e-9, $p_{c=0.75}$ = 1.4e-4; $p_{c=1}$ = 9.4e-4, *post hoc* Bonferroni-corrected paired t test; Table 1). However, there was limited sound-induced improvement in decoding opposite drifting grating directions (Fig. 9*E*; p(vis) = 4.2e-90, p(aud) = 0.87, p(interact) = 8.1e-6, two-way ANOVA; $p_{c=0}$ = 0.21, $p_{c=0.25}$ = 5.9e-4, $p_{c=0.5}$ = 2.5e-4, $p_{c=0.75}$ = 0.48, $p_{c=1}$ = 0.97, *post hoc* Bonferroni-corrected paired t test; Table 1).

Expanding on the pairwise discriminability approach, the MLE-based decoder allowed us to also perform 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 (Fig. 9*F*–*H*), before reaching asymptotic performance at full visual contrast (Fig. 9*H*; p(vis) = 8.7e-55, p(aud) = 4.2e-4, p(interact) = 3.3e-4, two-way ANOVA; $p_{c=0}$ = 0.011, $p_{c=0.25}$ = 2.9e-4, $p_{c=0.5}$ = 0.090, $p_{c=0.75}$ = 0.0054, $p_{c=1}$ = 0.57, *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 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 sensorimotor inputs shape V1 visual responses (Niell and Stryker, 2010; Mesik et al., 2019), and locomotion improves visual processing in V1 (Dadarlat and Stryker, 2017). Thus, we next tested whether the observed sound-induced improvement in visual stimulus representation (Figs. 8, 9) 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 (Fig. 7). 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

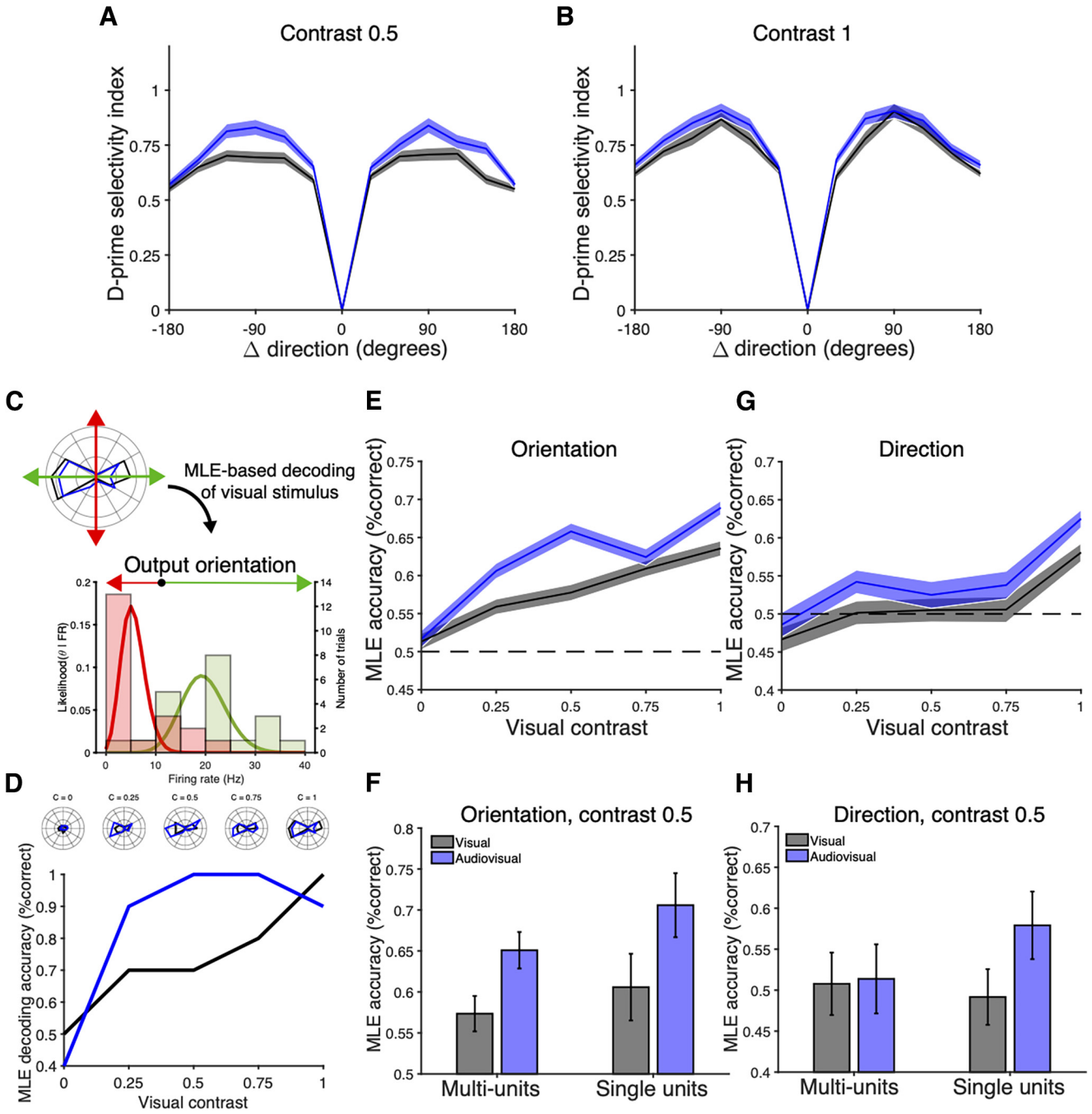

Figure 8. Sound improves decoding of drifting grating direction and orientation in individual neurons. ***A***, ***B***, The *d′* sensitivity index between neuronal responses to drifting grating directions, averaged across orientation-selective and direction-selective neurons. Enhancements are observed 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***, Absolute difference in decoding accuracy of preferred versus orthogonal orientations, averaged across sound-modulated orientation-selective neurons, demonstrating higher performance in the audiovisual condition ($n = 269$, p(vis) = 4.8e-112, p(aud) = 1.7e-11, p(interact) = 1.0e-5, paired two-way ANOVA). ***F***, Similar improvements in decoding were observed in both multiunites and single units. ***G***, Absolute difference in decoding accuracy of preferred versus anti-preferred directions, averaged across sound-modulated direction-selective neurons, again with improved performance on audiovisual trials compared with visual trials ($n = 144$, p(vis) = 1.2e-15, p(aud) = 6.9e-4, p(interact) = 0.82, paired two-way ANOVA). ***H***, The improvement in decoding probe trial direction was principally driven by single units, with limited effects observed in multiunits.

movements. We tested this hypothesis by expanding on the GLM-based classification of neurons described in Figure 7. Using the same GLM generated for each neuron, we independently modified either the sound or movement variables and their associated pairwise predictors to their lowest values, and then used the GLM coefficients and the exponential nonlinearity to estimate each neuron's audiovisual response magnitude (Fig. 10*A*; Materials and Methods). We then input these estimated trial-wise neuronal responses into the same MLE-based decoder described above (Figs. 8, 9). Using this approach, we found that in individual orientation-selective neurons, controlling for the effect of motion on audiovisual trials had little effect on the improvement in decoding accuracy on audiovisual trials (Fig. 10*B*,*C*; p(vis) = 6.2e-81, p(aud) = 5.3e-3, p(interact) = 0.15, paired

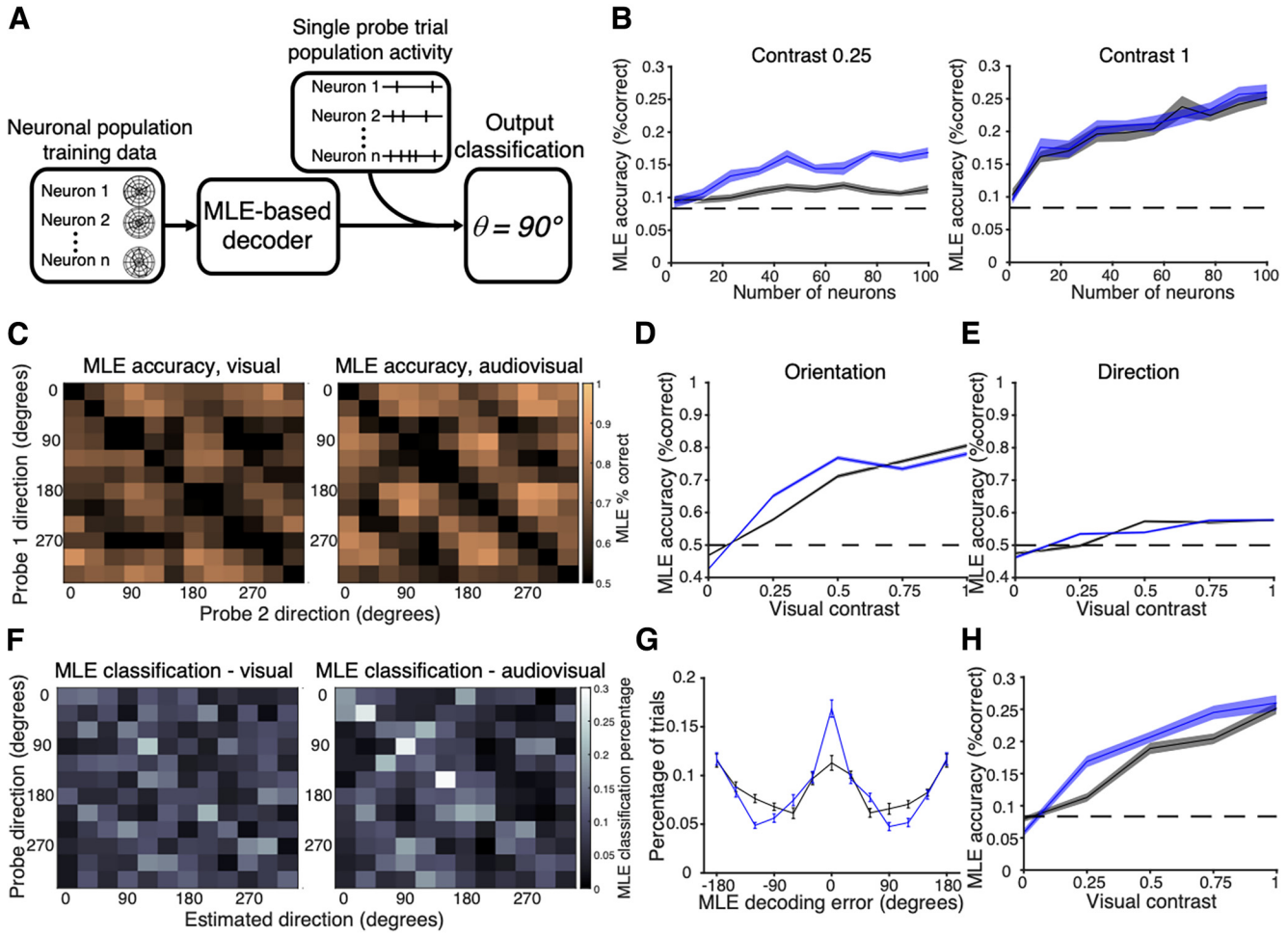

Figure 9. Sound improves accuracy of population-based visual stimulus decoding. ***A***, Schematic illustrating the decoding of the drifting grating direction using an MLE decoder trained on neuronal population activity. ***B***, Accuracy of MLE decoding 1-of-12 drifting grating options improved as the 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 pairwise classification of drifting gratings on visual (left) and audiovisual (right) trials, contrast 0.25. ***D***, MLE decoding accuracy when classifying orthogonal drifting grating orientations improved with sound ($n = 50$ randomizations, p(vis) = 2.6e-98, p(aud) = 0.098, p(interact) = 1.7e-82, two-way ANOVA, *post hoc* Bonferroni-corrected paired t test). ***E***, MLE decoding accuracy when classifying opposite drifting grating directions, demonstrating limited effect of sound on performance ($n = 50$ randomizations, p(vis) = 4.2e-90, p(aud) = 0.87, p(interact) = 8.1e-6, two-way ANOVA, *post hoc* Bonferroni-corrected paired t test). ***F***, Heat map of actual versus MLE-output directions under visual (left) and audiovisual (right) trials, contrast 0.25. MLE decoder could choose between all 12 drifting grating directions. ***G***, MLE decoder classification error, comparing estimated direction to actual direction. ***H***, Overall, decoding accuracy of MLE decoder when choosing between all 12 drifting grating directions improved with sound ($n = 10$ randomizations, p(vis) = 8.7e-55, p(aud) = 4.2e-4, p(interact) = 3.3e-4, two-way ANOVA, *post hoc* Bonferroni-corrected paired t test).

two-way ANOVA; Table 1). However, instead, regressing out sound from the audiovisual responses resulted in decoding accuracy that more closely resembled that of visual trials (Fig. 10*B*,*C*; p(vis) = 7.7e-72, p(aud) = 0.23, p(interact) = 0.11, paired two-way ANOVA; Table 1). These results in individual neurons suggest that sound and not movement primarily drives the improvement in decoding accuracy on 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 with that on visual trials (Fig. 10*D*,*E*; p(vis) = 2.5e-46, p(aud) = 2.1e-13, p(interact) = 9.3e-5, two-way ANOVA; $p_{c=0}$ = 0.94, $p_{c=0.25}$ = 1.0e-4, $p_{c=0.5}$ = 0.010, $p_{c=0.75}$ = 0.0023, $p_{c=1}$ = 0.021, Bonferroni-corrected paired t test). In contrast, alternatively regression out sound from audiovisual trials resulted in population decoding performance similar to that on visual trials (Fig. 10*D*,*E*; p(vis) = 3.3e-43, p(aud) = 0.88, p(interact) = 2.4e-6, two-way ANOVA; $p_{c=0}$ = 0.87, $p_{c=0.25}$ = 0.039, $p_{c=0.5}$ = 0.19, $p_{c=0.75}$ = 0.080, $p_{c=1}$ = 0.0025, Bonferroni-corrected paired t test). Additionally, we further refined our model by individually controlling for locomotion and whisking behaviors, as identified previously using Facemap software (Fig. 6). We again found that regressing out locomotion and whisking still resulted in MLE decoding performance that was significantly improved compared with visual trials (Fig. 10*E*; Locomotion: p(vis) = 7.8e-56, p(aud) = 1.3e-12, p(interact) = 1.3e-5, two-way ANOVA; $p_{c=0}$ = 0.010, $p_{c=0.25}$ = 3.5e-4, $p_{c=0.5}$ = 0.019, $p_{c=0.75}$ = 1.5e-3, $p_{c=1}$ = 2.3e-3, Bonferroni-corrected paired t test. Whisking: p(vis) = 1.1e-53, p(aud) = 1.3e-14, p(interact) = 2.3e-8, two-way ANOVA; $p_{c=0}$ = 4.1e-3, $p_{c=0.25}$ = 3.8e-4, $p_{c=0.5}$ = 7.3e-3, $p_{c=0.75}$ = 1.4e-4, $p_{c=1}$ = 2.1e-3, Bonferroni-corrected paired t test). 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

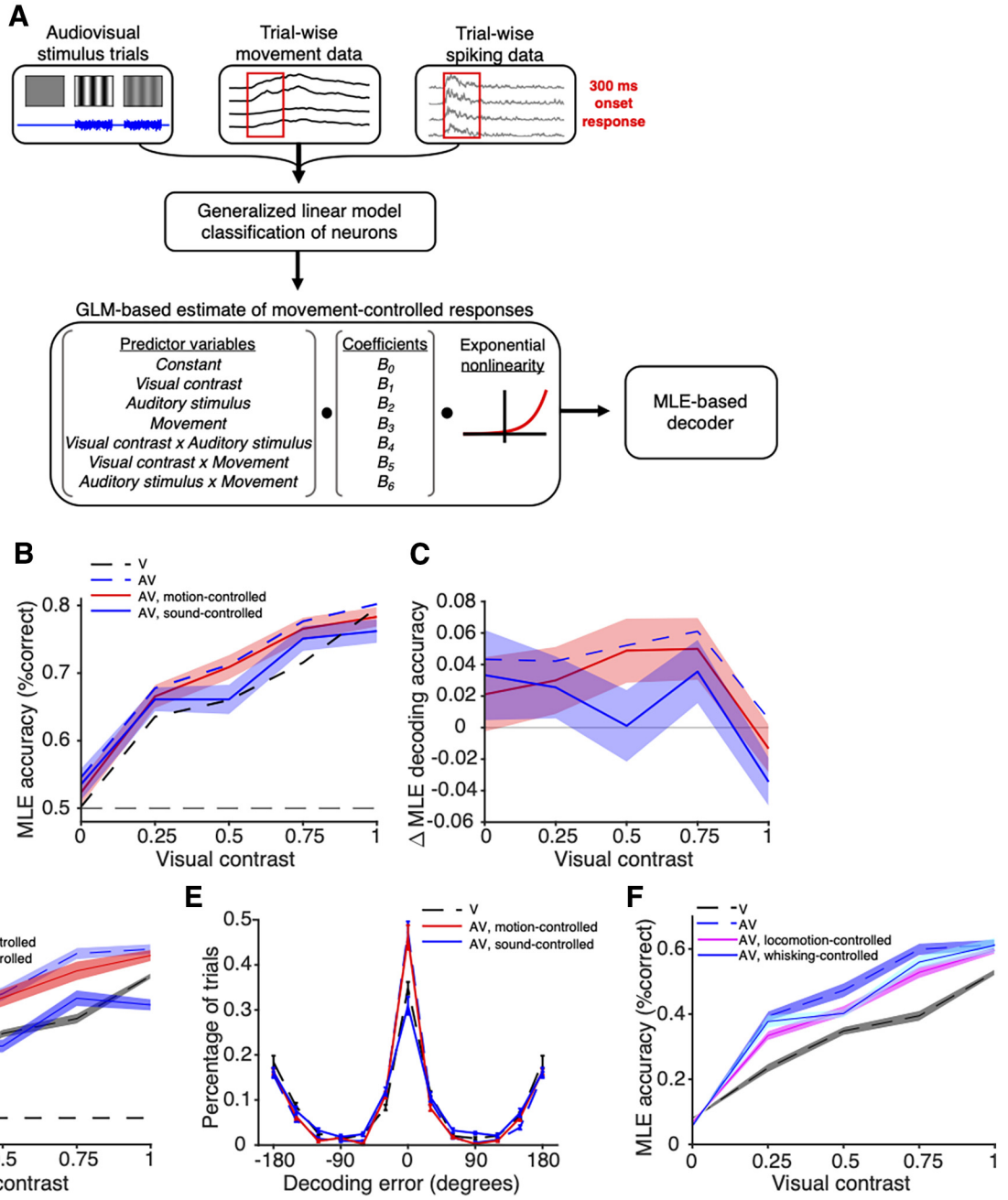

Figure 10. Sound improved decoding performance when controlling for motion. ***A***, Diagram illustrating the use of a GLM to calculate each predictor variable's coefficient. These are then used when varying the predictor variables to estimate trial-wise neuronal responses, which are then input into the MLE-based decoder. ***B***, Absolute accuracy of decoding orientation among orientation-selective, sound/motion-modulated light-responsive neurons, comparing visual responses (black, dotted) to audiovisual responses (blue, dotted), audiovisual responses when regressing out motion (red, solid) and audiovisual responses when regressing out sound (blue, solid). ***C***, Relative decoding accuracy compared with decoding on visual trials. Regressing out motion still preserved improved performance compared with visual trials ($n = 90$ neurons, p(vis) = 6.2e-81, p(aud) = 5.3e-3, p(interact) = 0.15, paired two-way ANOVA), whereas regressing out sound resulted in comparable performance to visual trials ($n = 90$ neurons, p(vis) = 7.7e-72, p(aud) = 0.23, p(interact) = 0.11, paired two-way ANOVA). ***D***, Population decoding accuracy of population-based decoder on audiovisual trials (blue, dotted) is preserved even when controlling for motion (red, solid; $n = 10$ randomizations, p(vis) = 2.5e-46, p(aud) = 2.1e-13, p(interact) = 9.3e-5, two-way ANOVA; Bonferroni-corrected paired t test), whereas controlling for sound (blue, solid) resembles decoding performance on visual trials (black dotted; $n = 10$ randomizations, p(vis) = 3.3e-43, p(aud) = 0.88, p(interact) = 2.4e-6, two-way ANOVA; Bonferroni-corrected paired t test). ***E***, MLE decoder classification percentage, comparing estimated direction to actual direction, contrast 0.5. Little difference is observed between audiovisual trials and audiovisual trials when controlling for motion, whereas both are more accurate than visual trials. ***F***, Population decoding accuracy on audiovisual trials (blue, dotted) is also preserved when controlling for locomotion (teal; $n = 10$ randomizations, p(vis) = 7.8e-56, p(aud) = 1.3e-12, p(interact) = 1.3e-5, two-way ANOVA; Bonferroni-corrected paired t test) and whisking (magenta; $n = 10$ randomizations, p(vis) = 1.1e-53, p(aud) = 1.3e-14, p(interact) = 2.3e-8, two-way ANOVA; Bonferroni-corrected paired t test).

used in everyday behaviors such as speech perception and object recognition (Fujisaki et al., 2014). The goal of the present study was to test whether sound drives improvement in encoding and decoding of visual stimuli 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 (Fig. 3). Notably, the effects we observed were stronger and more skewed toward response-enhancing than in previous studies (80.3% 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 represented a mixed effect of both sound and movement sensitivity (Fig. 6), an effect in which sound primarily enhances the onset response whereas movement complementarily enhances the sustained response (Fig. 7). We also found that the sound-induced changes in response magnitude and consistency combined to improve the discriminability of drifting grating orientation and direction in individual neurons (Fig. 8) and at a population level (Fig. 9). The improvements in neuronal encoding were most pronounced at low to intermediate visual contrast levels, a finding consistent with 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 (Fig. 10).

Auditory and locomotive inputs distinctly shape visual responses

We find that sound and movement have distinct and complementary effects on visual responses. Previous work found that locomotion modulates neuronal responses in the visual cortex in the presence of both sounds and visual stimuli but did not find an audio-specific interaction of locomotion's effect (McClure and Polack, 2019). Our results have revealed this component possibly because of the dynamics evoked by our sound stimulus, a white burst at a moderate sound pressure level which elicited a partial locomotive response. In our analysis, stimulus decoding relied largely on neuronal responses during the stimulus onset period. Therefore, despite robust differences in movement during visual and audiovisual trials, motion, which affected neuronal responses over slower time scales, only partially contributed to these changes in decoding (Fig. 10). 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 (Fig. 2; Dadarlat and Stryker, 2017). 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 (Niell and Stryker, 2010; Dadarlat 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 (Figs. 6*G*,*H*, 7*F*,*H*). 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 A1, 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 mice display robust movement. 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 (Fig. 4*E*,*F*). However, we also observed that sound reduced the CV of visual responses (Fig. 5), 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 (Fig. 8*A*,*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, despite the mild sound-associated reduction in OSI and DSI. The relevance of trial-wise variability is further supported by our observation of reliably orientation-selective and direction-selective neurons despite relatively low OSI and DSI (Fig. 4*A*–*D*), a metric agnostic to response variability. Prior studies using patch-clamp primarily in anesthetized animals approaches showed that V1 neurons sharpen their tuning profiles in response to sound a magnitude-based coding scheme, with some degree of recapitulation in the awake state (Ibrahim et al., 2016). The difference between these findings and those reported in the current study potentially indicate different coding schemes present in anesthetized and awake brains, additionally modified by unrestricted uninstructed movement during both visual and audiovisual trials, of which both factors are known to affect cortex-wide neuronal dynamics (Musall et al., 2019). 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).

Multiple studies found variable effects of sound presentation on visual responses in V1 (Meijer et al., 2017; McClure and Polack, 2019). Dependent on selection criteria of stimulus responsiveness, stimulus parameters, and visual contrast level, both studies found that sound presentation evoked either enhancement or suppression of V1 activity. Importantly, these studies found that highly selective neuronal responses to

visual stimuli are enhanced by sounds at intermediate visual contrast, which is in agreement with our analysis. Some differences could potentially be attributed to differences in sound presentation and stimulus selection. Meijer et al. (2017) find that visual responses are enhanced on average by sound stimuli that are congruent with the visual stimulus. It is possible that the congruent stimulus for the majority of neurons approximated the neurons' preferred temporally frequency, which activated neurons similarly to the white noise burst in our study. Meijer et al. (2017) also used a white noise burst and found that this auditory stimulus enhanced neuronal responses at intermediate visual contrast levels and suppressed them at high visual contrast levels. The difference between our results and those by McClure and Polack, 2019 may also be attributed to differences in the sound stimulus: McClure and Polack (2019) used pure tones that would activate neurons tuned to that particular frequency. Conversely, the duration of our white noise was matched to the visual stimulus, which were also longer in our study, therefor recruiting neurons in V1 across layers. Combined, our findings build on these two studies to explicitly test the role of locomotion and its sound-evoked qualities, providing a deeper understanding of the effects of locomotion through an expanded generalized linear model and of changes in variability as the key component to improved population-level decoding.

Inhibitory and disinhibitory effects may arise at different points within V1 with distinct effects on neuronal dynamics. Iurilli et al.(2012) found that presentation of sound burst drove a suppressive effect in V1, driven by cortico-cortical excitatory projections from AC to infragranular neurons. The effects in supragranular layers were predominantly suppressive. We note that this study differed from ours in large part because of the visual stimulus. Iurilli et al. (2012) presented a brief light flash, which would activate V1 neurons in a different fashion than the stimulus that ours and later studies used, a prolonged drifting grating. Whereas a single flash may evoke wide-spread adaptation and suppression in V1, a drifting grating is a stimulus that targets specific V1 neurons depending on their orientation. Furthermore, we observed that sound may suppress the sustained portion of the visual response in the absence of motion (Fig. 7*F*), suggesting that sound-induced excitation and inhibition may be temporally dependent as well.

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. Indeed, it is known that the audiovisual stimulus profile affects the degree to which sound is integrated with the visual signal (Bizley et al., 2016; Meijer et al., 2017; Atilgan et al., 2018). 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. 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 static white noise that is spatially consistent with the visual stimulus 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; Bremen 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., 2011), 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.

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 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.

In conclusion, in everyday life, we combine information across modalities in perception. For example, we watch the facial movement of our conversation partner to help with speech comprehension when significant background noise is present. Therefore, understanding how information across sensory modalities is combined is essential to the study of perception. We designed our study of audiovisual integration to determine whether sound improved V1 neurons' processing of visual inputs. The results revealed that sound improves neuronal encoding of the visual stimulus, especially at intermediate visual contrast levels, which we hypothesize would align with behavioral detection thresholds observed in psychophysics studies (Gleiss and Kayser, 2012; Meijer et al., 2018; Stein et al., 2020). The fact that this improvement in neuronal processing occurs in a passive manner in a primary sensory area, in the absence of any associated goal-directed task, likely underscores the importance of multisensory integration is to an organism's functioning (Stein et al., 2014). We additionally revealed a tripartite interaction through which movement shapes neuronal responses to audiovisual input. This additional sensorimotor interaction supports the idea that neuronal activity is influenced by global brain and bodily states (Musall et al.,

2019). Therefore, our perceptual experience is a reflection of a multitude of sensorimotor inputs and a product of dynamic, hierarchical, yet highly integrated processing performed by our brains.

References

Atilgan H, Town SM, Wood KC, Jones GP, Maddox RK, Lee A, Bizley JK (2018) Integration of visual information in auditory cortex promotes auditory scene analysis through multisensory binding. Neuron 97: 640–655.e4.

Berens P, Ecker AS, Cotton RJ, Ma WJ, Bethge M, Tolias AS (2012) A fast and simple population code for orientation in primate V1. J Neurosci 32:10618–10626.

Bergan JF, Knudsen EI (2009) Visual modulation of auditory responses in the owl inferior colliculus. J Neurophysiol 101:2924–2933.

Bigelow J, Poremba A (2016) Audiovisual integration facilitates monkeys' short-term memory. Anim Cogn 19:799–811.

Bigelow J, Morrill RJ, Dekloe J, Hasenstaub AR (2019) Movement and VIP interneuron activation differentially modulate encoding in mouse auditory cortex. eNeuro 6:ENEURO.0164-19.2019.

Bimbard C, Sit TPH, Lebedeva A, Harris KD, Reddy CB, Carandini M (2023) Behavioral origin of sound-evoked activity in mouse visual cortex. Nat Neurosci 26:251–258.

Bizley JK, Maddox RK, Lee AKC (2016) Defining auditory-visual objects: behavioral tests and physiological mechanisms. Trends Neurosci 39:74–85.

Borst A, Theunissen FE (1999) Information theory and neural coding. Nat Neurosci 2:947–957.

Bremen P, Massoudi R, Van Wanrooij MM, Van Opstal AJ (2017) Audiovisual integration in a redundant target paradigm: a comparison between rhesus macaque and man. Front Syst Neurosci 11:89.

Chen YC, Huang PC, Yeh SL, Spence C (2011) Synchronous sounds enhance visual sensitivity without reducing target uncertainty. Seeing Perceiving 24:623–638.

Churchland AK, Kiani R, Chaudhuri R, Wang X, Pouget A, Shadlen MN (2011) Variance as a signature of neural computations during decision making. Neuron 69:818–831.

Cohen L, Rothschild G, Mizrahi A (2011) Multisensory integration of natural odors and sounds in the auditory cortex. Neuron 72:357–369.

Colonius H, Diederich A (2017) Measuring multisensory integration: from reaction times to spike counts. Sci Rep 7:3023.

Dadarlat MC, Stryker MP (2017) Locomotion enhances neural encoding of visual stimuli in mouse V1. J Neurosci 37:3764–3775.

Deneux T, Harrell ER, Kempf A, Ceballo S, Filipchuk A, Bathellier B (2019) Context-dependent signaling of coincident auditory and visual events in primary visual cortex. Elife 8:e44006.

Denison RN, Driver J, Ruff CC (2013) Temporal structure and complexity affect audio-visual correspondence detection. Front Psychol 3:619.

Diederich A, Colonius H (2004) Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. Percept Psychophys 66:1388–1404.

Fahey PG, Muhammad T, Smoth C, Froudarakis E, Cobos E, Fu J, Walker EY, Yatsenko D, Sinz FH, Reimer J, Tolias AS (2019) A global map of orientation tuning in mouse visual cortex. bioRxiv 745323. https://doi.org/10.1101/745323.

Falchier A, Clavagnier S, Barone P, Kennedy H (2002) Anatomical evidence of multimodal integration in primate striate cortex. J Neurosci 22:5749–5759.

Fujisaki W, Goda N, Motoyoshi I, Komatsu H, Nishida S (2014) Audiovisual integration in the human perception of materials. J Vis 14:12.

Gingras G, Rowland BA, Stein BE (2009) The differing impact of multisensory and unisensory integration on behavior. J Neurosci 29:4897–4902.

Gleiss S, Kayser C (2012) Audio-visual detection benefits in the rat. PLoS One 7:e45677.

Gur M, Beylin A, Snodderly DM (1997) Response variability of neurons in primary visual cortex (V1) of alert monkeys. J Neurosci 17:2914–2920.

Hammond-Kenny A, Bajo VM, King AJ, Nodal FR (2017) Behavioural benefits of multisensory processing in ferrets. Eur J Neurosci 45:278–289.

Hwang J, Romanski L (2015) Prefrontal neuronal responses during audiovisual mnemonic processing. J Neurosci 35:960–971.

Ibrahim LA, Mesik L, Ji XY, Fang Q, Li HF, Li YT, Zingg B, Zhang LI, Tao HW (2016) Cross-modality sharpening of visual cortical processing through layer-1-mediated inhibition and disinhibition. Neuron 89:1031–1045.

Iurilli G, Ghezzi D, Olcese U, Lassi G, Nazzaro C, Tonini R, Tucci V, Benfenati F, Medini P (2012) Sound-driven synaptic inhibition in primary visual cortex. Neuron 73:814–828.

Knöpfel T, Sweeney Y, Radulescu CI, Zabouri N, Doostdar N, Clopath C, Barnes SJ (2019) Audio-visual experience strengthens multisensory assemblies in adult mouse visual cortex. Nat Commun 10:5684. https://doi.org/10.1038/s41467-019-13607-2

Lippert M, Logothetis NK, Kayser C (2007) Improvement of visual contrast detection by a simultaneous sound. Brain Res 1173:102–109.

Maddox RK, Atilgan H, Bizley JK, Lee AK (2015) Auditory selective attention is enhanced by a task-irrelevant temporally coherent visual stimulus in human listeners. Elife 4:e04995.

Marlin BJ, Mitre M, D'amour JA, Chao MV, Froemke RC (2015) Oxytocin enables maternal behaviour by balancing cortical inhibition. Nature 520:499–504.

Marucci M, Flumeri GD, Borghini G, Sciaraffa N, Scandola M, Pavone EF, Babiloni F, Betti V, Aricò P (2021) The impact of multisensory integration and perceptual load in virtual reality settings on performance, workload and presence. Sci Rep 11:4831.

Mazurek M, Kager M, Van Hooser SD (2014) Robust quantification of orientation selectivity and direction selectivity. Front Neural Circuits 8:92.

McClure JP Jr, Polack PO (2019) Pure tones modulate the representation of orientation and direction in the primary visual cortex. J Neurophysiol 121:2202–2214.

McGurk H, MacDonald J (1976) Hearing lips and seeing voices. Nature 264:746–748.

Meijer GT, Montijn JS, Pennartz C, Lansink CS (2017) Audiovisual modulation in mouse primary visual cortex depends on cross-modal stimulus configuration and congruency. J Neurosci 37:8783–8796.

Meijer GT, Pie JL, Dolman TL, Pennartz C, Lansink CS (2018) Audiovisual integration enhances stimulus detection performance in mice. Front Behav Neurosci 12:231.

Meijer GT, Mertens P, Pennartz C, Olcese U, Lansink CS (2019) The circuit architecture of cortical multisensory processing: distinct functions jointly operating within a common anatomical network. Prog Neurobiol 174: 1–15.

Mesik L, Huang JJ, Zhang LI, Tao HW (2019) Sensory- and motor-related responses in layer 1 neurons in mouse visual cortex. J Neurosci 39:10060–10070.

Métin C, Godement P, Imbert M (1988) The primary visual cortex in the mouse: receptive field properties and functional organization. Exp Brain Res 69:594–612.

Montijn JS, Vinck M, Pennartz CMA (2014) Population coding in mouse visual cortex: response reliability and dissociability of stimulus tuning and noise correlation. Front Comput Neurosci 8:58.

Morrill RJ, Hasenstaub AR (2018) Visual information present in infragranular layers of mouse auditory cortex. J Neurosci 38:2854–2862.

Musall S, Kaufman MT, Juavinett AL, Gluf S, Churchland AK (2019) Single-trial neural dynamics are dominated by richly varied movements. Nat Neurosci 22:1677–1686.

Nelson A, Schneider DM, Takatoh J, Sakurai K, Wang F, Mooney R (2013) A circuit for motor cortical modulation of auditory cortical activity. J Neurosci 33:14342–14353.

Niell CM, Stryker MP (2008) Highly receptive fields in mouse visual cortex. J Neurosci 28:7520–7536.

Niell CM, Stryker MP (2010) Modulation of visual responses by behavioral state in mouse visual cortex. Neuron 65:472–479.

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

Rochefort NL, Narushima M, Grienberger C, Marandi N, Hill DN, Konnerth A (2011) Development of direction selectivity in mouse cortical neurons. Neuron 71:425–432.

Schneider DM, Mooney R (2018) How movement modulates hearing. Annu Rev Neurosci 41:553–572.

Shams L, Kamitani Y, Shimojo S (2002) Visual illusion induced by sound. Brain Res Cogn Brain Res 14:147–152.
Stanislaw H, Todorov N (1999) Calculation of signal detection theory measures. Behav Res Methods Instrum Comput 31:137–149.
Stein BE, Stanford TR, Rowland BA (2014) Development of multisensory integration from the perspective of the individual neuron. Nat Rev Neurosci 15:520–535.
Stein BE, Stanford TR, Rowland BA (2020) Multisensory integration and the Society for Neuroscience: then and now. J Neurosci 40:3–11.
Stringer C, Pachitariu M, Steinmetz N, Reddy CB, Carandini M, Harris KD (2019) Spontaneous behaviors drive multidimensional brainwide activity. Science 364:255.
Tivadar RI, Gaglianese A, Murray MM (2020) Auditory enhancement of illusory countour perception. Multisens Res 34:1–15.
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. Psychol Aging 31:380–389.
von Trapp G, Buran BN, Sen K, Semple MN, Sanes DH (2016) A decline in response variability improves neural signal detection during auditory task performance. J Neurosci 36:11097–11106.
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.
Williams A, Angeloni C, Geffen M (2023) Sound improves neuronal encoding of visual stimuli in mouse primary visual cortex Dryad Dataset. https://doi.org/10.5061/dryad.sxksn033q
Zhao X, Chen H, Liu X, Cang J (2013) Orientation-selective responses in the mouse lateral geniculate nucleus. J Neurosci 33:12751–12763.