

1
2 **Title:** Sound improves neuronal encoding of visual stimuli in mouse primary visual
3 cortex
4
5
6

7 **Authors:** Aaron M. Williams^{1,2,3}, Christopher F. Angeloni^{1,4}, Maria N. Geffen^{1,2,3}
8
9

10
11 **Affiliations:**

12 ¹Department of Otorhinolaryngology, University of Pennsylvania, Philadelphia, United
13 States

14 ²Department of Neuroscience, University of Pennsylvania, Philadelphia, United States

15 ³Department of Neurology, University of Pennsylvania, Philadelphia, United States

16 ⁴Department of Psychology, University of Pennsylvania, Philadelphia, United States
17
18
19

20 **Corresponding author:**

Maria Geffen
Stemmler Hall G10
3450 Hamilton Walk
Philadelphia, PA 19104
mgeffen@pennmedicine.upenn.edu
21
22
23
24
25
26
27

28 **Keywords:** audiovisual integration, primary visual cortex, stimulus decoding,
29 electrophysiology
30

31 **Abstract**

32 In everyday life, we integrate visual and auditory information in routine tasks such as navigation
33 and communication. While it is known that concurrent sound can improve visual perception, the
34 neuronal correlates of this audiovisual integration are not fully understood. Specifically, it remains
35 unknown whether improvement due to sound of detection and discriminability of visual stimuli is
36 reflected in the neuronal firing patterns in the primary visual cortex (V1). Furthermore,
37 presentation of the sound can induce movement in the subject, but little is understood about
38 whether and how sound-induced movement contributes to V1 neuronal activity. Here, we
39 investigated how sound and movement interact to modulate V1 visual responses in awake, head-
40 fixed mice and whether this interaction improves neuronal encoding of the visual stimulus. We
41 presented visual drifting gratings with and without simultaneous auditory white noise to awake
42 mice while recording mouse movement and V1 neuronal activity. Sound modulated the light-
43 evoked activity of 80% of light-responsive neurons, with 95% of neurons exhibiting increased
44 activity when the auditory stimulus was present. Sound consistently induced movement. However,
45 a generalized linear model revealed that sound and movement had distinct and complementary
46 effects on the neuronal visual responses. Furthermore, decoding of the visual stimulus from the
47 neuronal activity was improved with sound, an effect that persisted even when controlling for
48 movement. These results demonstrate that sound and movement modulate visual responses in
49 complementary ways, resulting in improved neuronal representation of the visual stimulus. This
50 study clarifies the role of movement as a potential confound in neuronal audiovisual responses
51 and expands our knowledge of how multimodal processing is mediated at a neuronal level in the
52 awake brain.

53

54 **Significance statement**

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

63

64

65 **Introduction**

66 Our brains use incoming sensory information to generate a continuous perceptual experience.
67 The neuronal systems underlying sensory perceptions of different modalities interact in a way that
68 often improves perception of the complementary modality (Gingras et al., 2009; Gleiss and
69 Kayser, 2012; Bigelow and Poremba, 2016; Hammond-Kenny et al., 2017; Meijer et al., 2018;
70 Stein et al., 2020). In the audiovisual realm, it is often easiest to understand what someone is
71 saying in a crowded room by additionally relying on visual cues such as lip movement and facial
72 expression (Maddox et al., 2015; Tye-Murray et al., 2016). The McGurk effect and flash-beep
73 illusion are other common perceptual phenomena that demonstrate mutual interactions between
74 the auditory and visual systems (McGurk and MacDonald, 1976; Shams et al. 2002). Despite this
75 current awareness of audiovisual integration at a perceptual level, a detailed understanding of the
76 neuronal codes that mediate this improvement has proved elusive.

77

78

79 Previous studies of neuronal correlates of audiovisual integration found that the primary sensory
80 cortical areas participate in this process (Wang et al., 2008; Ibrahim et al., 2016; Meijer et al.,

81 2019; Deneux et al., 2019). The primary visual cortex (V1) contains neurons whose light-evoked
82 firing rates are modulated by sound, as well as neurons that are responsive to sound alone
83 (Knöpfel et al., 2019). Orientation and directional tuning of individual neurons are also affected by
84 sound. In anesthetized mice, layer 2/3 neurons in V1 exhibited sharpened tuning in the presence
85 of sound (Ibrahim et al., 2016). But another study in awake mice found no average differences in
86 visual tuning curve bandwidth with and without sound (Meijer et al., 2017). These contrasting
87 findings raise the question of whether the multisensory perceptual improvements described above
88 are reflected in individual V1 neurons in the awake brain. Furthermore, awake animals are subject
89 to brain-wide changes in neuronal activity due to stimulus-aligned, uninstructed movements
90 (Musall et al., 2019), a factor yet unaccounted for in most audiovisual studies.
91

92 Sound-induced movement represents a potential confound for audiovisual studies in awake
93 animals because whisking and locomotion modulate neuronal activity in the sensory cortical
94 areas. In V1, movement enhances neuronal visual responses and improves neuronal encoding
95 of the visual scene (Niell and Stryker, 2010; Dardalat and Stryker, 2017). Conversely, in the
96 auditory cortex (AC), locomotion generally suppresses neuronal spontaneous and auditory
97 responses (Nelson et al., 2013; Schneider and Mooney, 2018; Bigelow et al., 2019). Therefore,
98 movement is an important factor in neuronal sensory responses that often correlates with stimulus
99 features.
100

101 Thus, audiovisual integration in V1 may not simply represent afferent information from auditory
102 brain regions, as supported by studies demonstrating that V1 neurons are sensitive to the
103 optogenetic stimulation (Ibrahim et al., 2016) and pharmacologic suppression (Deneux et al.,
104 2019) of AC neurons. Indeed, the modulation of V1 activity may instead be a byproduct of
105 uninstructed sound-induced movements which themselves modulate visual responses (Bimbard
106 et al., 2021). However, because previous studies were either performed in anesthetized subjects
107 (Ibrahim et al., 2016), or trials during which the mouse moved were excluded from analysis
108 (Deneux et al., 2019) or pooled together (Iurilli et al., 2012; Meijer et al., 2017), these alternative
109 explanations have not been quantified. We tested to what extent locomotion contributed to
110 audiovisual integration in V1 by performing extracellular recordings of neuronal activity in V1
111 concurrent with monitoring movement in awake mice presented with audiovisual stimuli. We found
112 that the majority of neurons in V1 were responsive to visual and auditory stimuli. We found that
113 sound and movement exerted distinct yet complementary effects on shaping the visual
114 responses. Importantly, sound improved discriminability of the visual stimuli both in individual
115 neurons and at a population level, an effect that persisted when accounting for movement.
116
117

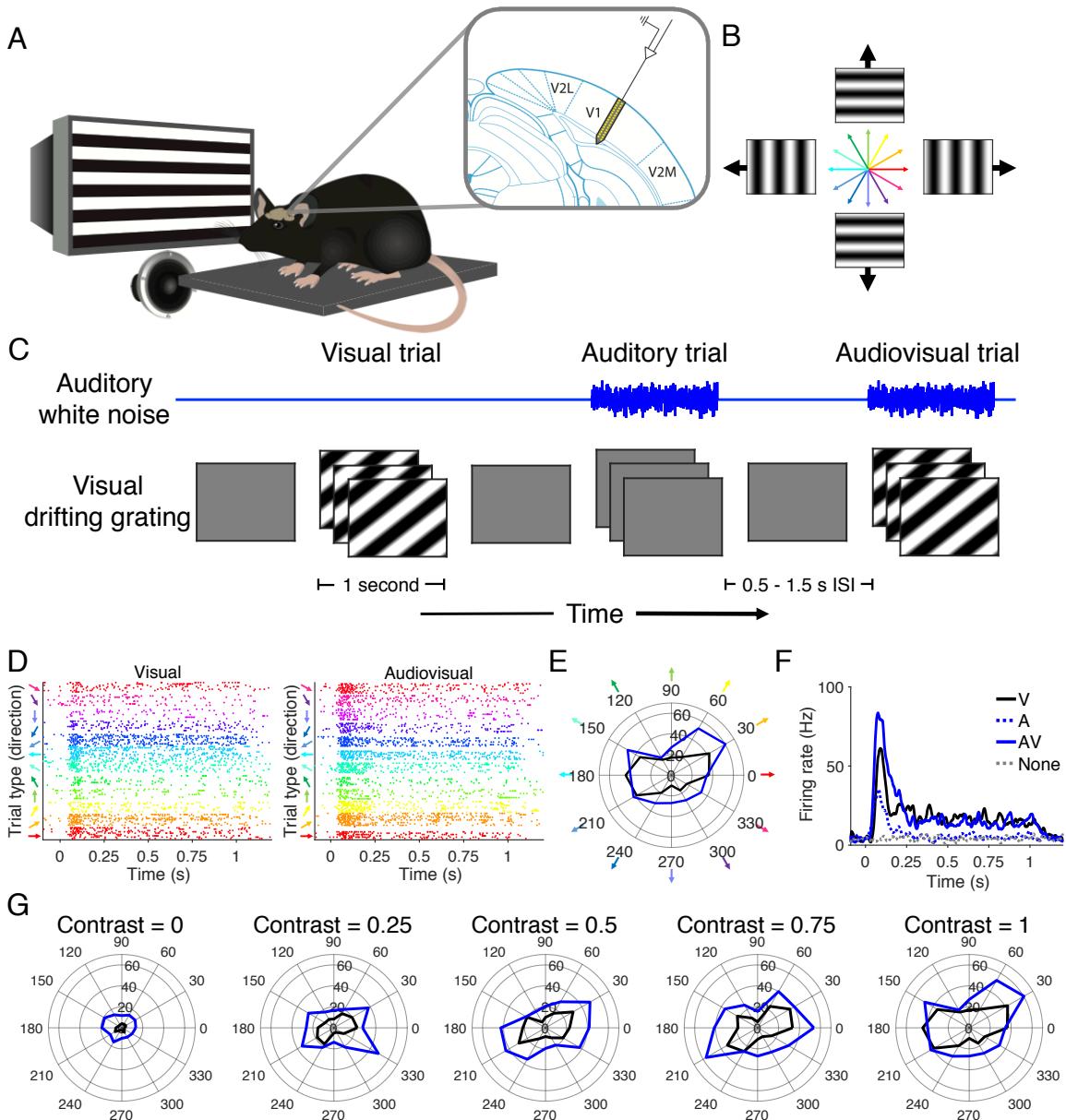
118 Results

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

121 Previous work identified that sound modulates visual responses in V1 (Ibrahim et al., 2016; Meijer
122 et al., 2018; McClure and Polack, 2019), yet how that interaction affects stimulus encoding in
123 individual neurons and as a population remains unclear. Furthermore, whether that interaction
124 can be exclusively attributed to sound or rather to sound-induced motion is controversial (Bimbard
125 et al., 2021). To elucidate the principles underlying audiovisual integration, we presented
126 audiovisual stimuli to awake mice while performing extracellular recordings in V1 (Figure 1A). The
127 visual stimulus consisted of drifting gratings in 12 directions presented at 5 visual contrast levels
128 (Figure 1B). On half of the trials, we paired the visual stimulus with a 70 dB burst of white noise
129 from a speaker positioned next to the screen (Figure 1C), affording 10 trials of each unique
130 audiovisual stimulus condition (Figure 1C). Twelve recording sessions across six mice were spike
131 sorted, and the responses of these sorted neurons were organized by trial type to compare across

132 audiovisual stimulus conditions. Figure 1D-G demonstrates an example unit tuned for gratings
133 aligned to the 30°-210° axis whose baseline and light-evoked firing rate are increased by the
134 sound.

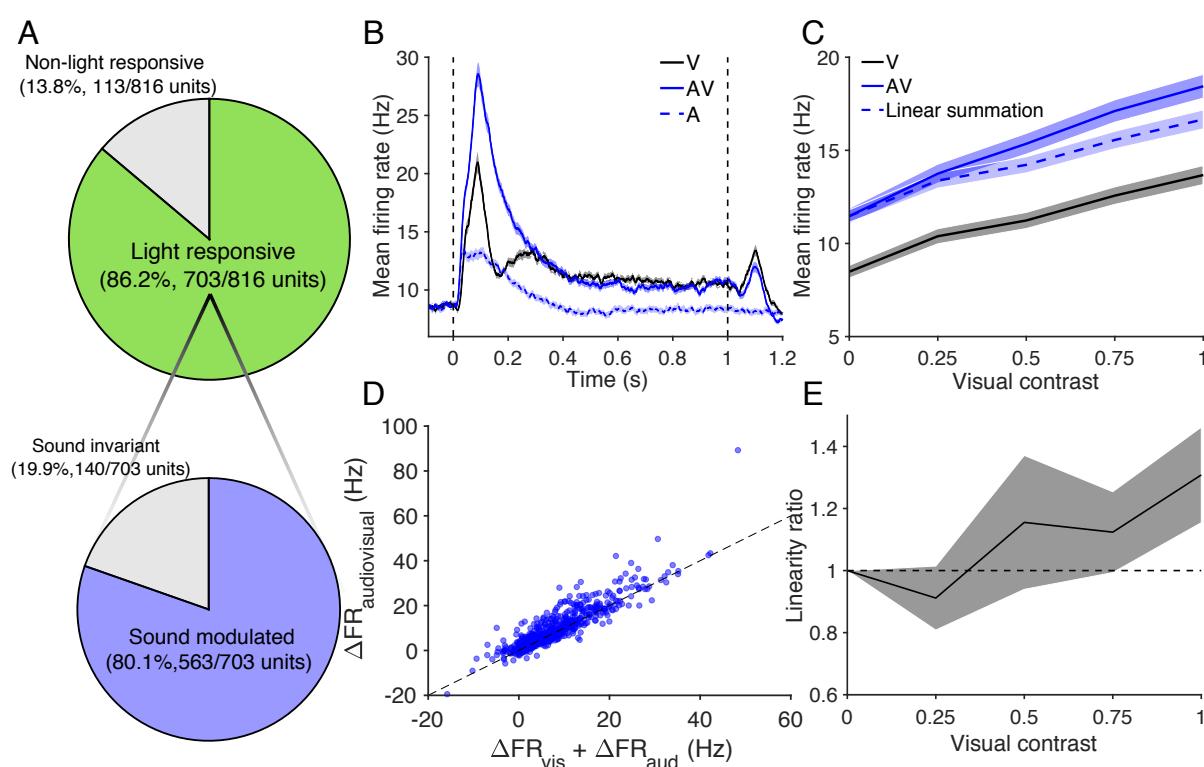
135
136
137



138
139
140
141
142
143
144
145
146

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

147 Sound modulated the activity of the majority of V1 neurons. We used a generalized linear model
148 (GLM) to classify neurons as light-responsive and/or sound-responsive based on their firing rate
149 at the onset (0-300 ms) of each trial. Using this classification method, we found that 86.2%
150 (703/816) of units were responsive to increasing visual stimulus contrast levels, and of these
151 visually responsive units, 80.1% (563/703 neurons, 12 recording sessions in 6 mice) were
152 significantly modulated by the presence of sound (Figure 2A). We constructed an average PSTH
153 from the response profiles of sound-modulated light-responsive neurons, which revealed that the
154 largest change in light-evoked firing rate occurs at the onset of the stimulus (Figure 2B). Averaged
155 across neurons, we found a robust increase in the magnitude of the visually evoked response
156 across visual contrast levels (Figure 2C; $p(\text{vis})=1.2\text{e-}100$, $p(\text{aud})=1.6\text{e-}88$, $p(\text{interact})=5.7\text{e-}4$,
157 paired 2-way ANOVA; $p_{c=0}=2.1\text{e-}51$, $p_{c=0.25}=2.6\text{e-}62$, $p_{c=0.5}=5.7\text{e-}75$, $p_{c=0.75}=1.1\text{e-}81$, $p_{c=1}=2.0\text{e-}81$,
158 post hoc Bonferroni-corrected paired t-test, Table 1). This difference was driven by the majority
159 of neurons (95%) that increased their firing rate in the presence of sound. However, some neurons
160 exhibited lower light-evoked and sound-evoked firing rates relative to baseline.
161
162
163



164
165 **Figure 2 | Sound enhances visual responses in a supra-linear manner** (A) Sound modulates visually evoked activity
166 in 80.1% of light-responsive neurons in V1. (B) Comparison of visual, auditory, and audiovisual PSTHs averaged across
167 all light-responsive sound-modulated neurons. Visual and audiovisual PSTHs correspond to the highest visual
168 contrast level. (C) The magnitude of audiovisual onset responses (0-300ms) is greater than that of the visual response
169 in light-responsive sound-modulated neurons ($n=563$, $p(\text{vis})=1.2\text{e-}100$, $p(\text{aud})=1.6\text{e-}88$, $p(\text{interact})=5.7\text{e-}4$, 2-way
170 repeated measures ANOVA; post hoc Bonferroni-corrected paired t-test). The expected linear sum of the unimodal
171 auditory and visual responses is included. (D) At full visual contrast, the observed audiovisual response in the
172 majority of neurons is greater than the linear sum of the unimodal auditory and visual responses. (E) A linearity ratio
173 above 1 demonstrates audiovisual responses in V1 represent supra-linear integration of the unimodal signals ($n=563$,
174 $p=1.6\text{e-}12$, Kruskal-Wallis test, post hoc Bonferroni-corrected Wilcoxon signed rank test).
175

176 This change in firing rate can be described as supra-linear or sub-linear based on whether the
177 audiovisual response is greater or less than, respectively, the sum of the unimodal light-evoked
178 and sound-evoked firing rates. At medium to high visual contrast levels, integration of the
179 audiovisual stimulus was predominantly supra-linear (Figure 2D-E; $p=1.6e-12$, Kruskal-Wallis
180 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
181 Wilcoxon signed rank test, Table 1). In summary, these results show that sound supra-linearly
182 increases the magnitude of the light-evoked response in the majority of V1 neurons.
183

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

185 Having observed sound-induced changes in the magnitude of the visual response, we next
186 assessed whether these changes in magnitude affected neuronal tuning. V1 neurons have
187 receptive fields tuned to a specific visual stimulus orientation and, to a lesser extent, stimulus
188 direction (Métin et al, 1988; Rochefort et al., 2011; Fahey et al., 2019). We first tested whether
189 sound altered tuning preferences of V1 neurons. In light-responsive neurons, we calculated the
190 orientation and direction-selective indices (OSI and DSI) as well as pseudo indices based on
191 random permutations of the trials (see Methods), and classified neurons in which the true indices
192 were >95% of the pseudo indices as “orientation-” or “direction-selective.” Using this stringent
193 selection criterion, we found that 13.9% (78/563) of neurons were orientation-selective, whereas
194 2.1% (12/563) were direction-selective. In these neurons, we observed shifts in the preferred
195 direction from the visual to audiovisual condition (Fig 2 Sup 1A). This shift in visual tuning
196 preference may be due to auditory input, or it may reflect noise in the neuronal responses. To test
197 this, we performed an additional permutation test by repeatedly sampling the visual responses.
198 We found that the resulting distribution of preferred direction shifts resembled the observed
199 distribution under the audiovisual condition (Fig 2 Sup 1B), and the observed mean shift in
200 degrees was within the limits of the sampled distribution (Fig 2 Sup 1C). Therefore, we cannot
201 conclude that the shift in directional tuning preferences is associated with the presence of sound.
202

203 In addition to testing a shift in preferred direction, we investigated whether sound altered the
204 neurons’ tuning selectivity. Tuning selectivity captures how strongly an individual neuron responds
205 to stimuli of a certain condition, e.g. grating orientation and drift direction, as compared to others.
206 We found a small reduction in the OSI from the visual to audiovisual conditions (Fig 2 Sup 1D-E;
207 $p=0.0018$, paired Student’s t-test), which may reflect disproportionate changes in firing rate at the
208 preferred versus orthogonal directions. We also found a reduction in the DSI in the presence of
209 sound (Fig 2 Sup 1F-G; $p=0.021$, paired Student’s t-test). Combined, these results suggest that
210 sound’s enhancement of the magnitude of light-evoked responses has minimal or potentially
211 diminishing effects on the tuning selectivity of neurons.
212

213 **Sound reduces the latency, increases onset duration, and decreases variability of visual 214 responses in neurons**

215 Behaviorally, certain cross-modal stimuli elicit shorter reaction times than their unimodal
216 counterparts (Diederich and Colonius, 2004; Colonius and Diederich, 2017; Meijer et al., 2018).
217 Therefore, we hypothesized that sound reduces the latency of the light-evoked response at a
218 neuronal level as well. For each neuron, we calculated the response latency as the first time bin
219 after stimulus onset at which the firing rate exceeded 1 standard deviation above baseline (Fig 2
220 Sup 2A), and found that sound reduced the response latency across contrast levels (Fig 2 Sup
221 2B; $p(\text{vis})=6.9e-4$, $p(\text{aud})=6.8e-15$, $p(\text{interact})=0.045$, paired 2-way ANOVA; $p_{c=0.25}=2.3e-4$,
222 $p_{c=0.5}=7.1e-12$, $p_{c=0.75}=4.6e-5$, $p_{c=1}=9.9e-4$, post hoc Bonferroni-corrected paired t-test, Table 1).
223 We additionally calculated the slope of the onset response of light-responsive sound-modulated
224 neurons, measured from trial onset until the time at which each neuron achieved its peak firing
225 rate (Fig 2 Sup 2C). We found that sound increased the slope of the onset response (Fig 2 Sup
226 2D; $p(\text{vis})=3.5e-121$, $p(\text{aud})=2.7e-15$, $p(\text{interact})=0.038$, paired 2-way ANOVA; $p_{c=0.25}=1.4e-4$,

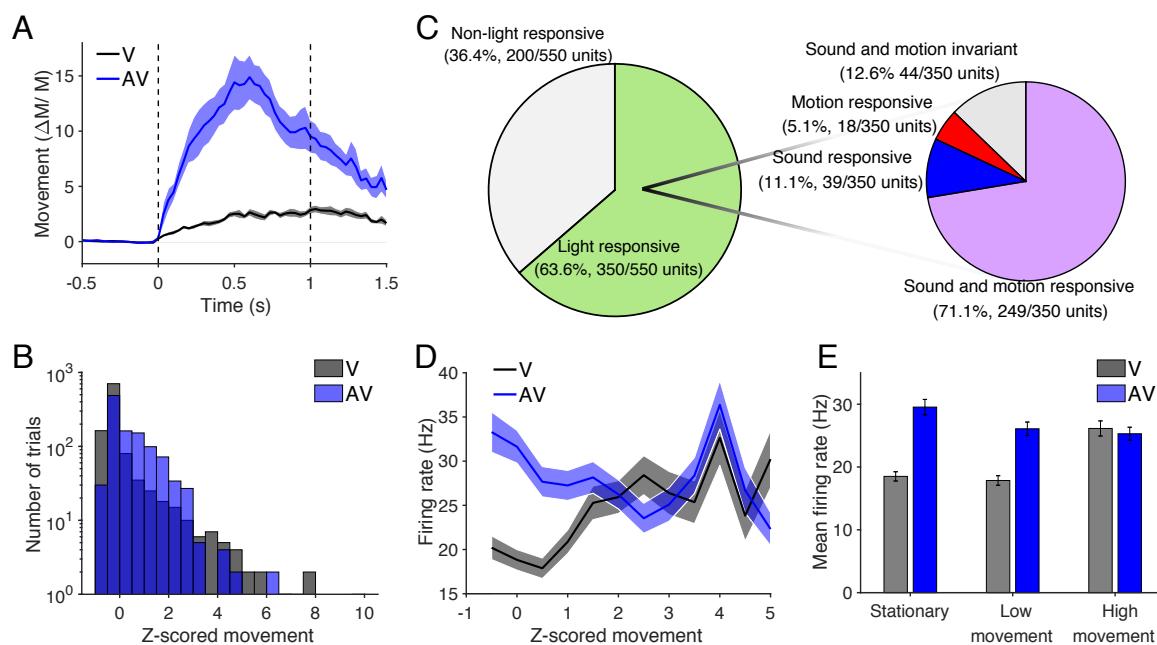
227 $p_{c=0.5}=8.9e-13$, $p_{c=0.75}=3.6e-12$, $p_{c=1}=5.5e-8$, post hoc Bonferroni-corrected paired t-test, Table 1),
228 both indicating that the response latency was reduced in the audiovisual condition compared to
229 the visual condition. Additionally, the duration of the light-evoked response, defined as the full
230 width at half maximum of the peak onset firing rate, increased in the presence of sound (Fig 2
231 Sup 2E,F; $p(\text{vis})=1.3e-10$, $p(\text{aud})=8.7e-98$, $p(\text{interact})=0.23$, paired 2-way ANOVA). Both of these
232 timing effects were relatively constant across contrast levels. Therefore, the latency and onset
233 duration of light-evoked responses in V1 neurons is enhanced by sound.
234

235 Having observed changes in response magnitude and timing, we next investigated the effect of
236 sound on the variability of light-evoked responses. If individual neurons encode the visual stimulus
237 using changes in their firing rate, a more consistent response would entail less spread in the
238 response magnitude relative to the mean response across trials of a single stimulus type. We
239 quantified this relationship using the coefficient of variation (CV) defined as the ratio of the
240 standard deviation to the response mean (Gur et al., 1997). We hypothesized that sound reduces
241 the CV of light-evoked responses, corresponding to reduced response variability and higher SNR.
242 Fig 2 Sup 2G depicts the relationship between response magnitude and CV in an example sound-
243 modulated light-responsive neuron, demonstrating that increased response magnitude correlates
244 with reduced CV. Consistent with sound increasing the visual response magnitude in the majority
245 of sound-modulated light-responsive neurons (Figure 2), we observed a reduction of CV in the
246 audiovisual condition relative to the visual condition when averaged across these neurons (Fig 2
247 Sup 2H; $p(\text{vis})=0.28$, $p(\text{aud})=4.2e-103$, $p(\text{interact})=0.38$, paired 2-way ANOVA). Taken together,
248 these results indicate that sound not only modulates the magnitude of the visual response (Figure
249 2), but also improves the timing and consistency of individual neurons' responses (Fig 2 Sup 2).
250

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

252 It is known that whisking and locomotive behaviors modulate neuronal activity in mouse visual
253 cortex (Niell and Stryker, 2010) and auditory cortex (Nelson et al., 2013; Schneider and Mooney,
254 2018; Bigelow et al., 2019). Therefore, having established that sound robustly modulates visual
255 responses (Figure 2), we tested whether these observed changes were more accurately
256 attributable to sound-induced movement. In an additional cohort of mice, we performed V1
257 extracellular recordings with the same audiovisual stimuli described above while recording
258 movement activity of the mice throughout stimulus presentation. We found that sound did evoke
259 whisking and locomotive behavior in mice, leading to increased movement on audiovisual trials
260 compared to visual trials (Figure 3A; $p=9.1e-5$, paired t-test). However, there were many visual
261 trials in which substantial movement occurred, as well as audiovisual trials in which little
262 movement was detected (Figure 3B). Because of this large variability in sound-induced
263 movement, we were able to control for movement when comparing visual and audiovisual activity
264 in the recorded neurons.
265

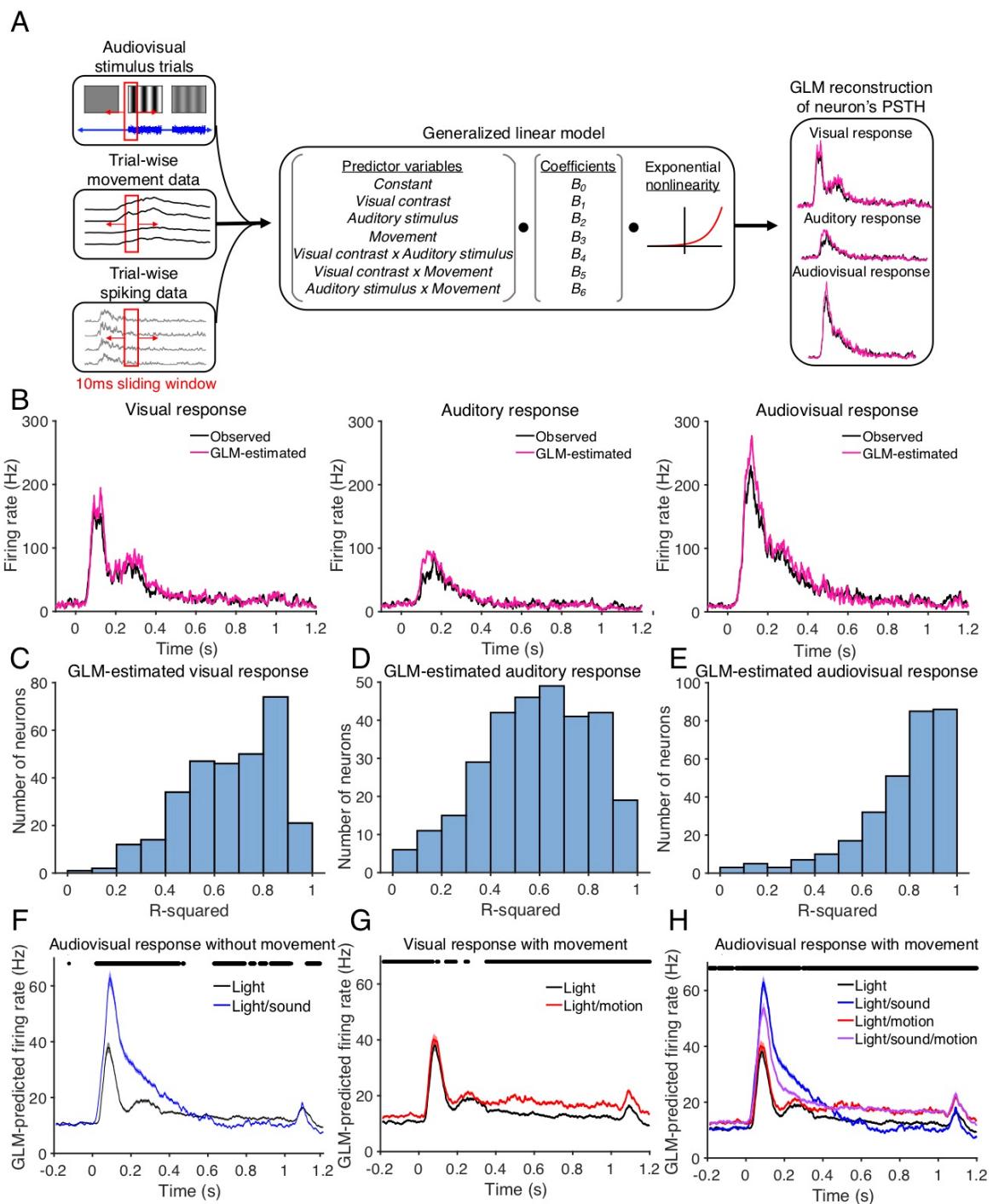
266 Similar to above, we used a GLM to classify each neuron as light-, sound-, and/or motion-
267 responsive based on the neuron's firing rate and mouse's movement activity during the onset (0-
268 300ms) of the trial. The vast majority of light-responsive neurons, 71.1% (249/350), displayed
269 both sound- and motion-modulated visual responses (Figure 3C). 11.1% (39/350) and 5.2%
270 (18/350) of light-responsive neurons were purely sound- or motion-modulated, respectively. An
271 additional 12.6% (44/350) were invariant to sound or motion. We then compared the visually and
272 audiovisually evoked firing rates of neurons when controlling for movement. Among sound- and
273 motion-modulated light-responsive neurons, the firing rate was higher on audiovisual trials than
274 visual trials when movement was held constant (Figure 3D), especially when mice showed limited
275 movement.
276



277
278 **Figure 3 | Sound modulates visual activity when controlling for stimulus-induced movement** (A) Mice displayed
279 more movement response to audiovisual trials than in visual trials ($n=9$ recording sessions; $p=9.1e-5$, paired t-test).
280 (B) Histogram of trials' z-scored movements show a range of levels of movement during both visual and audiovisual
281 trials. (C) Venn diagram demonstrating that 87% of light-responsive neurons exhibited some combination of sound-
282 and movement-responsiveness. (D) Comparison of firing rate of sound- and motion-modulated light-responsive
283 neurons across trials with a range of z-scored movement. (E) Responses to audiovisual stimuli evoke larger
284 magnitude responses than visual stimuli when mice were stationary ($z\text{-score} < -0.5$) or displayed low to moderate
285 movement ($-0.5 < z\text{-score} < 1.5$), but responses were not significantly different when mice displayed the highest
286 amount of movement ($z\text{-score} > 1.5$; $p(\text{motion})=0.001$, $p(\text{aud})=1.4e-13$, $p(\text{interact})=1.8e-8$, 2-way ANOVA, post hoc
287 Bonferroni-corrected two-sample t-test)

288
289 On trials in which the mice were largely stationary ($z\text{-score} < -0.5$, 43% of visual trials, 32% of
290 audiovisual trials) or displayed moderate levels of movement ($-0.5 < z\text{-score} < 1.5$, 51% of visual
291 trials, 57% of audiovisual trials), the mean firing rate of neurons was 54–62% higher when sound
292 was presented than when sound was absent. The firing rates under the two stimulus conditions
293 converged on trials in which the mice displayed high movement activity ($z\text{-score} > 1.5$, 4.8% of
294 visual trials, 11% of audiovisual trials; Figure 3D,E; $p(\text{move})=0.010$, $p(\text{aud})=1.4e-13$,
295 $p(\text{interact})=1.8e-8$, unbalanced 2-way ANOVA; $p_{\text{stationary}}=1.5e-14$, $p_{\text{low motion}}=7.1e-10$, $p_{\text{high motion}}=0.6$,
296 post hoc Bonferroni-corrected two-sample t-test, Table 1). Notably, increasing movement activity
297 was correlated with increased firing rates on visual trials, but was correlated with decreasing firing
298 rates among audiovisual trials (Figure 3E). These results indicate that sound modulated visually
299 evoked neuronal activity even when accounting for sound-induced movement in awake mice.
300

301
302 **Sound and movement have distinct and complementary effects on visual responses**
303 To further parse out the role of sound and movement on audiovisual responses, we used a
304 separate GLM to capture the time course of these parameters' effects on visual activity. For each
305 neuron, we used a GLM with a sliding 10ms window to reconstruct the PSTH based on the visual
306 contrast level, sound presence, and movement during that time window (Figure 4A). Figure 4B
307 shows an example neuron in which the GLM accurately captures the light-evoked, sound-evoked,
308 and audiovisually evoked PSTHs using the average movement for each trial type. Across
309 neurons, the GLM-estimated PSTHs accurately reconstructed observed PSTHs, with the highest



310
311
312
313
314
315
316
317
318
319

Figure 4 | 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 to audiovisual stimuli with and without movement. (B) Observed trial-averaged PSTHs for visual-only (left), auditory-only (middle), and audiovisual (right) trials overlaid with GLM estimates based on the selected stimulus features. (C-E) Histograms demonstrating R^2 values of the GLM-estimated PSTHs, averaged across sound- and motion-modulated light-responsive neurons. Moderate to high R^2 values across the population indicate a good ability for the GLM to estimate neuronal firing rates. (F-H) GLM-predicted visually evoked PSTHs with and without sound and motion. Asterisks indicate time windows in which there was a significant difference between the *light* prediction and the

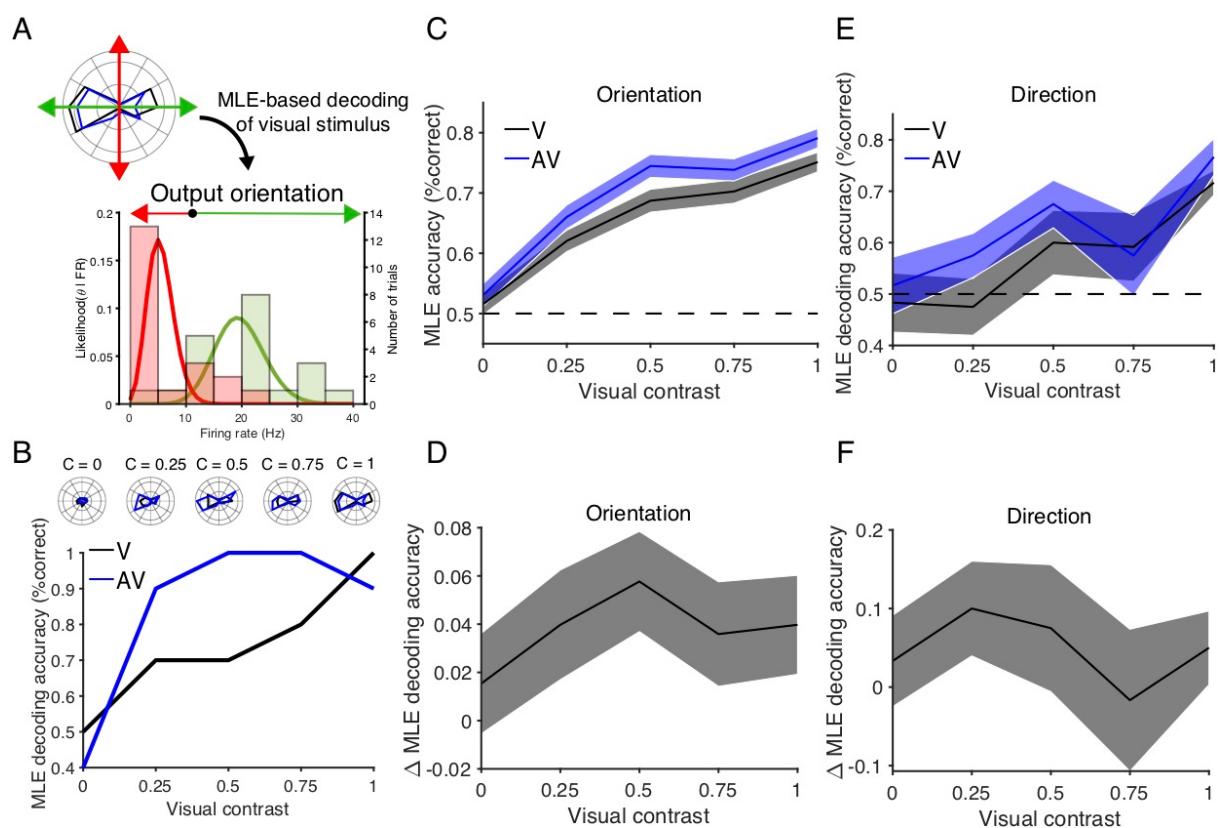
320 *light+sound*, *light+motion*, and *light+sound+motion* predictions, respectively. (F) Excluding motion highlights that
321 sound primarily enhances the onset response. Asterisks indicate time windows in which there was a significant
322 difference ($n=295$ fitted neurons; paired t-test, $\alpha=3.6e-5$). (G) Excluding sound highlights that motion primarily
323 enhances the sustained portion of the response. Asterisks indicate time windows in which there was a significant
324 difference ($n=295$ fitted neurons; paired t-test, $\alpha=3.6e-5$). (H) Sound and motion together enhance both the onset
325 and sustained periods of the visually evoked response. ($n=295$ fitted neurons; paired t-test, $\alpha=3.6e-5$).

326
327

328 correlation when all parameters were included in the estimate (Figure 4C-E). We leveraged the
329 coefficients fit to each neuron (Figure 4A) to estimate the unique contribution of each predictor to
330 the firing rates as a function of time (see Materials and Methods). In the absence of movement,
331 sound predominantly enhanced neuronal activity at the onset of the visual response and
332 suppressed activity during the response's sustained period (Figure 4F; $n=295$ fitted neurons,
333 paired t-test at each time window [1391], $\alpha=3.6e-5$). Conversely, movement had little effect on
334 the onset activity in the absence of sound, but rather enhanced firing rates during the response's
335 sustained period (Figure 4G; $n=295$ fitted neurons, paired t-test at each time window [1391],
336 $\alpha=3.6e-5$). Together, sound and movement have complementary effects in which both the onset
337 and sustained portions of the visual response are enhanced (Figure 4H; $n=295$ fitted neurons,
338 paired t-test at each time window [1391], $\alpha=3.6e-5$). Again notably, the peak onset response
339 under the audiovisual condition was lower when movement was included in the estimate (Figure
340 4H). These findings indicate not only that movement is unable to account for the changes in onset
341 response reported above, but also that sound and motion have distinct and complementary
342 effects on the time course of visually evoked activity in V1.
343

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

345 Behaviorally, sound can improve the detection and discriminability of visual responses, however
346 whether that improved visual acuity is reflected in V1 audiovisual responses is unknown. Despite
347 many studies reporting neuronal correlates of audiovisual integration in V1, whether sound
348 improves neuronal encoding of the visual stimulus has yet to be demonstrated. The increase in
349 response magnitude and decrease in CV suggest that sound may improve visual stimulus
350 discriminability in individual V1 neurons. Consistent with these changes in response magnitude
351 and variability, we observed sound-induced improvements in the d' sensitivity index between
352 responses to low contrast drifting grating directions among orientation- and direction-selective
353 neurons (Fig 5 Sup 1), further indicating improved orientation and directional discriminability in
354 individual neurons. To directly test this hypothesis, we used the neuronal responses of individual
355 neurons to estimate the visual stimulus drifting grating orientation and direction. We trained a
356 maximum likelihood estimate (MLE)-based decoder (Montijn et al., 2014; Meijer et al., 2017) on
357 trials from the preferred and orthogonal orientations in orientation-selective neurons and on trials
358 from the preferred and anti-preferred directions in direction-selective neurons. We used leave-
359 one-out cross-validation and cycled the probe trial through the repeated trials of the stimulus
360 condition in order to calculate the mean decoding performance. The MLE decoder's output was
361 the orientation or direction with the maximum posterior likelihood based on the training data
362 (Figure 5A). This decoding technique achieves high decoding accuracy (Figure 5B). When
363 averaged across sound-modulated orientation-selective neurons, decoding performance was
364 improved on audiovisual trials compared to visual trials (Figure 5C; $p(\text{vis})=4.8e-112$, $p(\text{aud})=7.8e-4$,
365 $p(\text{interact})=0.71$, paired 2-way ANOVA), with the greatest improvements at low to intermediate
366 contrast levels (Figure 5D). We applied this approach to sound-modulated direction-selective
367 units and found similar trends towards improvements at low contrast levels (Figure 5E,F;
368 $p(\text{vis})=2.1e-4$, $p(\text{aud})=0.18$, $p(\text{interact})=0.78$, paired 2-way ANOVA), limited by fewer and weaker
369 direction-selective neurons in V1. These results demonstrate that sound-induced changes in

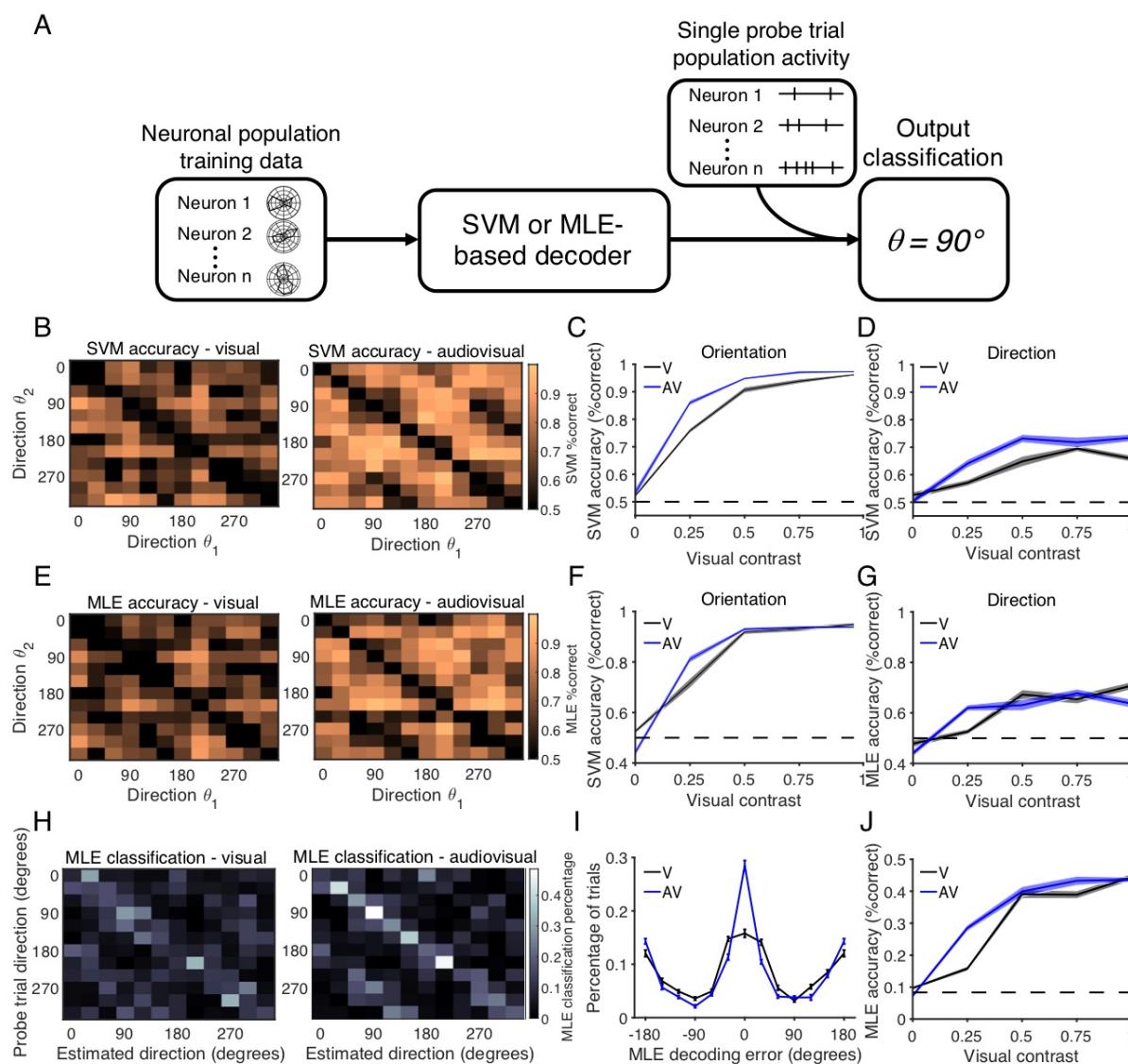


370
371 **Figure 5 | Sound improves decoding of drifting grating direction and orientation in individual neurons** (A) Diagram
372 illustrating MLE-based decoding of an individual neuron's preferred versus orthogonal orientations. (B) Performance
373 of the MLE decoder, trained on an example orientation-selective neuron, in decoding the neuron's preferred versus
374 orthogonal orientations. The neuron's polar plots are shown in the above inset. (C-D) Absolute (C) and difference (D)
375 in decoding accuracy of preferred versus orthogonal orientations, averaged across sound-modulated orientation-
376 selective neurons, demonstrating higher performance in the audiovisual condition ($n=78$, $p(\text{vis})=4.8e-12$,
377 $p(\text{aud})=7.8e-4$, $p(\text{interact})=0.71$, paired 2-way ANOVA). (E-F) Absolute (E) and difference (F) in decoding accuracy of
378 preferred versus anti-preferred directions, averaged across sound-modulated direction-selective neurons. No
379 significant effect of sound on decoding accuracy was observed ($n=12$, $p(\text{vis})=2.1e-4$, $p(\text{aud})=0.18$, $p(\text{interact})=0.78$,
380 paired 2-way ANOVA).

381
382
383 response magnitude and consistency interact in order to improve neuronal representation of the
384 visual stimulus in individual neurons.

385 386 **Population-based decoding of the visual stimulus improves with sound**

387 V1 uses population coding to relay information about the various stimulus dimensions to
388 downstream visual areas (Montijn et al., 2014, Berens et al., 2012), so we next tested whether
389 these improvements in visual stimulus encoding in individual neurons extended to the population
390 level. We began by training a support vector machine (SVM) to perform pairwise classification of
391 visual drifting grating directions based on neuronal population activity. We again used a leave-
392 one-out cross-validation approach when training and testing the SVM (Figure 6A). Decoding
393 accuracy improved as more neurons were included in the population (Fig 6 Sup 1A), achieving
394 an accuracy of ~90% when averaged across all pairwise orientation comparisons. At full visual
395 contrast, there was little difference between the performance on visual and audiovisual trials.
396



397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415

Figure 6 | Sound improves accuracy of population-based visual stimulus decoding (A) Schematic illustrating the decoding of the drifting grating direction using either an SVM or MLE decoder trained on neuronal population activity. (B) Accuracy of SVM pairwise classification of drifting grating directions on visual (left) and audiovisual (right) trials, contrast 0.25. (C) SVM decoding accuracy improved with sound when classifying orthogonal drifting grating orientations ($n=10$ randomizations, $p(\text{vis})=1.8e-61$, $p(\text{aud})=1.9e-8$, $p(\text{interact})=2.4e-4$, 2-way ANOVA, post hoc Bonferroni-corrected paired t-test). (D) SVM decoding accuracy when classifying opposite drifting grating directions, demonstrating improved performance with sound ($n=10$ randomizations, $p(\text{vis})=1.1e-21$, $p(\text{aud})=9.0e-9$, $p(\text{interact})=0.0019$, 2-way ANOVA, post hoc Bonferroni-corrected paired t-test). (E) Accuracy of MLE pairwise classification of drifting gratings on visual (left) and audiovisual (right) trials, contrast 0.25. (F) MLE decoding accuracy when classifying orthogonal drifting grating orientations improved with sound ($n=10$ randomizations, $p(\text{vis})=2.3e-66$, $p(\text{aud})=0.61$, $p(\text{interact})=9.6e-11$, 2-way ANOVA, post hoc Bonferroni-corrected paired t-test). (G) MLE decoding accuracy when classifying opposite drifting grating directions, demonstrating less effect of sound on performance ($n=10$ randomizations, $p(\text{vis})=4.6e-26$, $p(\text{aud})=0.51$, $p(\text{interact})=4.1e-6$, 2-way ANOVA, post hoc Bonferroni-corrected paired t-test). (H) Heat map of actual vs MLE-output directions under visual (left) and audiovisual (right) trials, contrast 0.25. MLE decoder could choose between all 12 drifting grating directions. (I) MLE decoder classification percentage, comparing estimated direction to actual direction. (J) Overall decoding accuracy of MLE decoder when choosing between all 12 drifting grating directions improved with sound ($n=20$ randomizations, $p(\text{vis})=2.2e-92$, $p(\text{aud})=1.9e-5$, $p(\text{interact})=2.7e-11$, 2-way ANOVA, post hoc Bonferroni-corrected paired t-test).

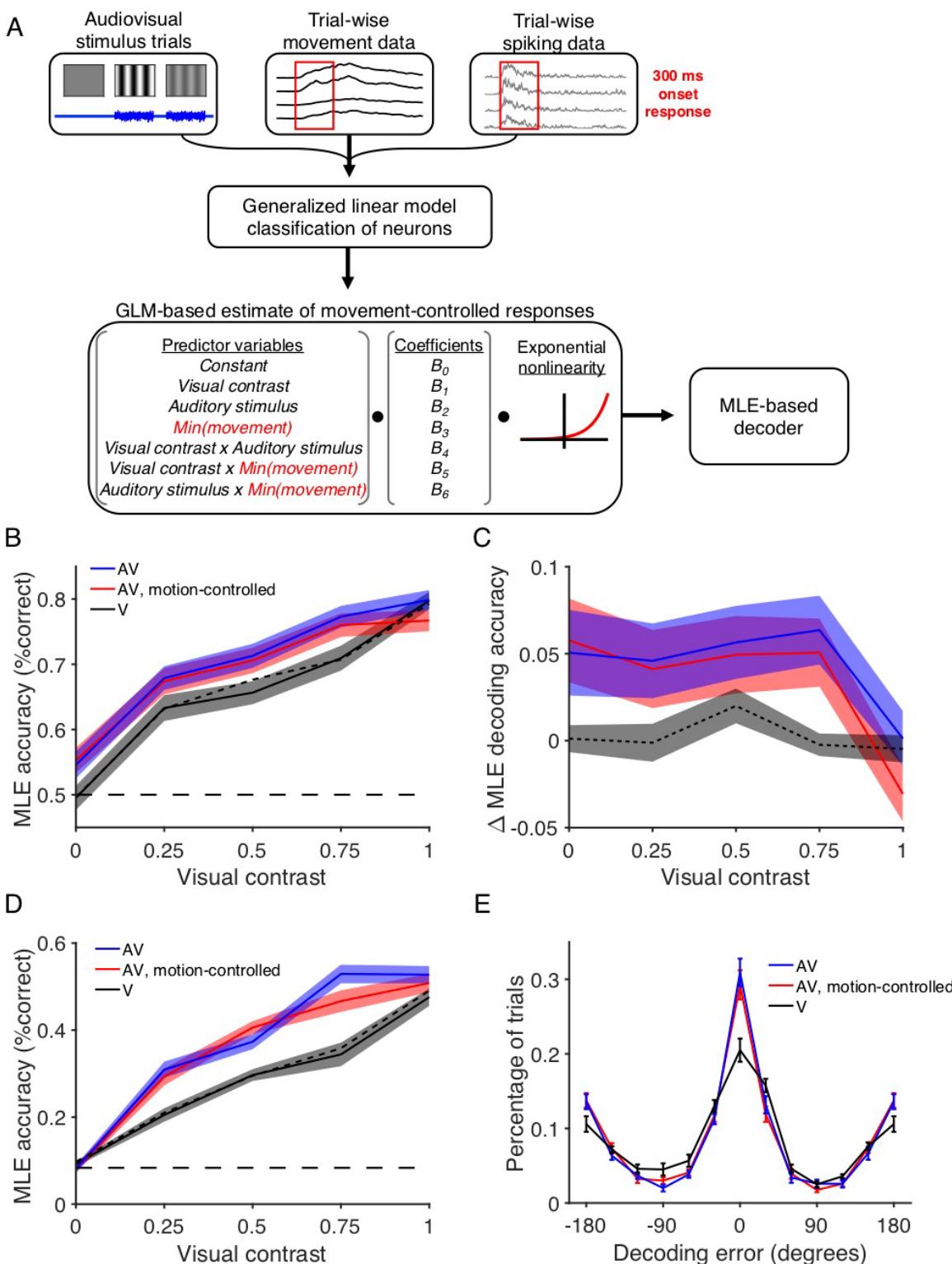
416 However, at low to intermediate visual contrast levels, classification performance robustly
417 increased on audiovisual trials as compared to visual trials (Figure 6B). This improvement in
418 performance was greatest when comparing orthogonal drifting grating orientations (Figure 6C;
419 $p(\text{vis})=1.8\text{e-}61$, $p(\text{aud})=1.9\text{e-}8$, $p(\text{interact}) = 2.4\text{e-}4$, 2-way ANOVA; $p_{c=0}=0.12$, $p_{c=0.25}=0.0016$,
420 $p_{c=0.5}=0.0014$, $p_{c=0.75}=0.0023$; $p_{c=1}=1$, post hoc Bonferroni-corrected paired t-test, Table 1).
421 However, a similar improvement was also observed in decoding opposite drifting grating
422 directions (Figure 6D, $p(\text{vis})=1.1\text{e-}21$, $p(\text{aud})=9.0\text{e-}9$, $p(\text{interact})=0.0019$, 2-way ANOVA;
423 $p_{c=0}=0.55$, $p_{c=0.25}=5.3\text{e-}5$, $p_{c=0.5}=0.0036$, $p_{c=0.75}=0.17$, $p_{c=1}=0.0036$, post hoc Bonferroni-corrected
424 paired t-test, Table 1). These results indicate that sound improves neuronal population encoding
425 of grating orientation and drift direction.
426

427 Similar performance levels were also observed when decoding drifting grating orientation and
428 direction using an MLE-based population decoder, indicating that the results were not specific to
429 the decoding algorithm. Again, performance improved with increasing population sizes (Fig 6 Sup
430 1B), and accuracy was higher on audiovisual trials than visual trials (Figure 6E-G; orientation:
431 $p(\text{vis})=2.3\text{e-}66$, $p(\text{aud})=0.61$, $p(\text{interact})=9.6\text{e-}11$, 2-way ANOVA; $p_{c=0}=5.8\text{e-}4$, $p_{c=0.25}=1.8\text{e-}4$,
432 $p_{c=0.5}=0.3$, $p_{c=0.75}=0.53$, $p_{c=1}=0.15$, post hoc Bonferroni-corrected paired t-test, Table 1; direction:
433 $p(\text{vis})=4.6\text{e-}26$, $p(\text{aud})=0.51$, $p(\text{interact})=4.1\text{e-}6$, 2-way ANOVA; $p_{c=0}=0.037$, $p_{c=0.25}=6.4\text{e-}6$,
434 $p_{c=0.5}=0.036$, $p_{c=0.75}=0.16$, $p_{c=1}=0.14$, post hoc Bonferroni-corrected paired t-test, Table 1).
435

436 Expanding on the SVM approach, the MLE-based decoder allowed us to perform not only pairwise
437 classification, but also classification of 1 out of all 12 drifting grating directions. When trained and
438 tested in this fashion, MLE decoding performance again improved at low to intermediate contrast
439 levels on audiovisual trials (Figure 6H-I), before reaching asymptotic performance of ~45% at full
440 visual contrast (Figure 6J; $p(\text{vis})=2.2\text{e-}92$, $p(\text{aud})=1.9\text{e-}5$, $p(\text{interact})=2.7\text{e-}11$, 2-way ANOVA;
441 $p_{c=0}=0.012$, $p_{c=0.25}=1.4\text{e-}10$, $p_{c=0.5}=0.48$, $p_{c=0.75}=0.0013$, $p_{c=1}=0.5$, post hoc Bonferroni-corrected
442 paired t-test, Table 1). Taken together, these results indicate that sound improves neuronal
443 encoding of the visual stimulus both in individual neurons and at a population level, especially at
444 intermediate visual contrast levels.
445

446 **Sound improves stimulus decoding when controlling for sound-induced movements**

447 It is known that locomotion improves visual processing in V1 (Dardalat and Stryker, 2017). We
448 next tested whether the sound-induced improvement in visual stimulus representation (Figure 6)
449 was attributable to sound's effect on visual responses or indirectly via sound-induced movement.
450 We observed previously that sound was primarily responsible for enhancing the visual response
451 onset, whereas motion enhanced the sustained portion (Figure 4). We therefore hypothesized
452 that the improvement on MLE decoding performance, based on the visual response onset, would
453 be present even when accounting for sound-induced uninstructed movements. We tested this
454 hypothesis by expanding on the GLM-based classification of neurons described in Figure 3. Using
455 the same GLM generated for each neuron, we modified the movement variable and its
456 corresponding pairwise predictors to the lowest observed value, and then used the GLM
457 coefficients and the exponential nonlinearity to estimate each neuron's audiovisual response
458 magnitude when regressing out the effect of motion (Figure 7A, Materials and Methods). We then
459 input these estimated trial-wise neuronal responses into the same MLE-based decoder described
460 above. Using this approach, we found that in individual orientation-selective neurons, controlling
461 for the effect of motion on audiovisual trials minimally changed the accuracy of the population
462 decoder across contrast levels (Figure 7B-C; $p(\text{vis})=7.7\text{e-}93$, $p(\text{aud})=0.055$, $p(\text{interact})=0.058$,
463 paired 2-way ANOVA, Table 1). However, regressing out both sound and motion from the
464 audiovisual responses resulted in decoding accuracy that resembled that on visual trials (Figure
465 7B-C; $p(\text{vis})=8.1\text{e-}95$, $p(\text{aud}) = 0.55$, $p(\text{interact})=0.24$, paired 2-way ANOVA, Table 1). These
466 results in individual neurons indicate that sound and not movement primarily drives the



467
468
469
470
471
472
473

Figure 7 | Sound improved decoding performance when controlling for motion. (A) Diagram illustrating the use of a GLM to calculate each predictor variable's coefficient. These are then used when varying the predictor variables to estimate trial-wise neuronal responses, which are then into the MLE-based decoder. (B) Absolute accuracy of decoding orientation among orientation-selective, sound/motion-modulated light-responsive neurons, comparing visual responses (black, solid) to audiovisual responses (blue) and audiovisual responses when regressing out motion (red). The finely dotted line represents audiovisual responses when controlling for the effects of both motion and

474 sound. (C) Relative decoding accuracy compared to decoding on visual trials. Regressing out motion did not reduce
475 performance compared to audiovisual trials ($n=85$ neurons, $p(\text{vis})=7.7e-93$, $p(\text{aud})=0.055$, $p(\text{interact})=0.058$, paired
476 2-way ANOVA), whereas regressing out both motion and sound resulted in comparable performance to visual trials
477 ($n=85$ neurons, $p(\text{vis})=8.1e-95$, $p(\text{aud})=0.55$, $p(\text{interact})=0.24$, paired 2-way ANOVA). (D) Population decoding
478 accuracy of population-based decoder on audiovisual trials (blue) is preserved even when controlling for motion
479 (red) compared to decoding of visual trials (black; $n=10$ randomizations, $p(\text{vis}) = 1.4e-38$, $p(\text{aud})=6.0e-8$,
480 $p(\text{interact})=0.0015$, 2-way ANOVA; $p_{c=0}=0.30$, $p_{c=0.25}=0.0012$, $p_{c=0.5}=0.0022$, $p_{c=0.75}=0.0044$, $p_{c=1}=0.35$, Bonferroni-
481 corrected paired t-test). The finely black dotted line represents decoding accuracy when regressing out both sound
482 and motion. (E) MLE decoder classification percentage, comparing estimated direction to actual direction, contrast
483 0.25. Little difference is observed between audiovisual trials and audiovisual trials when controlling for motion,
484 whereas both are more accurate than visual trials.
485
486

487 improvements in decoding accuracy in audiovisual trials. We found similar results when
488 implementing this approach in the MLE-based population decoder. We again found that that
489 decoding performance on audiovisual trials when regressing out motion was still significantly
490 improved compared to that on visual trials (Figure 7D-E; $p(\text{vis})=1.4e-38$, $p(\text{aud})=6.0e-8$,
491 $p(\text{interact})=0.0015$, 2-way ANOVA; $p_{c=0}=0.30$, $p_{c=0.25}=0.0012$, $p_{c=0.5}=0.0022$, $p_{c=0.75}=0.0044$,
492 $p_{c=1}=0.35$, Bonferroni-corrected paired t-test). Furthermore, regression of both sound and
493 movement from audiovisual trials resulted in population decoding performance similar to that on
494 visual trials (Figure 7D-E; $p(\text{vis})=2.5e-39$, $p(\text{aud})=0.48$, $p(\text{interact})=0.99$, 2-way ANOVA). These
495 results demonstrate that at both an individual neuron and population level, sound improves visual
496 stimulus decoding on audiovisual trials even when controlling for sound-induced motion.
497
498

499 Discussion

500
501 Audiovisual integration is an essential aspect of sensory processing (Stein et al., 2020). In
502 humans, audiovisual integration is used in everyday behaviors such as speech perception and
503 object recognition (Fujisaki et al., 2014). In animal models, audiovisual integration improves the
504 detection and discriminability of unisensory auditory and visual stimuli (Gleiss and Kayser, 2012;
505 Meijer et al., 2018). However, the neuronal mechanisms underlying these behavioral
506 improvements are still being revealed. Specifically, it remains unclear how sound-induced
507 changes in neuronal activity affect encoding of the visual stimulus. Furthermore, whether the
508 reported audiovisual integration can more accurately be attributed to sound-induced movement
509 has yet to be studied.
510

511 The goal of the present study was to test the hypothesis that sound improves neuronal encoding
512 of visual stimuli in V1 independent of sound-induced movement. We performed extracellular
513 recordings in V1 while presenting combinations of visual drifting gratings and auditory white noise
514 and recording movement of awake mice. The drifting gratings were presented at a range of visual
515 contrast levels to determine the threshold levels at which sound is most effective. As in previous
516 studies, we found neurons in V1 whose spontaneous and visually evoked firing rates are
517 modulated by sound (Figure 2). Notably, the effects we observed were stronger and more positive
518 than in previous studies (80.1% of neurons were modulated by sound, with ~95% exhibiting
519 sound-induced increases in firing rate). When accounting for movement in awake animal subjects,
520 we found that the neurons' audiovisual responses actually represented a mixed effect of both
521 sound- and movement-sensitivity (Figure 3), an effect in which sound primarily enhances the
522 onset response whereas movement complementarily enhances the sustained response (Figure
523 4). We also found that sound-induced changes in response magnitude and consistency combined
524 to improve the discriminability of drifting grating orientation and direction in individual neurons and

525 at a population level (Figure 5,6). The improvements in neuronal encoding were most pronounced
526 at low to intermediate visual contrast levels, a finding that supports the current understanding that
527 audiovisual integration is most beneficial for behavioral performance under ambiguous
528 unisensory conditions (Gleiss and Kayser, 2012; Meijer et al., 2018; Stein et al., 2020).
529 Importantly, the improvement in neuronal encoding was based on firing at the onset of the visual
530 response, indicating that the auditory signal itself is responsible for improvements in visual
531 encoding and not attributable to uninstructed movements. This was directly demonstrated by the
532 persistence of sound-induced improvements in stimulus decoding, even when controlling for the
533 effect of motion (Figure 7).

534

535 **Auditory and locomotive inputs distinctly shape visual responses**

536 We present the novel finding that sound and movement have distinct and complementary effects
537 on visual response. Specifically, we found that sound primarily enhances the firing rate at the
538 onset of the visual response, whereas motion enhances the firing rate during the sustained period
539 of the visual response (Figure 4F-H). Our initial classification of sound-modulated neurons and
540 the subsequent decoding analyses were based on firing rates during the onset period. Therefore,
541 despite robust differences in movement during visual and audiovisual trials, motion was unable
542 to account for the sound-induced changes in neuronal responses that resulted in improved
543 neuronal encoding (Figure 7). The distinct effects that sound and locomotion have on visual
544 responses also adds nuance to our understanding of how motion affects visual processing, as
545 other groups have predominantly used responses averaged across the duration of the stimulus
546 presentation in categorizing motion responsive neurons in V1 (Neil and Stryker, 2010; Dardalat
547 and Stryker, 2017). Our findings indicate that the timing of cross-sensory interactions is an
548 important factor in the classification and quantification of multisensory effects.

549 We also observed that motion decreases the magnitude of the enhancing effect that sound has
550 on the onset of the visual response (Figure 3E, 4H). This finding suggests a degree of suppressive
551 effect that motion has on this audiovisual interaction. A potential mechanism for this result may
552 relate to the circuits underlying audiovisual integration in V1. Other groups have shown using
553 retrograde tracing, optogenetics and pharmacology that the AC projects directly to V1 and is
554 responsible for the auditory signal in this region (Falchier et al., 2002; Ibrahim et al., 2016; Deneux
555 et al., 2019). It is currently understood that unlike in V1, in other primary sensory cortical areas
556 including the AC movement suppresses sensory evoked activity (Nelson et al., 2013; Schneider
557 and Mooney, 2018; Bigelow et al., 2019). Therefore, one explanation for this observation is that
558 despite motion enhancing the visual response magnitude in the absence of sound, the
559 suppressive effect that motion has on sound-evoked responses in the AC leads to weaker AC
560 enhancement of visual activity on trials in which the mice move. A detailed experimental approach
561 using optogenetics or pharmacology would be required to test this hypothesis of a tripartite
562 interaction and would also reveal the potential contribution of other auditory regions.

563

564 **Enhanced response magnitude and consistency combine to improve neuronal encoding**

565 Signal detection theory indicates that improved encoding can be mediated both by enhanced
566 signal magnitude as well as reduced levels of noise (von Trapp et al., 2016). When using purely
567 magnitude-based metrics of discriminability, OSI and DSI, we found a small reduction from the
568 visual to audiovisual conditions (Fig 2 Sup 1). However, we also observed that sound reduced
569 the CV of visual responses (Fig 2 Sup 2), a measure of the trial-to-trial variability in response.
570 When we measured the d' sensitivity index of neuronal responses, a measure that factors in both
571 the response magnitude and distribution, we found that sound improved the discriminability of
572 drifting grating orientation and direction (Fig 4 Sup 1). These findings indicate that the improved
573 discriminability of visual responses in individual neurons was mediated not only by changes in
574 response magnitude but also by the associated improvement in response consistency between

576 trials. Therefore, it is important to consider response variability in addition to magnitude-based
577 metrics when quantifying tuning and discriminability in neurons (Churchland et al., 2011).
578

579 Prior studies using calcium imaging found equivocal results when investigating whether sound-
580 induced changes in visual responses led to improved population encoding of the visual stimulus
581 (Meijer et al., 2017). The improved discriminability of grating orientation and direction by individual
582 neurons supports our finding that the presence of sound enhances population encoding of the
583 visual stimulus. One explanation for this difference may be the recording modality and analysis
584 parameters. We performed electrophysiological recordings of spiking activity and limited our
585 quantification to the onset of the stimulus (0-300 ms), the time window in which there was the
586 greatest change in firing rate across neurons. Calcium imaging, on the other hand, may lack the
587 temporal resolution required to detect the trial-by-trial differences in spiking activity associated
588 with improved neuronal discriminability. Additionally, extracellular electrophysiology allowed us to
589 take advantage of large numbers of neurons in awake animals to include in the population
590 analysis, as opposed to patch-clamp approaches with a limited number of neurons (Ibrahim et al.,
591 2016). Finally, presenting a wide range of visual contrast levels allowed use to demonstrate that
592 sound improves neuronal encoding at low to intermediate contrasts, above which further
593 improvement is difficult to demonstrate due to already reliable encoding in the absence of sound.
594

595 **Stimulus parameters relevant to audiovisual integration**

596 Sensory neurons are often tuned to specific features of unisensory auditory and visual stimuli,
597 and these features are relevant to cross-sensory integration of the signals. In the current study
598 we paired the visual drifting gratings with a static burst of auditory white noise as a basic well-
599 controlled stimulus. Previous studies found that temporally congruent audiovisual stimuli, e.g.
600 amplitude-modulated sounds accompanying visual drifting gratings, evoke larger changes in
601 response than temporally incongruent stimuli in the mouse visual cortex (Meijer et al., 2017), and
602 therefore using such stimuli would potentially result in even stronger effects than we observed.
603 Auditory pure tones can also induce changes in V1 visual responses (McClure and Polack, 2019).
604 However, in other brain regions such as the inferior colliculus, audiovisual integration is highly
605 dependent on spatial congruency between the unimodal inputs (Bergan and Knudsen, 2009). Our
606 results show that spatially congruent, static white noise is sufficient to improve the neuronal
607 response magnitude and latency to light-evoked response. However, additional studies are
608 needed to explore the full range of auditory stimulus parameters relevant to visual responses in
609 V1. Additionally, visual drifting gratings are often used to evoke robust responses in V1, but it
610 would be valuable to determine whether sound is also capable of modulating responses to
611 looming stimuli and more complex visual patterns as well.
612

613 **Neuronal correlates of multisensory behavior**

614 Our findings of multisensory improvements in neuronal performance are supported by numerous
615 published behavioral studies in humans and various model organisms (Gleiss and Kayser, 2012;
616 Meijer et al., 2018; Stein et al., 2020). Training mice to detect or discriminate audiovisual stimuli
617 allows the generation of psychometric performance curves in the presence and absence of sound.
618 We would hypothesize that the intermediate visual contrast levels in which we see improvements
619 in neural encoding would align with behavioral detection threshold levels. One could also correlate
620 the trial-by-trial neural decoding of the visual stimulus with the behavioral response on a stimulus
621 discriminability task, an analysis that could provide information about the proximity of the V1
622 responses to the behavioral perception and decision. Additionally, a behavioral task could allow
623 the comparison of neural responses between passive and active observing, helping to reveal the
624 role of attention on how informative or distracting one stimulus is about the other.
625
626

627 **Multisensory integration in other systems**

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

639

640 **Acknowledgements**

641 The authors thank Gabrielle Samulewicz for assistance with experiments and members of the
642 Geffen laboratory for helpful discussions, as well as Dr. Jay Gottfried and Dr. Yale Cohen at the
643 University of Pennsylvania. This work was supported by funding by the National Institute on
644 Deafness and Other Communication Disorders at the National Institute of Health grants
645 5T32DC016903 to AMW, F31DC016524 to CFA, and R01DC015527, R01DC014479, and
646 R01NS113241 to MNG.

647

648

649 **Materials and methods**

650

651 **Mice**

652 All experimental procedures were in accordance with NIH guidelines and approved by the IACUC
653 at the University of Pennsylvania. Mice were acquired from Jackson Laboratories (5 male, 6
654 female, aged 10-18 weeks at time of recording; B6.Cast-Cdh23^{Ahl/+} mice [Stock No: 018399]) and
655 were housed at 28°C in a room with a reversed light cycle and food provided ad libitum.
656 Experiments were carried out during the dark period. Mice were housed individually after
657 headplate implantation. Euthanasia was performed using CO₂, consistent with the
658 recommendations of the American Veterinary Medical Association (AVMA) Guidelines on
659 Euthanasia. All procedures were approved by the University of Pennsylvania IACUC and the
660 AALAC Guide on Animal Research. We made every attempt to minimize the number of animals
661 used and to reduce pain or discomfort.

662

663 **Surgical procedures**

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

671

672 **Electrophysiological recordings**

673 All recordings were carried out inside a custom-built acoustic isolation booth. 1-2 weeks following
674 the headplate and ground pin attachment surgery, we habituated the mice to the recording booth
675 for increasing durations (5, 15, 30 minutes) over the course of 3 days. On the day of recording,
676 mice were placed in the recording booth and anesthetized with 2.5% isoflurane. We then

677 performed a small craniotomy above the left primary visual cortex (V1, 2.5mm lateral of midline,
678 0-0.5 mm posterior of the lambdoid suture). Mice were then allowed adequate time to recover
679 from anesthesia. Activity of neurons were recorded using a 32-channel silicon probe (NeuroNexus
680 A1x32-Poly2-5mm-50s-177). The electrode was lowered into the primary visual cortex via a
681 stereotactic instrument to a depth of 775-1000 μ m. Following the audiovisual stimulus
682 presentation, electrophysiological data from all 32 channels were filtered between 600 and 6000
683 Hz, and spikes belonging to single neurons and multi-units were identified in a semi-automated
684 manner using KiloSort2 (Pachitariu et al., 2016).

685

686 **Audiovisual stimuli**

687 The audiovisual stimuli were generated using MATLAB (MathWorks, USA), and presented to mice
688 on a 12" LCD monitor (Eyoyo) and through a magnetic speaker (Tucker-Davis Technologies)
689 placed to the right of the mouse. The visual stimulus was generated using the PsychToolBox
690 package for MATLAB and consisted of square wave drifting gratings 1 s in duration, 4-Hz temporal
691 frequency, and 0.1 cycles/ $^{\circ}$. The gratings moved in 12 directions, evenly spaced 0 $^{\circ}$ -360 $^{\circ}$, and
692 were scaled to a range of 5 different visual contrast levels (0, 0.25, 0.5, 0.75, 1), totaling 60 unique
693 visual stimuli. The auditory stimulus was sampled at 400 kHz and consisted of a 1 s burst of 70
694 dB white noise. The visual grating was accompanied by the auditory noise on half of trials (120
695 unique trial types, 10 repeats each), with simultaneous onset and offset. The auditory-only
696 condition corresponded to the trials with a visual contrast of 0. The trial order was randomized
697 and was different for each recording.

698

699 **Data analysis and statistical procedures**

700 Spiking data from each recorded unit was organized by trial type and aligned to the trial onset.
701 The number of spikes during each trial's first 0-300ms was input into a generalized linear model
702 (GLM; predictor variables: visual contrast [continuous variable 0, 0.25, 0.5, 0.75 ,1], sound [0 or
703 1]; response variable: number of spikes during 0-300ms; Poisson distribution, log link function),
704 allowing the classification of each neuron's responses as having a main effect ($p < 0.05$) of light,
705 sound, and/or a light-sound interaction. Neurons that were responsive to both light and sound or
706 had a significant light-sound interaction term were classified as "light-responsive sound-
707 modulated." To quantify the supra- or sub-linear integration of the auditory and visual responses,
708 we calculated the linearity ratio of neurons' audiovisual responses. This ratio was defined as FR_{AV}
709 / ($FR_V + FR_A$), and the sound-only response FR_A was calculated using the trials with a visual
710 contrast of 0.

711

712 We quantified changes in response timing by calculating response latency, onset slope, and onset
713 response duration. First, mean peristimulus time histograms (PSTH) were constructed for each
714 trial type using a 10 ms sliding window. The latency was calculated as the first time bin after
715 stimulus onset in which the mean firing rate at full contrast exceeded 1 standard deviation above
716 baseline. The slope Hz/ms slope was calculated from the trial onset to the time of the peak
717 absolute value firing rate. The response duration was calculated using the full width at half
718 maximum of the peak firing rate at stimulus onset (limited to 0-300 ms).

719

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

725

$$726 \text{OSI} = \frac{FR_{pref} - FR_{ortho}}{FR_{pref} + FR_{ortho}} \quad DSI = \frac{FR_{pref} - FR_{antipref}}{FR_{pref} + FR_{antipref}}$$

727
728 where FR_{ortho} and $FR_{antipref}$ are the mean firing rates in the orthogonal (90°) and anti-preferred
729 (180°) directions, respectively. One-tailed permutation testing was performed by comparing these
730 OSI and DSI values to pseudo OSI and DSI values obtained by 200 random shuffles of the firing
731 rates from the pooled preferred and orthogonal or anti-preferred trials. If a neuron's actual OSI or
732 DSI value was $>95\%$ of shuffled OSI or DSI values, the neuron was classified as "orientation-" or
733 "direction-selective," respectively. To determine whether there were statistically significant
734 changes in the preferred direction from the visual to audiovisual conditions, we applied a
735 bootstrapping procedure, subsampling the visual trials for each neuron 1000 times and creating
736 a confidence interval of the mean shift in preferred direction (degrees) for each population
737 randomization.
738

739 We assessed and controlled for sound-induced movement as a potential confound for the
740 audiovisual effects observed. During a subset of V1 recordings (9 recordings, 5 mice), mouse
741 movement was tracked throughout stimulus presentation. Video recording was performed using
742 a Raspberry Pi 4 Model B computer system with an 8MP infrared Raspberry Pi NoIR Camera V2
743 attachment. The video was converted to MP4 format, and motion was quantified by calculating
744 the frame-by-frame difference, an approach that captured both whisking and locomotive behavior.
745 This movement value for each recording was then aligned to the trials of the audiovisual stimulus
746 from the recording trials for further analysis.
747

748 Similar to above, a GLM (predictor variables: visual contrast level, sound presence, average
749 motion during each trial; response variable: trial spikes during 0-300ms; Poisson distribution, log
750 link function) classified each neuron as having a main effect ($p < 0.05$) of light, sound, or motion,
751 as well as the pairwise interactions of these parameters. Light-responsive sound-modulated
752 neurons, according to the above definition, that additionally displayed either a main effect of
753 motion or significant light-motion or sound-motion interaction terms were classified as "motion-
754 modulated" and were included for further analysis.
755

756 In order to reconstruct peristimulus time histograms of light-responsive, sound-modulated,
757 motion-modulated neurons, we used a separate GLM. Using a 10ms sliding window across all
758 trials, we input the visual contrast level, sound presence, and motion during that window
759 (discretized into five bins) as predictor variables, and the number of spikes during that window as
760 response variables, into the GLM (Poisson distribution, log link function) to calculate coefficients
761 for light, sound, motion, and their pairwise interactions. This approach allowed us to reconstruct
762 the mean PSTH of individual neurons observed during each trial type by calculating:
763

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

764 where the spikes in time window t are determined by the values p and coefficients c of predictor
765 variable i . From there, we used this same equation to estimate the shape of the PSTHs when
766 varying sound and motion in order to determine differential effects these parameters had on the
767 temporal trajectory of neurons' visual responses.
768

769 The d' sensitivity index (Stanislaw and Todorov, 1999; von Trapp et al., 2016) was used to
770 calculate the directional discriminability of direction-selective neurons. The d' sensitivity index
771 between two directions θ_1 and θ_2 is calculated as:
772
773

774

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

775

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

780

781

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

784

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

785

786 For decoding using individual neurons, the likelihood $P(A_{trial}|\theta)$ for each orientation or direction
787 was computed based on the Poisson response distribution across all trials of that orientation or
788 direction, with a leave-one-out cross-validation technique in which the probe trial (A_{trial}) was
789 excluded from the training data. The prior $P(\theta)$ was uniform, and the normalization term $P(A_{trial})$
790 was similarly applied to all directions. Therefore, the posterior probability $P(\theta|A_{trial})$ was
791 proportional to and based on evaluating the likelihood function at the value of the probe trial. For
792 orientation-selective neurons, decoding was performed between the preferred and orthogonal
793 orientations, and for direction-selective neurons, decoding was performed between the preferred
794 and anti-preferred directions. For decoding using populations of neurons, neurons were pooled
795 across recording sessions. A similar approach was used; however, here, the posterior probability
796 $P(\theta|A_{pop})$ was proportional to the joint likelihood $P(A_{pop}|\theta)$ of the single-trial activity across all N
797 neurons in the population (A_{pop}):

798

799

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

800

801

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

805

806

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

812

813

814

815

816

817

Statistics

Figure data are displayed as means with standard error of the mean (SEM), unless otherwise noted. Shapiro-Wilk tests were used to assess normality, and the statistical tests performed are indicated in the text, figures, and Table 1. For multi-group and multivariate analysis (e.g., ANOVA and Kruskal-Wallis tests) in which a significant ($p<0.05$) interaction was detected, we subsequently performed a post hoc Bonferroni-corrected test. P-values reported as 0 are too

818 small to be accurately calculated by Matlab ($p < 2.2 \times 10^{-301}$), due to characteristically large data sets.
819 See Table 1 for a detailed summary of statistical results and post hoc comparisons.
820

References

- 822
823 Atilgan, H., Town, S. M., Wood, K. C., Jones, G. P., Maddox, R. K., Lee, A., & Bizley, J. K.
824 (2018). Integration of Visual Information in Auditory Cortex Promotes Auditory Scene Analysis
825 through Multisensory Binding. *Neuron*, 97(3), 640–655.e4.
826 <https://doi.org/10.1016/j.neuron.2017.12.034>

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

829 Berens, P., Ecker, A.S., Cotton, R.J., Ma, W.J., Bethge, M., Tolias, A.S. (2012) A fast and
830 simple population code for orientation in primate V1. *J Neurosci*, 32(31): 10618-26. doi:
831 10.1523/JNEUROSCI.1335-12.2012

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

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

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

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

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

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

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

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

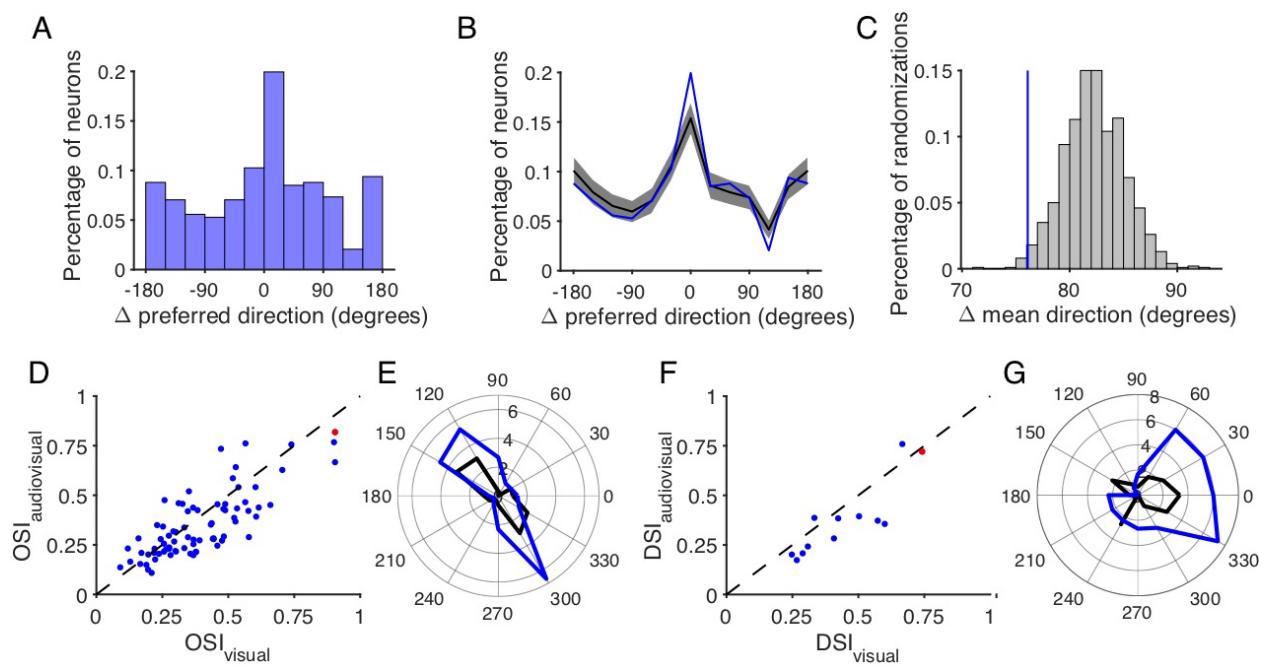
853 Deneux T., Harrell E.R., Kempf A., Ceballo S., Filipchuk A., Bathellier B. (2019) Context-
854 dependent signaling of coincident auditory and visual events in primary visual cortex. *eLife* 8:
855 e44006. doi: [10.7554/eLife.44006](https://doi.org/10.7554/eLife.44006)

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

- 859 Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: effects
860 of stimulus onset and intensity on reaction time. *Perception & psychophysics*, 66(8), 1388–
861 1404. <https://doi.org/10.3758/bf03195006>
- 862 Fahey, P.G., Muhammad, T., Smoth, C., Froudarakis, E., Cobos, E., Fu, J., Walker, E.Y.,
863 Yatsenko, D., Sinz, F.H., Reimer, J., Tolias, A.S. (2019). A global map of orientation tuning in
864 mouse visual cortex. BioRXiv.
- 865 Falchier, A., Clavagnier, S., Barone, P., Kennedy, H. (2002) Anatomical evidence of
866 multimodal integration in primate striate cortex. *J Neurosci*, 22(13): 5749-59. Doi:
867 10.1523/JNEUROSCI.22-13-05749.2002.
- 868 Fujisaki, W., Goda, N., Motoyoshi, I., Komatsu, H., Nishida, S. (2014) Audiovisual integration
869 in the human perception of materials. *J Vis*, 14(4):12. doi: 10.1167/14.4.12
- 870 Gingras, G., Rowland, B. A., & Stein, B. E. (2009). The differing impact of multisensory and
871 unisensory integration on behavior. *The Journal of neuroscience : the official journal of the
872 Society for Neuroscience*, 29(15), 4897–4902. <https://doi.org/10.1523/JNEUROSCI.4120-08.2009>
- 874 Gleiss, S., & Kayser, C. (2012). Audio-visual detection benefits in the rat. *PLoS one*, 7(9),
875 e45677. <https://doi.org/10.1371/journal.pone.0045677>
- 876 Gur, M., Beylin, A., Snodderly, D.M. (1997) Response variability of neurons in primary visual
877 cortex (V1) of alert monkeys. *J Neurosci*, 17(8):2914-20. doi: 10.1523/JNEUROSCI.17-08-
878 02914.1997
- 879 Hammond-Kenny, A., Bajo, V. M., King, A. J., & Nodal, F. R. (2017). Behavioural benefits of
880 multisensory processing in ferrets. *The European journal of neuroscience*, 45(2), 278–289.
881 <https://doi.org/10.1111/ejn.13440>
- 882 Hirokawa, J., Sadakane, O., Sakata, S., Bosch, M., Sakurai, Y., & Yamamori, T. (2011).
883 Multisensory information facilitates reaction speed by enlarging activity difference between
884 superior colliculus hemispheres in rats. *PLoS one*, 6(9), e25283.
885 <https://doi.org/10.1371/journal.pone.0025283>
- 886 Ibrahim, L. A., Mesik, L., Ji, X. Y., Fang, Q., Li, H. F., Li, Y. T., Zingg, B., Zhang, L. I., & Tao,
887 H. W. (2016). Cross-Modality Sharpening of Visual Cortical Processing through Layer-1-
888 Mediated Inhibition and Disinhibition. *Neuron*, 89(5), 1031–1045.
889 <https://doi.org/10.1016/j.neuron.2016.01.027>
- 890 Iurilli, G., Ghezzi, D., Olcese, U., Lassi, G., Nozzaro, C., Tonini, R., Tucci, V., Benfenati, F.,
891 Medini, P. (2012) Sound-driven synaptic inhibition in primary visual cortex. *Neuron*, 73(4):
892 814-28. Doi: 10.1016/j.neuron.2011.12.026
- 893 Knöpfel, T., Sweeney, Y., Radulescu, C. I., Zabouri, N., Doostdar, N., Clopath, C., & Barnes,
894 S. J. (2019). Audio-visual experience strengthens multisensory assemblies in adult mouse
895 visual cortex. *Nature communications*, 10(1), 5684. [https://doi.org/10.1038/s41467-019-13607-2](https://doi.org/10.1038/s41467-019-
896 13607-2)

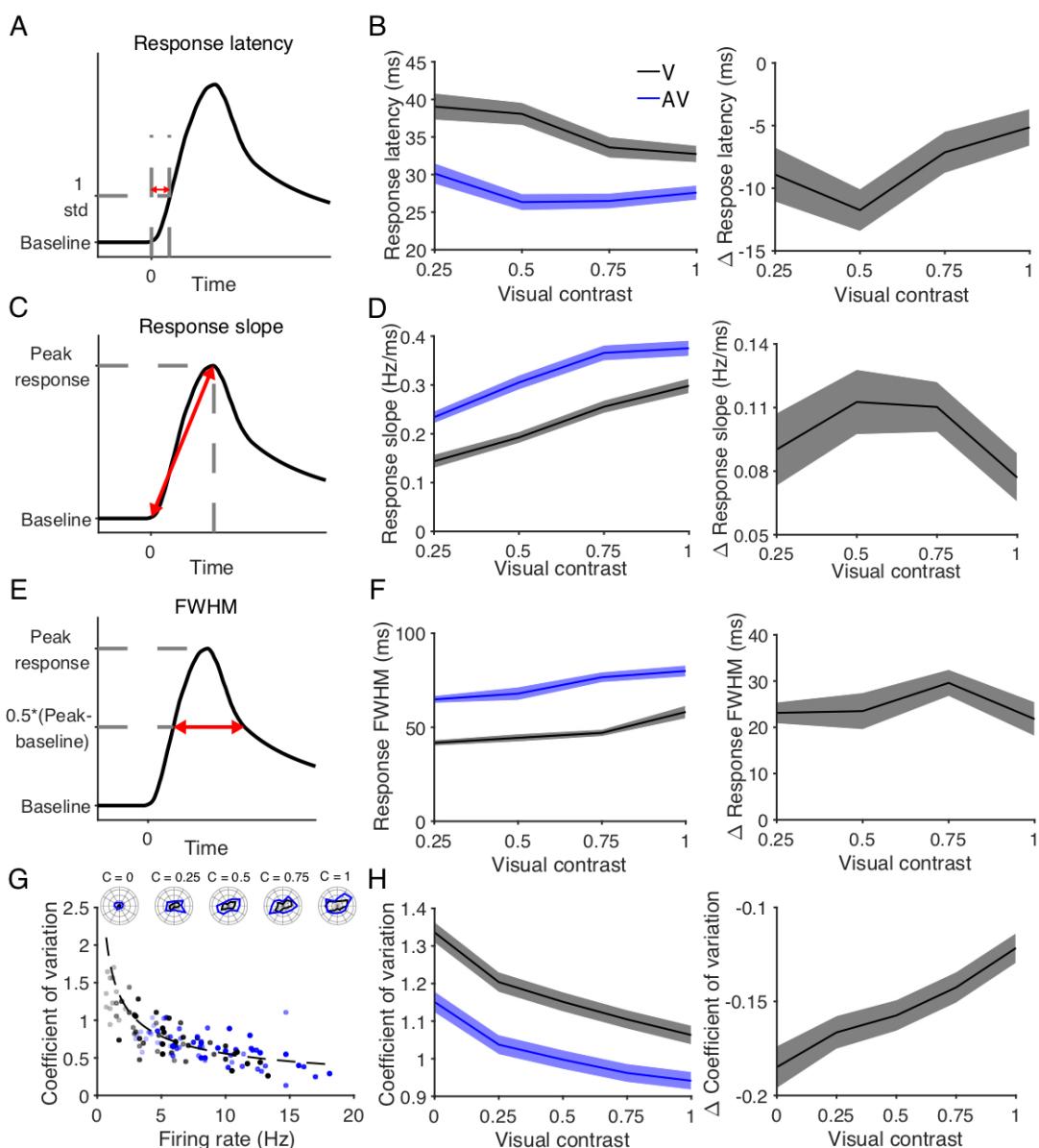
- 897 Maddox, R. K., Atilgan, H., Bizley, J. K., & Lee, A. K. (2015). Auditory selective attention is
898 enhanced by a task-irrelevant temporally coherent visual stimulus in human listeners. *eLife*, 4,
899 e04995. <https://doi.org/10.7554/eLife.04995>
- 900 Marlin, B.J., Mitre, M., D'amour, J.A., Chao, M.V., Froemke, R.C. (2015). Oxytocin enables
901 maternal behavior by balancing cortical inhibition. *Nature*, 520(7528): 499-504. doi:
902 10.1038/nature14402
- 903 McClure, J. P., Jr, & Polack, P. O. (2019). Pure tones modulate the representation of
904 orientation and direction in the primary visual cortex. *Journal of neurophysiology*, 121(6),
905 2202–2214. <https://doi.org/10.1152/jn.00069.2019>
- 906 McGurk, H., and Macdonald, J. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748.
907 doi: 10.1038/264746a0
- 908 Meijer, G. T., Montijn, J. S., Pennartz, C., & Lansink, C. S. (2017). Audiovisual Modulation in
909 Mouse Primary Visual Cortex Depends on Cross-Modal Stimulus Configuration and
910 Congruency. *The Journal of neuroscience : the official journal of the Society for
911 Neuroscience*, 37(36), 8783–8796. <https://doi.org/10.1523/JNEUROSCI.0468-17.2017>
- 912 Meijer, G. T., Pie, J. L., Dolman, T. L., Pennartz, C., & Lansink, C. S. (2018). Audiovisual
913 Integration Enhances Stimulus Detection Performance in Mice. *Frontiers in behavioral
914 neuroscience*, 12, 231. <https://doi.org/10.3389/fnbeh.2018.00231>
- 915 Meijer, G. T., Mertens, P., Pennartz, C., Olcese, U., & Lansink, C. S. (2019). The circuit
916 architecture of cortical multisensory processing: Distinct functions jointly operating within a
917 common anatomical network. *Progress in neurobiology*, 174, 1–15.
918 <https://doi.org/10.1016/j.pneurobio.2019.01.004>
- 919 Métin, C., Godement, P., & Imbert, M. (1988). The primary visual cortex in the mouse:
920 receptive field properties and functional organization. *Experimental brain research*, 69(3),
921 594–612. <https://doi.org/10.1007/BF00247312>
- 922 Montijn J.S., Vinck M., Pennartz C.M.A. (2014) Population coding in mouse visual cortex:
923 response reliability and dissociability of stimulus tuning and noise correlation. *Front Comput
924 Neurosci* 8:58. Doi: 10.3389/fncom.2014.00058
- 925 Morrill, R. J., & Hasenstaub, A. R. (2018). Visual Information Present in Infragranular Layers
926 of Mouse Auditory Cortex. *The Journal of neuroscience : the official journal of the Society for
927 Neuroscience*, 38(11), 2854–2862. <https://doi.org/10.1523/JNEUROSCI.3102-17.2018>
- 928 Musall, S., Kaufman, M.T., Juavinett, A.L., Gluf, S., Churchland, A.K. (2019) Single-trial neural
929 dynamics are dominated by richly varied movements. *Nat Neurosci*, 22(10):1677-1686. Doi:
930 10.1038/s41593-093-019-0502-4.
- 931 Nelson A., Schneider D.M., Takatoh J., Sakurai K., Wang F., Mooney R. (2013) A circuit for
932 motor cortical modulation of auditory cortical activity. *J Neurosci*, 33(36):14342-53. Doi:
933 10.1523/JNEUROSCI.2275-13.2013.
- 934 Niell C.M., Stryker M.P. (2010) Modulation of visual responses by behavioral state in mouse
935 visual cortex. *Neuron*, 65(4):472-9. Doi: 10.1016/j.neuron.2010.01.033.

- 936 Quak, M., London, R. E., & Talsma, D. (2015). A multisensory perspective of working
937 memory. *Frontiers in human neuroscience*, 9, 197. <https://doi.org/10.3389/fnhum.2015.00197>
- 938 Pachitariu M., Steinmetz N., Kadir S., Carandini M., Harris K.D. (2016) Kilosort: realtime spike-
939 sorting for extracellular electrophysiology with hundreds of channels. *BioRxiv*. Doi:
940 <https://doi.org/10.1101/061481>
- 941 Rochefort, N. L., Narushima, M., Grienberger, C., Marandi, N., Hill, D. N., & Konnerth, A.
942 (2011). Development of direction selectivity in mouse cortical neurons. *Neuron*, 71(3), 425–
943 432. <https://doi.org/10.1016/j.neuron.2011.06.013>
- 944 Schneider D.M., Mooney R. (2018) How movement modulates hearing. *Annu Rev Neurosci*,
945 41:553-72. Doi: 10.1146/annurev-neuro-072116-031215.
- 946 Shams L, Kamitani Y, Shimojo S (2002) Visual illusion induced by sound. *Brain Res Cogn
947 Brain Res* 14(1):147-52. Doi: 10.1016/s0926-6410(02)00069-1
- 948 Stanislaw, H., Todorov, N. (1999) Calculation of signal detection theory measures. *Behav Res
949 Methods Instrum Comput.* 31(1):137-49. Doi: 10.3758/bf03207704
- 950 Stein, B. E., Stanford, T. R., & Rowland, B. A. (2020). Multisensory Integration and the Society
951 for Neuroscience: Then and Now. *The Journal of neuroscience : the official journal of the
952 Society for Neuroscience*, 40(1), 3–11. <https://doi.org/10.1523/JNEUROSCI.0737-19.2019>
- 953 Tye-Murray, N., Spehar, B., Myerson, J., Hale, S., & Sommers, M. (2016). Lipreading and
954 audiovisual speech recognition across the adult lifespan: Implications for audiovisual
955 integration. *Psychology and aging*, 31(4), 380–389. <https://doi.org/10.1037/pag0000094>
- 956 von Trapp, G., Buran, B. N., Sen, K., Semple, M. N., & Sanes, D. H. (2016). A Decline in
957 Response Variability Improves Neural Signal Detection during Auditory Task
958 Performance. *The Journal of neuroscience : the official journal of the Society for
959 Neuroscience*, 36(43), 11097–11106. <https://doi.org/10.1523/JNEUROSCI.1302-16.2016>
- 960 Wang, Y., Celebrini, S., Trotter, Y., Barone, P. (2008) Visuo-auditory interactions in the
961 primary visual cortex of the behaving monkey: electrophysiological evidence. *BMC Neurosci*,
962 9:79. Doi: 10.1186/1471-2202-9-79
- 963 Zhao X., Chen H., Liu X., Cang J. (2013) Orientation-selective responses in the mouse lateral
964 geniculate nucleus. *J Neurosci* 33(31):12751-763. Doi: [10.1523/JNEUROSCI.0095-13.2013](https://doi.org/10.1523/JNEUROSCI.0095-13.2013)
- 965



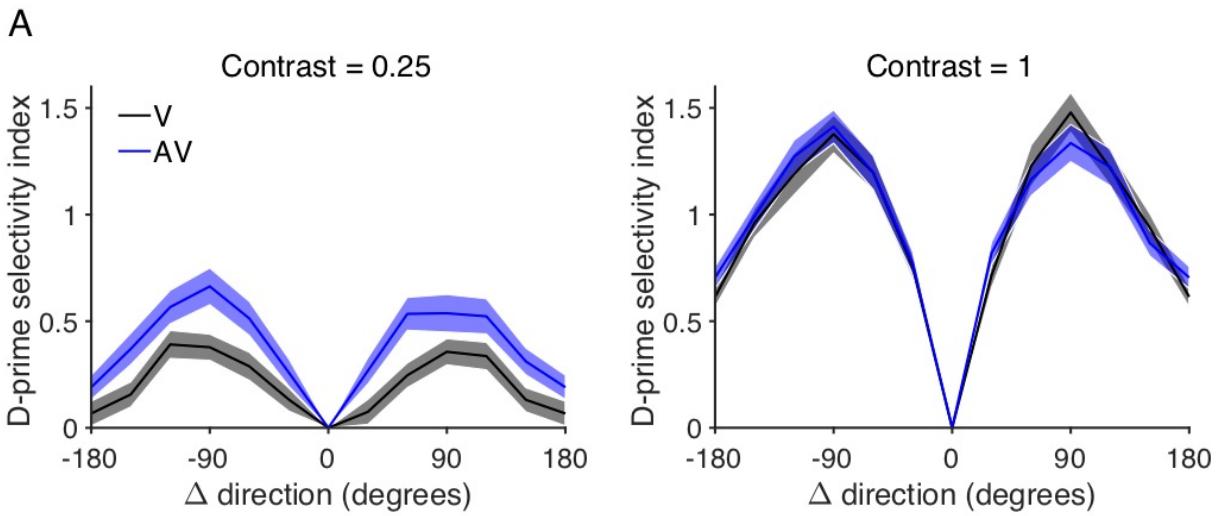
966
967
968
969
970
971
972
973
974
975
976
977

Figure 2 Supplementary 1 | Sound minimally reduces tuning selectivity in individual neurons (A) Histogram depiction of changes in preferred drifting grating directions with sound in orientation-selective neuron. (B) Observed changes in preferred direction (blue) compared to shuffled permutations (black) using the mean and standard deviation of observed responses. (C) The observed mean change in preferred direction (blue) is within the expected distribution (gray) based on visual response variability. (D,E) A slight reduction in the orientation selectivity index was observed in orientation-selective neurons ($n=78$, $p=0.0018$, paired t-test). The visual tuning of the red data point in D is displayed in E. (F,G) A slight reduction in the direction selectivity index was also observed in direction-selective neurons ($n=12$, $p=0.021$, paired t-test), with the tuning of the red data point in F displayed in G.



978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994

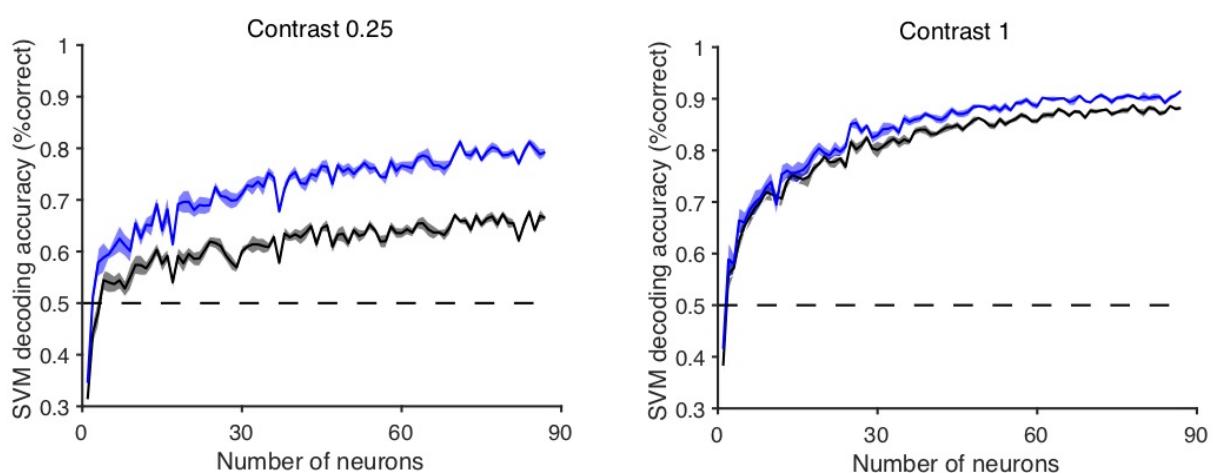
Figure 2 Supplementary 2 | Sound reduces the latency, increases duration, and reduces variability of light-evoked responses in individual neurons (A) Diagram of the calculation of response latency, the first time bin in which the FR exceeds 1 std above baseline. (B) Response latency is reduced by sound (left: absolute, right: difference; $p(\text{vis})=6.9e-4$, $p(\text{aud})=6.8e-15$, $p(\text{interact})=0.045$, paired 2-way ANOVA, post hoc Bonferroni-corrected paired t-test, Table 1). (C) Diagram of the calculation of response onset slope, the peak change in FR over the latency to peak response. (D) Sound increases the slope of the onset response (left: absolute, right: difference; $n=563$, $p(\text{vis})=3.5e-121$, $p(\text{aud})=2.7e-15$, $p(\text{interact})=0.038$, paired 2-way ANOVA, post hoc Bonferroni-corrected paired t-test). (E) Diagram of the calculation of FWHM, the width of the onset response at half maximum FR. (F) Sound increases the FWHM duration of the onset response (left: absolute, right: difference; $n=367$, $p(\text{vis})=1.3e-10$, $p(\text{aud})=8.7e-98$, $p(\text{interact})=0.23$ paired 2-way ANOVA). (G) An example neuron demonstrating that increased response magnitude corresponds to lower CV according to an inverse square root relationship. The black and blue dots represent visual and audiovisual responses, respectively, and the dot transparency corresponds to visual contrast level. The dotted lines are fitted $y=c/\sqrt{x}$ curves, where c is a constant. The above inset is the polar plots corresponding to the example neuron. (H) Lower coefficient of variation indicates reduced response variability in audiovisual compared to visual responses (left: absolute, right: difference; $n=563$, $p(\text{vis})=0.28$, $p(\text{aud})=4.2e-103$, $p(\text{interact})=0.38$, paired 2-way ANOVA).



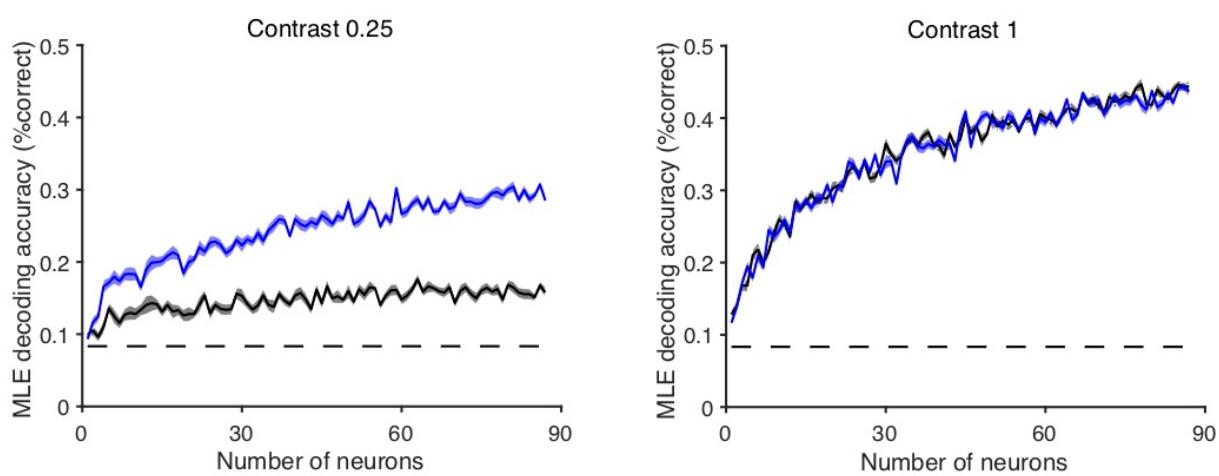
995
996 **Figure 5 Supplementary 1 | Sound enhances the d' sensitivity index at low contrast levels** (A) The d' sensitivity
997 index between neuronal responses to drifting grating directions, averaged across orientation- and direction-selective
998 neurons. Enhancements are observed at low visual contrast (left), whereas minimal changes are present at full
999 contrast (right).

1000
1001

A



B



1002
1003
1004
1005
1006
1007
1008

Figure 6 Supplementary 1 | Decoding accuracy increases with population size (A) Accuracy of SVM pairwise classification, average across all direction pairs, as the neuronal population size included in the decoder increases. Visual contrast 0.25 is on the left, and full visual contrast is on the right. (B) Accuracy of MLE decoding 1 of 12 drifting grating options, as the neuronal population size increases. Again, visual contrast 0.25 is on the left, and full visual contrast is on the right.

Table 1: Statistical comparisons

Comparison	Fig	Test	Test statistic	N	df	p-value	Post hoc test	Post hoc α	Post hoc comparison	Post hoc p-value
Mean firing rate, V vs AV	2C	Paired 2-way ANOVA	F(vis)=340 F(aud)=506 F(interact)=75	565 neurons	vis=4 aud=1 interact = 4	p(vis) = 1.2e-100 p(aud) = 1.6e-88 p(interact) = 5.7e-4	Bonferroni-corrected paired t-test	0.01	Contrast 0, V vs AV	2.1e-50
									Contrast 0.25, V vs AV	2.6e-62
									Contrast 0.5, V vs AV	5.7e-75
									Contrast 0.75, V vs AV	1.1e-81
									Contrast 1, V vs AV	2.0e-81
Linearity ratio, V vs AV	2E	Kruskal-Wallis test	Chi-sq = 61	555 neurons	4	p = 1.6e-12	Bonferroni-corrected Wilcoxon signed rank test	0.0125	Contrast 0 vs 0.25	0.053
									Contrast 0 vs 0.5	0.0040
									Contrast 0 vs 0.75	4.6e-8
									Contrast 0 vs 1	2.1e-5
Sound induced movement	3A	Paired t-test	t-stat = -7.2	9 recording sessions	8	p = 9.1e-5				
Firing rate across movement range, V vs AV	3E	Unbalanced 2-way ANOVA	F(motion)=6.9 F(sound)=55 F(interact)=18	Vari-able trial count	mot=2 aud=1 Interact=2	p(motion) = 0.001 p(sound) = 1.4e-13 p(interact) = 1.8e-8	Bonferroni corrected two-sample t-test	0.016	Stationary, V vs AV	1.5e-14
									Low motion, V vs AV	7.1e-10
									High motion, V vs AV	0.60
PSTH, light vs light/sound	4F	Paired t-test	1391 unique t-stats	295 neurons	294	1391 unique p-values, $\alpha=0.05/1391= 3.6e-5$				
PSTH, light vs light/motion	4G	Paired t-test	1391 unique t-stats	295 neurons	294	1391 unique p-values, $\alpha=0.05/1391= 3.6e-5$				
PSTH, light/sound vs light/sound/motion	4H	Paired t-test	1391 unique t-stats	295 neurons	294	1391 unique p-values, $\alpha=0.05/1391= 3.6e-5$				
Orientation selectivity index, V vs AV	Fig 2 Sup 1D	Paired t-test	t-stat = 3.2	78 neurons	77	p = 0.0018				
Direction selectivity index, V vs AV	Fig 2 Sup 1F	Paired t-test	t-stat = 2.7	12 neurons	11	p = 0.0206				
Onset response latency, V vs AV	Fig 2 Sup 2B	Paired 2-way ANOVA	F(vis)=5.7 F(aud)=64 F(interact)=2.7	517 neurons	vis=3 aud=1 interact=3	p(vis)=6.9e-4 p(aud)=6.8e-18 p(interact)=0.045	Bonferroni-corrected paired t-test	0.01	Contrast 0.25, V vs AV	2.3e-4
									Contrast 0.5, V vs AV	7.1e-12
									Contrast 0.75, V vs AV	4.6e-5
									Contrast 1, V vs AV	9.9e-4
Onset response slope, V vs AV	Fig 2	Paired 2-way ANOVA	F(vis)=70 F(aud)=66 F(interact)=2.8	563 neurons	vis=3 aud=1	p(vis)=3.5e-121 p(aud) = 2.7e-15 p(interact) = 0.038	Bonferroni-corrected paired t-test	0.01	Contrast 0.25, V vs AV	1.4e-4
									Contrast 0.5, V vs AV	8.9e-13
									Contrast 0.75, V vs AV	3.6e-12

	Sup 2D				intera ct=3			Contrast 1, V vs AV	5.5e-8
Onset response duration, V vs AV	Fig 2 Sup 2F	Paired 2-way ANOVA	F(vis)=17 F(aud)=129 F(interact)=1.4	367 neurons	vis=3 aud= 1 Intera ct=3	p(vis)=1.3e-10 p(aud) = 8.7e-98 p(interact) = 0.23			
Response coefficient of variation, V vs AV	Fig 2 Sup 2H	Paired 2-way ANOVA	F(vis)=1.3 F(aud)=834 F(interact)=1.0	564 neurons	vis=4 aud= 1 Intera ct=4	p(vis) = 0.28 p(aud) = 4.2e-103 p(interact) = 0.38			
Orientation decoding accuracy, individual neurons, V vs AV	Fig 5C	Paired 2-way ANOVA	F(vis)=67 F(aud)=12 F(interact)=0.54	78 neurons	vis=4 aud= 1 intera ct=4	p(vis)=4.8e-112 p(aud)=7.8e-4 p(interact) = 0.71			
Direction decoding accuracy, individual neurons, V vs AV	5E	Paired 2-way ANOVA	F(vis)=6.9 F(aud)=2.0 F(interact)=0.43	12 neurons	vis=4 aud= 1 intera ct=4	p(vis)=2.1e-4 p(aud)=0.18 p(interact)=0.78			
Orientation decoding accuracy, SVM, population, V vs AV	6C	2-way ANOVA	F(vis)=526 F(aud)=38 F(interact)=6	10 repeats	vis=4 aud= 1 intera ct=4	p(vis) = 1.8e-61 p(aud) = 1.9e-8 p(interact) = 2.4e-4	Bonferroni-corrected paired t-test	0.01	Contrast 0, V vs AV 0.12 Contrast 0.25, V vs AV 0.0016 Contrast 0.5, V vs AV 0.0014 Contrast 0.75, V vs AV 0.0023 Contrast 1, V vs AV 1
Direction decoding accuracy, SVM, population, V vs AV	6D	2-way ANOVA	F(vis)=48 F(aud)=40 F(interact)=4.6	10 repeats	vis=4 aud= 1 intera ct=4	p(vis) = 1.1e-21 p(aud) = 9.0e-9 p(interact) = 0.0019	Bonferroni-corrected paired t-test	0.01	Contrast 0, V vs AV 0.55 Contrast 0.25, V vs AV 5.3e-5 Contrast 0.5, V vs AV 0.0036 Contrast 0.75, V vs AV 0.17 Contrast 1, V vs AV 0.0036
Orientation decoding accuracy, MLE, population, V vs AV	6F	2-way ANOVA	F(vis)=682 F(aud)=0.27 F(interact)=18	10 repeats	vis=4 aud= 1 intera ct=4	p(vis)=2.3e-66 p(aud)=0.61 p(interact) =9.6e-11	Bonferroni-corrected paired t-test	0.01	Contrast 0, V vs AV 5.8e-4 Contrast 0.25, V vs AV 1.8e-4 Contrast 0.5, V vs AV 0.30 Contrast 0.75, V vs AV 0.53 Contrast 1, V vs AV 0.15
Direction decoding accuracy, MLE, population, V vs AV	6G	2-way ANOVA	F(vis)=67 F(aud)=0.43 F(interact)=8.9	10 repeats	vis=4 aud= 1 intera ct=4	p(vis)=4.6e-26 p(aud)=0.51 p(interact) =4.1e-6	Bonferroni-corrected paired t-test	0.01	Contrast 0, V vs AV 0.037 Contrast 0.25, V vs AV 6.4e-6 Contrast 0.5, V vs AV 0.036 Contrast 0.75, V vs AV 0.16 Contrast 1, V vs AV 0.014
Overall decoding accuracy, MLE, population, V vs AV	6J	2-way ANOVA	F(vis)=411 F(aud)=19 F(interact)=16	20 repeats	vis=4 aud= 1 intera ct=4	p(vis)=2.2e-92 p(aud)=1.9e-5 p(interact)=2.7e-11	Bonferroni -corrected paired t-test	0.01	Contrast 0, V vs AV 0.012 Contrast 0.25, V vs AV 1.4e-10 Contrast 0.5, V vs AV 0.48 Contrast 0.75, V vs AV 0.0013 Contrast 1, V vs AV 0.50

Orientation decoding accuracy, individual neurons, V vs AV	7B	Paired 2-way ANOVA	F(vis) = 74 F(aud) = 19 F(interact) = 1.5	85 neurons	vis=4 aud=1 interact=4	p(vis) =0 p(aud)=3.5e-5 p(interact)=0.21				
Orientation decoding accuracy, individual neurons, V vs motion-corrected AV	7B	Paired 2-way ANOVA	F(vis) = 64 F(aud) = 13 F(interact) = 3	85 neurons	vis=4 aud=1 interact=4	p(vis) =0 p(aud)=5.9e-4 p(interact)=0.019	Bonferroni-corrected paired t-test	0.01	Contrast 0, V vs AV Contrast 0.25, V vs AV Contrast 0.5, V vs AV Contrast 0.75, V vs AV Contrast 1, V vs AV	0.019 0.071 0.029 0.011 0.0602
Orientation decoding accuracy, individual neurons, AV vs motion-corrected AV	7B	Paired 2-way ANOVA	F(vis) = 34 F(aud) = 3.8 F(interact) = 2.4	85 neurons	vis=4 aud=1 interact=4	p(vis) = 7.7e-93 p(aud) = 0.055 p(interact) = 0.058				
Orientation decoding accuracy, individual neurons, V vs motion/sound-corrected AV	7B	Paired 2-way ANOVA	F(vis) = 56 F(aud) = 0.36 F(interact) = 1.4	85 neurons	vis=4 aud=1 interact=4	p(vis)=8.1e-95 p(aud)=0.55 p(interact)=0.24				
Population decoding accuracy, V vs AV	7D	2-way ANOVA	F(vis) = 166 F(aud) = 52 F(interact) = 8.2	10 repeats	vis=4 aud=1 interact=4	p(vis)=1.1e-40 p(aud)=1.6e-10 p(interact)=1.1e-5	Bonferroni-corrected paired t-test	0.01	Contrast 0, V vs AV Contrast 0.25, V vs AV Contrast 0.5, V vs AV Contrast 0.75, V vs AV Contrast 1, V vs AV	0.34 2.2e-5 0.0019 8.7e-6 0.013
Population decoding accuracy, V vs motion-corrected AV	7D	2-way ANOVA	F(vis) = 147 F(aud) = 35 F(interact) = 4.8	10 repeats	vis=4 aud=1 interact=4	p(vis)=1.4e-38 p(aud)=6.0e-8 p(interact)=0.0015	Bonferroni-corrected paired t-test	0.01	Contrast 0, V vs AV Contrast 0.25, V vs AV Contrast 0.5, V vs AV Contrast 0.75, V vs AV Contrast 1, V vs AV	0.30 0.0012 0.0022 0.0044 0.35
Population decoding accuracy, V vs motion/sound-corrected AV	7D	2-way ANOVA	F(vis) = 154 F(aud) = 0.50 F(interact) = 0.088	10 repeats	vis=4 aud=1 interact=4	p(vis)=2.5e-39 p(aud) = 0.48 p(interact) = 0.99				