MULTIMODAL MISMATCH RESPONSES IN MOUSE AUDITORY CORTEX

Magdalena Solyga[1] & Georg B. Keller[1,2]

[1]*Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland.*
[2]*Faculty of Science, University of Basel, Basel, Switzerland.*

SUMMARY

Our movements result in predictable sensory feedback that is often multimodal. Based on deviations between predictions and actual sensory input, primary sensory areas of cortex have been shown to compute sensorimotor prediction errors. How prediction errors in one sensory modality influence the computation of prediction errors in another modality is still unclear. To investigate multimodal prediction errors in mouse auditory cortex (ACx), we used a virtual environment to experimentally couple running to both self-generated auditory and visual feedback. Using two-photon microscopy, we first characterized responses of layer 2/3 (L2/3) neurons to sounds, visual stimuli, and running onsets and found responses to all three stimuli. Probing responses evoked by audiomotor mismatches, we found that they closely resemble visuomotor mismatch responses in visual cortex (V1). Finally, testing for cross modal influence on audiomotor mismatch responses by coupling both sound amplitude and visual flow speed to the speed of running, we found that audiomotor mismatch responses were amplified when paired with concurrent visuomotor mismatches. Our results demonstrate that multimodal and non-hierarchical interactions shape prediction error responses in cortical L2/3.

**** Dear reader, please note this manuscript is formatted in a standard submission format, and all statistical information is provided in Table S1. ****

INTRODUCTION

Neuronal responses consistent with prediction errors have been described in a variety of different cortical areas (Audette and Schneider, 2023; Ayaz et al., 2019; Han and Helmchen, 2023; Heindorf et al., 2018; Keller et al., 2012; Liu and Kanold, 2022) and across different species (Eliades and Wang, 2008; Keller and Hahnloser, 2009; Stanley and Miall, 2007). In V1, these responses are thought to be learned with experience (Attinger et al., 2017) and depend on local plasticity in cortex (Widmer et al., 2022). Prediction errors signal the unanticipated appearance or absence of a sensory input, and are thought to be computed as a deviation between top-down predictions and bottom-up sensory inputs (Rao and Ballard, 1999). One

type of top-down signals that conveys a prediction of sensory input are motor-related signals (Leinweber et al., 2017). In auditory cortex, motor-related signals can modulate responses to self-generated vocalizations (Eliades and Wang, 2008) or to sounds coupled to locomotion (Schneider et al., 2018). Auditory cortex is thought to use these motor-related signals to compute audiomotor prediction error responses (Audette et al., 2022; Audette and Schneider, 2023; Eliades and Wang, 2008; Keller and Hahnloser, 2009; Liu and Kanold, 2022). These audiomotor prediction errors can be described in a hierarchical variant of predictive processing, in which top-down motor-related signals function as predictions of bottom-up sensory input. While parts of both auditory and visual processing streams are well described by a hierarchy, the cortical network as a whole does not easily map onto a hierarchical architecture, anatomically (Markov et al., 2013) or functionally (St-Yves et al., 2023; Suzuki et al., 2023). One of the connections that does not neatly fit into a hierarchical model is the surprisingly dense reciprocal connection between ACx and V1 (Clavagnier et al., 2004; Falchier et al., 2002; Ibrahim et al., 2016; Leinweber et al., 2017; Zhao et al., 2022). From ACx to V1 this connection conveys a prediction of visual input given sound (Garner and Keller, 2022). What the reciprocal projection from V1 to ACx conveys, is still unclear. In proposals for hierarchical implementations of predictive processing there are no such lateral connections, and there is no reason to assume prediction error computations in different modalities should directly interact at the level of primary sensory areas. Thus, we argued that the lateral interaction between V1 and ACx is a good starting point to investigate how non-hierarchical interactions are involved in the computation of prediction errors, and how multimodal interactions shape sensorimotor prediction errors.

Based on this idea, we designed an experiment in which we could couple and transiently decouple running speed in a virtual environment to both self-generated auditory feedback and self-generated visual flow feedback. While doing this, we recorded activity in L2/3 neurons of ACx using two-photon calcium imaging. Using this approach, we first confirmed that a substantial subset of L2/3 neurons in ACx responds to either auditory, visual (Sharma et al., 2021), or motor-related inputs (Henschke et al., 2021; Morandell et al., 2023; Vivaldo et al., 2023). While we found that L2/3 neurons in ACx responded to audiomotor mismatches in a way that closely resembles visuomotor mismatch responses found in V1 (Keller et al., 2012), we found no evidence of responses to visuomotor mismatch in ACx. However, when coupling both visual flow and auditory feedback to running, we found that L2/3 neurons in ACx non-linearly combine information about visuomotor and audiomotor mismatches. Overall, our results demonstrate that prediction errors can be potentiated by multimodal interactions in primary sensory cortices.

RESULTS

Auditory, visual, and motor-related signals were intermixed in L2/3 of ACx.

To investigate auditory, visual, and motor-related signals in mouse ACx, we combined an audiovisual virtual reality (VR) system with two-photon calcium imaging in L2/3 ACx neurons (**Figure 1A**). We used an adeno-associated viral (AAV) vector to express a genetically encoded calcium indicator (AAV2/1-EF1α-GCaMP6f-WPRE) in ACx (**Figures 1B-D**). Following recovery from surgery, mice were habituated to the virtual reality setup (**Figure 1C**). We first mapped the location of the primary auditory cortex (A1) and the anterior auditory field (AAF) using widefield calcium imaging (**Figures 1C and 1E**). Based on these maps, we then chose recording locations for two-photon imaging of L2/3 neurons in either A1 or AAF. For the purposes of this work, we did not distinguish between A1 and AAF and will refer to these two areas here as ACx. To characterize basic sensory and motor-related responses, we recorded neuronal responses to pure tones, full-field moving gratings, and running onsets. Throughout all experiments mice were free to run, and auditory and visual responses were pooled across both sitting and running conditions (unless explicitly stated otherwise). We first assessed population responses of L2/3 neurons evoked by sounds (pure tones presented at 4 kHz, 8 kHz, 16 kHz, or 32 kHz at either 60 dB or 75 dB sound pressure level (SPL); **Figure 1F**) presented while the VR was off. While pure tones resulted in both increases or decreases in calcium activity in individual neurons (**Figure 1G**), the average population response exhibited a significant decrease in activity (**Figure 1H**). Next, we analyzed visual responses evoked by full-field drifting gratings (see Methods; **Figure 1I**). Visual stimulation resulted in a diverse response across the population of L2/3 neurons (**Figure 1J**) that was initially positive at the population level (**Figure 1K**). To quantify motor-related inputs, we analyzed activity during running onsets collected across all experimental conditions (**Figure 1L**). We found that the majority of neurons increased their activity during running onsets (**Figure 1M**), which was also reflected in a significant positive response on the population level (**Figure 1N**). Finally, we investigated how running modulates auditory and visual responses in ACx. In V1, running strongly increases responses to visual stimuli (Niell and Stryker, 2010), while in auditory cortex running has been shown to modulate auditory responses in a variety of different ways (Audette et al., 2022; Bigelow et al., 2019; Henschke et al., 2021; McGinley et al., 2015; Morandell et al., 2023; Schneider et al., 2014; Vivaldo et al., 2023; Yavorska and Wehr, 2021; Zhou et al., 2014). Separating auditory responses by running state, we found that sound evoked responses of ACx neurons were overall similar during sitting and running, but exhibited a smaller decrease in activity when the mouse was sitting (**Figure S1A**). Visual responses

consisted of an increase of activity during running, and a decrease of activity during sitting (**Figure S1B**). This is reminiscent of the running modulation effect observed on visual responses in V1 (Niell and Stryker, 2010). Thus, running appears to moderately and differentially modulate auditory and visual responses in L2/3 ACx neurons. Consistent with previous work, these results demonstrate that auditory, visual, and motor-related signals are all present in ACx and that running modulation influences L2/3 ACx neurons differently than in V1.

L2/3 neurons in ACx responded to audiomotor mismatch.

To test whether auditory, visual, and motor-related signals are integrated in L2/3 neurons of ACx to compute prediction errors, we first probed for responses to audiomotor (AM) mismatches. A mismatch in this context, is the absence of a sensory input that the brain predicts to receive from the environment, and thus a specific type of negative prediction error. We experimentally generated a coupling between movement and sensory feedback and then used movement as a proxy for what the mouse predicts to receive as sensory feedback. To do this with an auditory stimulus, we coupled the sound amplitude of an 8 kHz pure tone to the running speed of the mouse on the spherical treadmill such that sound amplitude was proportional to locomotion speed (**Figures 2A and 2B**)**.** In this paradigm, a running speed of 0 corresponded to a sound amplitude of 0, while 30 cm/s running speed corresponded to a sound amplitude of 60 dB SPL. We refer to this type of session as closed loop. We then introduced AM mismatches by setting the sound amplitude to 0 for 1 s at random times (on average every 15 s). An alternative approach to introduce AM mismatches would have been to clamp the sound amplitude to a constant value. However, based on an assumed analogy between sound amplitude and visual flow speed in visuomotor (VM) mismatch paradigms, where we induce VM mismatch by setting visual flow speed to 0 (Keller et al., 2012), we chose the former. We found that AM mismatch resulted in a strong population response (**Figures 2C and 2D**). AM mismatch responses likely cannot be attributed to changes in running speed and do not result in a pupil response. We found no evidence of AM mismatch induced changes in running speed (**Figure S2A**) or pupil dilation (**Figure S2B**). Interestingly, the AM mismatch response was already apparent in the first closed loop session with audiomotor coupling that the mice ever experienced, suggesting that this coupling is learned very rapidly (**Figure S2C**). To test whether AM mismatch responses can be explained by a sound offset response, we performed recordings in open loop sessions that consisted of a replay of the sound profile the mouse had self-generated in the preceding closed loop session. Mice were free to run during this session and did so at similar levels as during the closed loop

session (**Figure S2D**). The average response to the playback of sound halt during the open loop session was significantly less strong than the average response to AM mismatch (**Figure 2D**), but in contrast to visual playback halt responses in V1 (Vasilevskaya et al., 2023), we found no evidence of a running modulation of the response to the playback halt (**Figure S2E**). Thus, L2/3 neurons in ACx respond to AM mismatch in a way similar to how the L2/3 neurons in V1 respond to visuomotor (VM) mismatch.

Assuming that AM mismatch responses are computed as a difference between an excitatory motor-related prediction and an auditory stimulus driven inhibition, we would expect the neurons with high AM mismatch responses to exhibit opposing influence of motor-related and auditory input. To test this, we selected the 5% of neurons with the strongest responses to AM mismatch and quantified the responses of these neurons to sound stimulation and running onsets (as for the sound and running onset responses shown in **Figure 1**). Consistent with a model of a subtractive computation of prediction errors, we found that AM mismatch neurons exhibited a strong reduction in activity in response to sound stimulation and an increase of activity on running onsets (**Figure 2E**). Given that mismatch responses are likely enriched in the superficial part of L2/3 (O'Toole et al., 2023), and that in our two-photon imaging experiments we also preferentially recorded from more superficial neurons, we suspect that our population is enriched for mismatch neurons. Consistent with this interpretation, we observed a strong population response to AM mismatches (**Figures 2C and 2D**) and a decrease in population activity in response to sound stimulation (**Figure 1H**). Nevertheless, sound evoked responses were significantly more negative in neurons strongly responsive to AM mismatch, than for the remainder of the L2/3 neuronal population (**Figure 2F**). This effect was similar when we used different thresholds for the selection of AM mismatch neurons (10% or 20% of neurons with the strongest response to AM mismatch; **Figure S3**). Consistent with a sound driven reduction of activity and running related increase of activity in AM mismatch neurons, the correlation of calcium activity of AM mismatch neurons was predominantly negative with sound amplitude and positive with running speed in open loop sessions (**Figure 2G**). This again resembles the properties of VM mismatch neurons in V1 (Attinger et al., 2017). If AM mismatch responses are computed as a difference between a locomotion driven excitation and a sound driven inhibition, we could also expect to find a correlation between the strength of mismatch response and the strength of sound playback halt responses. Even in the absence of locomotion driven excitation, a relief from sound driven inhibition could trigger an increase in calcium activity. When comparing AM mismatch responses with playback sound halt responses for all neurons, we do indeed find a positive correlation between the two (**Figure S2F**). Finally, again assuming a locomotion driven excitation and a sound driven inhibition, we should find systematic differences in

closed and open loop running onsets. In closed loop running onsets in which the sound was coupled to movement, we indeed observed only a transient running onset response (**Figure 2H**). In contrast, in the open loop condition, running onset responses were sustained, possibly reflecting motor related input that is not cancelled out by sensory input. This analysis highlights the difference in input processing depending on whether the animal is in the coupled or uncoupled condition and is consistent with subtractive interactions between prediction and sensory signals. Overall, these results suggest that the implementation of sensorimotor prediction error computation generalizes beyond V1 to other primary cortices and might be a canonical cortical computation in L2/3.

We found no evidence of visuomotor mismatch responses in L2/3 of ACx.

Visuomotor mismatch responses are likely calculated in V1 (Jordan and Keller, 2020), and spread across dorsal cortex from there (Heindorf and Keller, 2023). To investigate multimodal mismatch responses, we first quantified the strength of these VM mismatch responses, which are independent of auditory input, in ACx. In these experiments, the running speed of the mouse was experimentally coupled to the visual flow speed in a virtual corridor, but not to sound feedback (**Figures 3A and 3B**). Note, there are several remaining sources of naturally occurring sound that increase with increasing running speed of the mouse and are still coupled (sound of mouse footsteps, sound of the treadmill rotating, etc.). However, these cannot be easily experimentally manipulated. We introduced VM mismatches by halting visual flow for 1 s at random times while the mice were running, as previously described (Keller et al., 2012; Zmarz and Keller, 2016). To control for visual responses independent of visuomotor coupling, we used an open loop replay of the visual flow generated in the previous session (see Methods). We found that neither VM mismatches nor visual flow playback halts, which the mouse experienced in open loop sessions, resulted in a measurable population response in ACx (**Figures 3C and 3D**). No significant changes in locomotion speed (**Figure S4A**) or pupil diameter (**Figure S4B**) were detected upon VM mismatch presentation. Selecting the 5% of neurons with the strongest responses to VM mismatches and quantifying their responses to grating presentations and running onsets, we found that these neurons exhibited positive responses to running onset and no significant response to grating stimuli (**Figure 3E**). These responses were not different from the population responses of the remainder of the neurons (**Figure 3F**). Quantifying the correlation of calcium activity with visual flow speed and running speed in the open loop session, we found that VM mismatch responsive neurons exhibited a distribution not different from chance (**Figure 3G**). Running onset responses were not different in closed and open loop conditions (**Figure 3H**). This

181 suggests that motor-related predictions are not cancelled out by visual inputs in the auditory cortex. We
182 also found no evidence of a correlation between VM mismatch responses and playback halt responses
183 (**Figure S4C**). Thus, while there may be a small subset of VM mismatch responsive neurons in L2/3 of ACx,
184 we find no evidence of a VM mismatch response at the level of the L2/3 population.

185 **Mismatch responses were potentiated by multimodal interactions.**

186 Finally, we explored how multimodal coupling of both auditory and visual feedback to running speed
187 influenced mismatch responses in L2/3 of ACx. To do this, we coupled both sound amplitude and visual
188 flow speed to the running speed of the mouse in an audiovisual virtual environment (**Figures 4A and 4B**).
189 We then introduced mismatch events by halting both sound and visual flow for 1 s to trigger a concurrent
190 audiomotor and visuomotor [AM + VM] mismatch (**Figure 4B**). The nomenclature here is such that the
191 first letter in the pair denotes the sensory input that is being predicted, while the second letter denotes
192 the putative predictor – the square brackets are used to denote that the two events happen concurrently.
193 By putative predictor, we mean an information source available to the mouse that would, in principle,
194 allow it to predict another input, given the current experimental environment. Thus, in the case of a [AM
195 + VM] mismatch both the movement related prediction of visual flow and sound amplitude are violated.
196 The [AM + VM] mismatch resulted in a significant response on the population level (**Figure 4C**). As with
197 [AM] and [VM] mismatch, we found no evidence of a change in running speed or pupil diameter following
198 [AM + VM] mismatch (**Figures S5A and S5B**). The concurrent experience of mismatch between multiple
199 modalities could simply be the result of a linear combination of the responses to the different mismatch
200 stimuli or could be the result of a non-linear combination. To test whether we find evidence of a non-
201 linear combination of mismatch responses, we compared the [AM+VM] mismatch to [AM] and [VM]
202 mismatch events presented alone. We found that the presentation of a [AM+VM] mismatch led to a
203 significantly larger response than either an [AM] or a [VM] mismatch in isolation (**Figure 4D**). To test
204 whether the linear summation of [AM] + [VM] mismatch responses could explain the response to the
205 concurrent presentation [AM+VM], we compared the two directly, and found that the concurrent
206 presentation [AM+VM] elicited a significantly larger response than the linear sum of [AM] + [VM]
207 mismatch responses (**Figures 4E and S5C**). Plotting the [AM+VM] mismatch responses against the linear
208 sum of the [AM] + [VM] mismatch responses for each neuron, we found that while there is some
209 correlation between the two, there is a subset of neurons (13.7%; red dots, **Figure 4F**) that selectively
210 respond to the concurrent [AM+VM] mismatch, while a different subset of neurons (11.2%; yellow dots,

211 **Figure 4F**) selectively responds to the mismatch responses in isolation. This demonstrates that mismatch
212 responses in different modalities can interact non-linearly.

213 **DISCUSSION**

214 Consistent with previous reports, we found that auditory, visual, and motor-related signals are intermixed
215 in the population of L2/3 neurons in ACx (**Figures 1H, 1K and 1N**). Responses to both motor-related
216 (McGinley et al., 2015; Schneider et al., 2014; Vivaldo et al., 2023; Yavorska and Wehr, 2021; Zhou et al.,
217 2014) and visual signals (Bigelow et al., 2022; Morrill and Hasenstaub, 2018; Sharma et al., 2021) have
218 been reported across layers in ACx, with the strongest running modulation effect found in L2/3 (Schneider
219 et al., 2014). Also, consistent with previous reports, we found that a subset of L2/3 neurons in ACx respond
220 to audiomotor prediction errors (**Figures 2C and 2D**) (Audette et al., 2022; Liu and Kanold, 2022). In V1, it
221 has been demonstrated that neurons signaling prediction errors exhibit opposing influence of bottom-up
222 visual and top-down motor-related inputs. This has been speculated to be the consequence of a
223 subtractive computation of prediction errors (Jordan and Keller, 2020; Keller et al., 2012; Leinweber et al.,
224 2017). Our findings now reveal a similar pattern of opposing influence in prediction error neurons in
225 primary ACx that exhibit a positive correlation with motor-related input and a negative correlation with
226 auditory input (**Figure 2G**). This would be consistent with the idea that both visuomotor and audiomotor
227 prediction errors are computed as a subtractive difference between bottom-up and top-down inputs.
228 Based on this, it is conceivable that this type of computation extends also beyond primary sensory areas
229 of cortex and may be a more general computational principle implemented in L2/3 of cortex.

230 While there are many similarities between the AM mismatch responses we describe here in auditory
231 cortex and previously described visuomotor mismatch responses in visual cortex (Attinger et al., 2017;
232 Keller et al., 2012; Widmer et al., 2022; Zmarz and Keller, 2016), there are several notable differences.
233 Specifically, AM responses in auditory cortex (**Figure 2D**) appear to be more sustained than VM mismatch
234 responses in visual cortex. We have no explanation for why this might be the case. Additionally, unlike the
235 observed visual flow playback halt responses in V1 and their increase during running (Vasilevskaya et al.,
236 2023), we found only very weak responses to sound playback halts and no evidence of an influence of
237 running on these responses (**Figure S2E**). A decreased dependence of playback halt responses on running
238 may be caused by the fact that the mouse generates additional sounds while running on an air-supported
239 treadmill (e.g., treadmill rotation, changes in airflow, footsteps). These sounds correlate with running

speed intensity and may reduce the relative salience of sound playback halts during running. Thus, differences in the experimental control of the sensorimotor coupling for a different sensory modalities could account for some of the observed differences between the AM and VM mismatch responses.

Finally, we found that concurrent prediction errors in multiple modalities result in an increase in prediction error response that exceeds a linear combination of the prediction error responses in single modalities (**Figure 4E**), with a subset of neurons selectively responding only to the combination of prediction error responses (**Figure 4F**). A similar non-linear relationship has been described between auditory and visual oddball responses in both ACx and V1 (Shiramatsu et al., 2021). At this point, it should be kept in mind that deviations from linearity in terms of spiking responses are difficult to assess using calcium imaging data. However, given that the difference between the concurrent presentation and the linear sum of the two individual mismatch responses was approximately a factor of two (**Figure 4E**), and the fact that we found a population of neurons that responds selectively to the concurrent presentation of both mismatches, we suspect that also the underlying spiking responses are non-linear. What are the mechanisms that could underlie this interaction? Neurons in ACx have access to information about VM mismatches from at least two sources. In widefield calcium imaging, VM mismatch responses are detectable across most of dorsal cortex (Heindorf and Keller, 2023). Thus, we speculate that VM mismatch responses are present in long-range cortico-cortical axons from V1, or possibly in L4, L5, or L6 neurons in ACx. Alternative sources of VM mismatch input are neuromodulatory signals. Locus coeruleus, for example, drives noradrenergic signals in response to VM mismatches across the entire dorsal cortex (Jordan and Keller, 2023). However, given that noradrenergic signals only weakly modulate responses in L2/3 neurons in V1 (Jordan and Keller, 2023), it is unclear if the broadcasted noradrenergic signals could non-linearly potentiate the AM mismatch responses of ACx neurons. We speculate that cholinergic signals are also unlikely to contribute to this effect. In V1 there are no cholinergic responses to VM mismatch (Yogesh and Keller, 2023). However, given that ACx and V1 receive cholinergic innervation from different sources (Kim et al., 2016), we cannot rule out the possibility that cholinergic signals in ACx respond to VM mismatch. In sum, given that the [AM+VM] mismatch responses do not simply appear to be an amplified variant of the [AM] mismatch responses (**Figure 4F**), we speculate that [AM+VM] mismatch responses are primarily driven by long-range cortico-cortical input from V1 that interacts with a local computation of [AM] mismatch responses in the L2/3 ACx circuit.

Lateral interactions in the computation of prediction errors between sensory streams are not accounted for by hierarchical variants of predictive processing. In these hierarchical variants, prediction errors are computed as a comparison between top-down and bottom-up inputs (Rao and Ballard, 1999). To explain the lateral interactions between prediction errors likely computed in ACx (AM mismatch responses) and prediction errors likely computed in V1 (VM mismatch responses) that we describe here, we will need new variants of predictive processing models that include lateral and non-hierarchical interactions. Thus, our results demonstrate that mismatch responses in different modalities interact non-linearly and can potentiate each other. The circuit mechanisms that underlie this form of multimodal integration of mismatch responses are still unclear and will require further investigation. However, we would argue that the relatively strong multimodal interaction demonstrates that unimodal and hierarchical variants of predictive processing are insufficient to explain cortical mismatch responses - if predictive processing aims to be a general theory of cortical function, we will need to explore non-hierarchical variants of predictive processing.

FIGURES

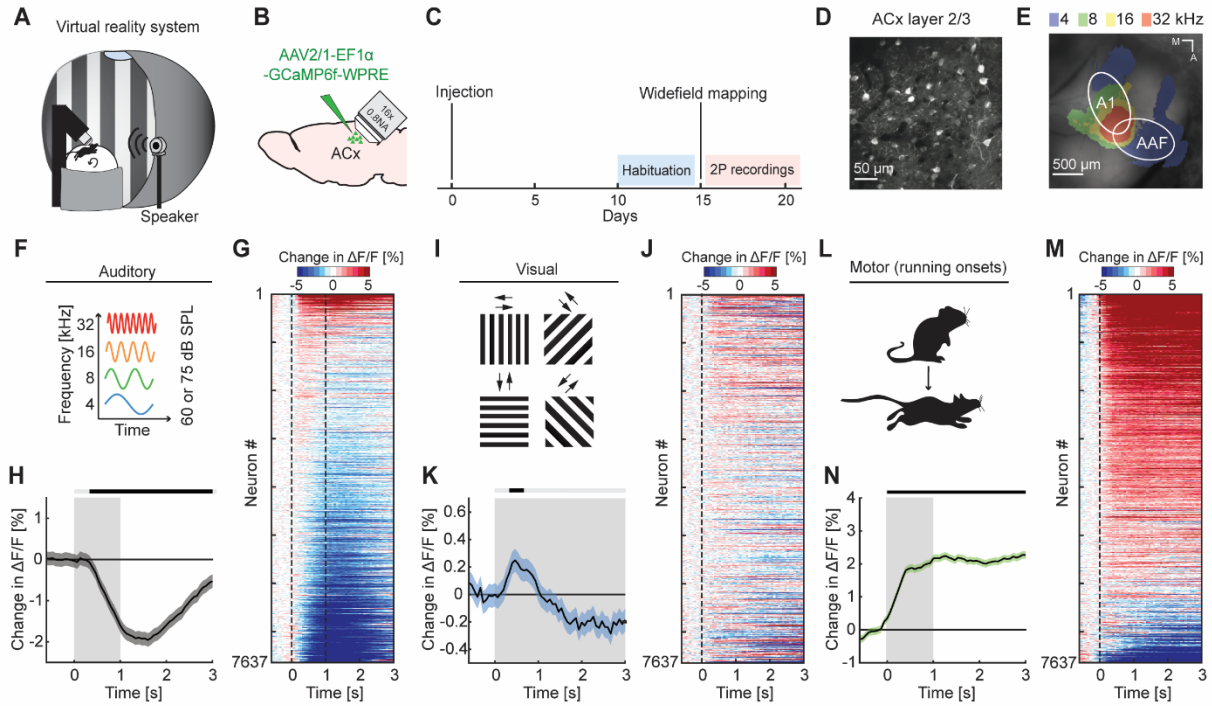


Figure 1. Auditory, visual, and motor-related signals were present in L2/3 of ACx.

(**A**) Schematic of the virtual reality system. For imaging experiments, mice were head-fixed and free to run on an air-supported spherical treadmill. For all recordings, the microscope was tilted 45 degrees to the left to image left ACx.

(**B**) Strategy for two-photon imaging of L2/3 ACx neurons. We injected an AAV vector to express a genetically encoded calcium indicator in ACx.

(**C**) Timeline of the experiment. Starting 10 days after viral injection and window-implantation surgery mice were habituated to the virtual reality setup without any visual or auditory stimulation for 5 days. We mapped ACx with widefield imaging to be able to target two-photon (2P) imaging to ACx. In 1 to 6 recording sessions, 1 session per day, we recorded from 7637 neurons in 17 mice.

(**D**) Example two-photon image in L2/3 of ACx.

(**E**) Example widefield mapping of ACx. Response maps reflect regions with the strongest response for each tested sound frequency.

(**F**) The sound stimuli were 1 s long pure tones of 4 kHz, 8 kHz, 16 kHz, or 32 kHz played at 60 dB or 75 dB sound pressure level (SPL), presented with randomized inter-stimulus intervals.

(**G**) The average sound evoked response of all L2/3 ACx neurons across all tested frequencies and sound levels. Sound is presented from 0 s to 1 s. Red indicates an increase in activity, while blue indicates a decrease in activity. All responses are baseline subtracted. To avoid regression to the mean artifacts in plotting, the response heatmap is generated by splitting data in two halves by trials. The responses from the first half of trials are used to sort neurons by response strength and the average responses of the

306 second half of trials are plotted for each neuron. To prevent graphical aliasing, the heatmaps are
307 smoothed over 10 neurons for plotting.

308 (**H**) The average sound evoked population response of all ACx L2/3 neurons across all tested frequencies
309 and sound levels (7637 neurons). Stimulus duration was 1 s (gray shading). Here and in subsequent panels,
310 solid black lines represent mean and shading SEM. The horizontal bar above the plot marks time bins in
311 which the response is statistically different from 0 (gray: not significant, black: $p<0.05$; see Methods).

312 (**I**) The visual stimuli we used were full-field drifting gratings of 8 different directions, presented for 4 s to
313 8 s with randomized inter-stimulus intervals.

314 (**J**) As in **G**, but for gratings onsets responses averaged across all orientations.

315 (**K**) As in **H**, but for the population response to grating onsets averaged across all orientations.

316 (**L**) Motor-related activity was assessed based on responses upon running onsets.

317 (**M**) As in **G**, but for running onset responses.

318 (**N**) As in **H**, but for the average population response to running onsets. Only data from running onsets in
319 which the mouse ran for at least 1 s (gray shading) were included.

320

321

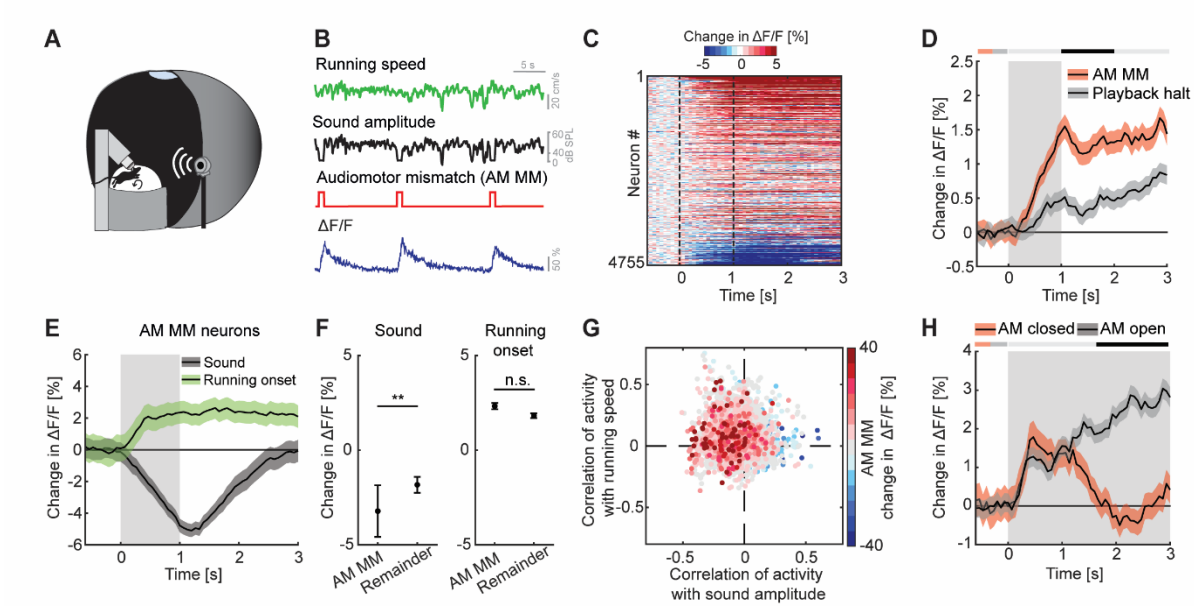


Figure 2. L2/3 neurons of ACx responded to audiomotor mismatch events.

(**A**) Schematic of the virtual reality system used to study responses to audiomotor (AM) mismatches. The sound amplitude of an 8 kHz pure tone was coupled to the running speed of the mouse on a spherical treadmill. These experiments were performed with the VR switched off.

(**B**) In closed loop sessions, the running speed of the mouse was coupled to the sound amplitude. AM mismatches were introduced by briefly setting the sound amplitude to 0 for 1 s. Below, the calcium response of an example neuron to AM mismatch events.

(**C**) Responses of all L2/3 ACx neurons to audiomotor mismatches. The response heatmap is generated as described in **Figure 1G**.

(**D**) The average population response of all L2/3 neurons to AM mismatches and sound playback halts (4755 neurons). AM mismatch duration was 1 s (gray shading). The horizontal bar above the plot marks time bins in which the AM mismatch response is statistically different from the playback halt response (gray: not significant., black: $p<0.05$; see Methods).

(**E**) The average population response of AM mismatch neurons (5% of strongest responders) to sound stimulation (black) and running onsets (green). Same data as in **Figures 1H and 1N**, but subselected for AM MM neurons. Sound stimulation was 1 s (gray shading).

(**F**) Comparison of the response strength of AM mismatch (MM) neurons to sound stimulation (left) and running onsets (right) compared to those of the remainder of the neuron population. Error bars indicate SEM. Here and elsewhere, n.s.: not significant; *: $p<0.05$; **: $p<0.01$; ***: $p<0.001$. See Table S1 for all statistical information.

(**G**) Scatter plot of the correlations of calcium activity with sound amplitude (x-axis) and running speed (y-axis) in open loop sessions, for all neurons. The color-code reflects the strength of responses to AM mismatch in the closed loop session. Note, AM mismatch-responsive neurons are enriched in the upper left quadrant.

(**H**) The average population response to running onsets in closed (red) and open (black) loop sessions (only data from neurons for which we had at least 2 running onsets in both closed and open loop sessions are included here, see **Table S1**).

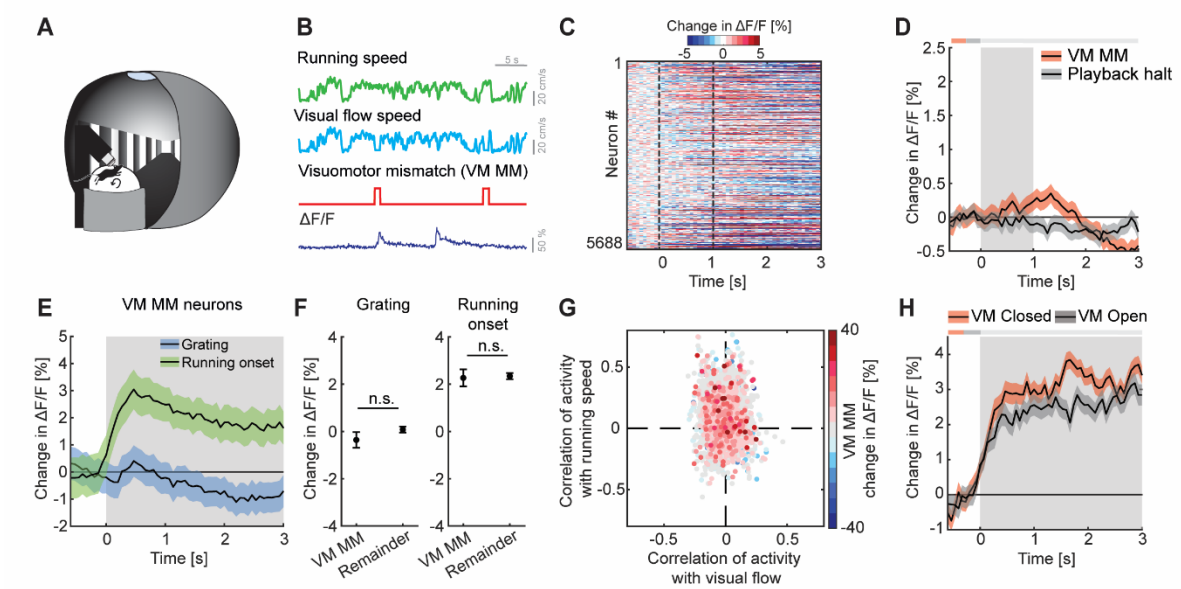


Figure 3. We found no evidence of visuomotor mismatch responses in ACx.

(**A**) Schematic of the virtual reality system used to measure VM mismatch responses. The visual flow of the virtual corridor was coupled to the running speed of the mouse on a spherical treadmill. There was no sound stimulus present in these experiments.

(**B**) In closed loop sessions, the running speed of the mouse was coupled to the movement in a virtual corridor. VM mismatches were introduced by briefly setting visual flow speed to 0 for 1 s. Below, the calcium response of an example neuron to VM mismatch events.

(**C**) Responses of all L2/3 ACx neurons to VM mismatches. The response heatmap is generated as described in **Figure 1G**.

(**D**) The average population response of all L2/3 neurons to VM mismatches and visual flow playback halts (5688 neurons). Gray shading marks the duration of both stimuli. The horizontal bar above the plot marks time bins in which the VM mismatch response is statistically different from the playback halt response (gray: n.s., black: $p<0.05$; see Methods).

(**E**) The average population response of VM mismatch neurons (5% of strongest responders) to grating stimulation (blue) and running onsets (green). Same data as in **Figures 1H and 1N**, but subselected for VM MM neurons. Stimulus duration was 4 s to 8 s (gray shading).

(**F**) Comparison of the response strength of VM mismatch (MM) neurons to visual stimulation (left) and running onsets (right) compared to those of the remainder of the neuron population. Error bars indicate SEM. Here and elsewhere, n.s.: not significant; *: $p<0.05$; **: $p<0.01$; ***: $p<0.001$. See Table S1 for all statistical information.

(**G**) Scatter plot of the correlation of calcium activity with visual flow speed (x-axis) and running speed (y-axis) in open loop sessions for all neurons. The color-code reflects the strength of responses to VM mismatch in the closed loop session. Note, VM mismatch responsive neurons are scattered randomly.

(**H**) The averaged population response to running onsets in closed (red) and open (black) loop session. (only data from neurons for which we had at least 2 running onsets in both closed and open loop sessions are included here, see **Table S1**).

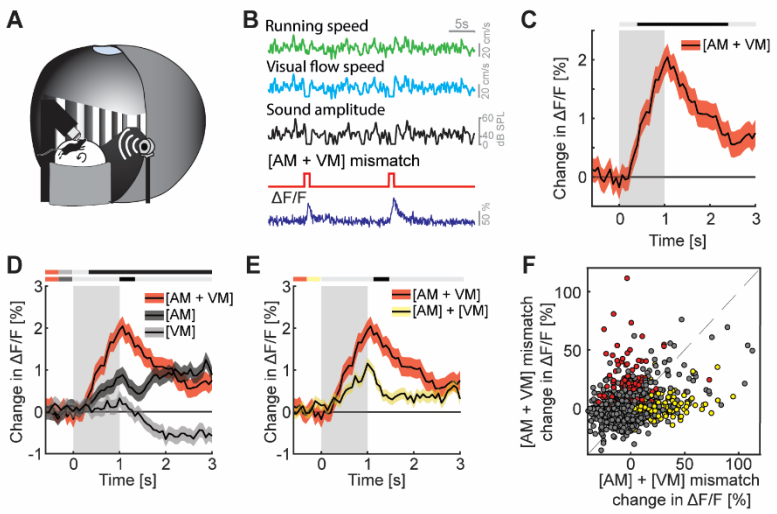


Figure 4. Mismatch responses were potentiated by multimodal interactions in L2/3 ACx neurons.

(**A**) Schematic of the virtual reality system used to measure multimodal mismatch responses. Both the sound amplitude of an 8 kHz pure tone and the movement of the virtual corridor were coupled to the running speed of the mouse on a spherical treadmill.

(**B**) Concurrent audiomotor and visuomotor [AM + VM] mismatches were introduced by simultaneously setting both sound amplitude and visual flow speed to 0 for 1 s.

(**C**) The average population response of all L2/3 ACx neurons to concurrent [AM + VM] mismatches (3289 neurons). Gray shading marks the duration of the mismatch stimulus. The horizontal bar above the plot marks time bins in which the [AM + VM] mismatch response is statistically different from 0 (gray: n.s., black: $p<0.05$; see Methods).

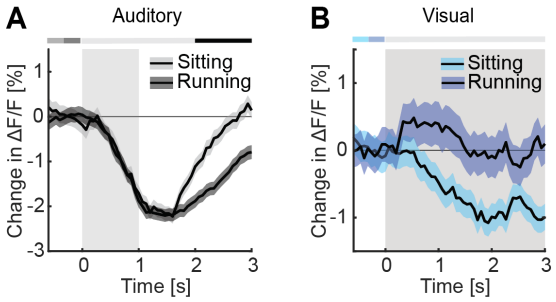
(**D**) Average population response of all L2/3 neurons to concurrent [AM + VM] mismatch and responses evoked by [AM] and [VM] mismatches presented in isolation (same data as shown in **Figures 2 and 3** but sub-selected to match the neurons that were recorded in both unimodal and multimodal mismatch paradigms; 3289 neurons). The horizontal bars above the plot mark time bins in which the [AM + VM] mismatch response is larger than the [AM] or the [VM] mismatch response (gray: n.s., black: $p<0.05$; see Methods). The two short horizontal color bars to the left of time 0 indicate which two responses are being compared. Gray shading indicates the duration of the stimulus.

(**E**) Average population response of all L2/3 neurons to a concurrent [AM + VM] mismatches compared to the linear sum of the responses evoked by [AM] and [VM] mismatches presented in isolation. Gray shading indicates the duration of the stimuli.

(**F**) Scatter plot of the responses of all neurons to the concurrent [AM + VM] mismatches against the linear sum of the responses evoked by [AM] and [VM] mismatches presented in isolation. In red, the subset of neurons (13.7%) that exhibited selective responses to the concurrent [AM+VM] mismatch, and in yellow, the subset of neurons (11.2%) in which the linear sum of the responses to [AM] + [VM] mismatches presented in isolation was significant, while response to their concurrent presentation was not. Neurons without a significant response, or that are responsive to both, are shown in gray.

407 **SUPPLEMENTARY FIGURES**

408



409

410 **Figure S1. Running exhibited differential effects on the responses to sound presentation and moving**
411 **grating onsets. Related to Figure 1.**

412 (**A**) The average population response of ACx L2/3 neurons to sound presentation (4390 neurons) during
413 sitting (light gray) and running (dark gray). Stimulus duration was 1 s (gray shading). The horizontal bar
414 above the plot marks time bins in which the responses are statistically different from each other (gray:
415 n.s., black: $p<0.05$; see Methods).

416 (**B**) As in **A**, but for responses to moving gratings (3701 neurons) during sitting (light blue) and running
417 (dark blue). Stimulus duration was 4 s to 8 s (gray shading).

418

419

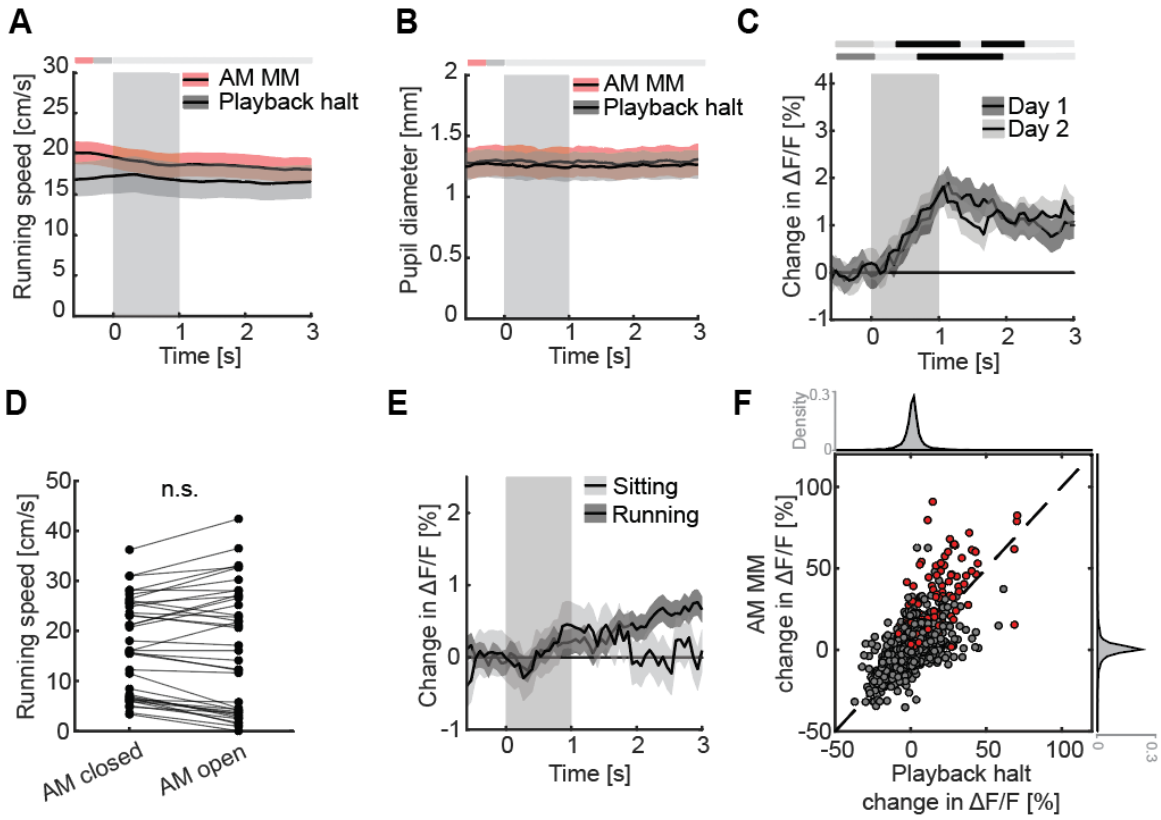


Figure S2. Controls for audiomotor mismatch responses. Related to Figure 2.

(**A**) The average running speed of mice (38 sessions, 12 mice) upon AM mismatches (red) and sound playback halts (black). AM mismatch duration was 1 s (gray shading).

(**B**) Changes in pupil diameter upon the AM mismatches (red) and sound playback halts (black) (32 sessions, 11 mice).

(**C**) The average population response of L2/3 ACx neurons to AM mismatches as a function of experience with audiomotor coupling. AM mismatch duration was 1 s (gray shading). The data shown are from the first two audiomotor closed loop sessions. Each closed loop session lasted 5.5 minutes and mice experienced one such session per day (Day 1: 1271, Day 2: 904). Note, AM mismatch responses are already present in the first closed loop session.

(**D**) Comparison of running speeds during either closed or open loop audiomotor sessions. Dots are different recording sessions. Here and elsewhere, n.s.: not significant; *: p<0.05; **: p<0.01; ***: p<0.001. See Table S1 for all statistical information.

(**E**) Comparison of the average population response of L2/3 neurons to sound playback halts while mice were running (dark gray, 4017 neurons) or sitting (light gray, 1878 neurons) in open loop session. AM mismatch duration was 1 s (gray shading).

(**F**) Scatter plot of the responses to AM mismatch and sound playback halt for all neurons. Neurons that exhibited significant ($p < 0.05$) positive responses to AM mismatch are shown in red (13.7%). Black dashed line marks unity

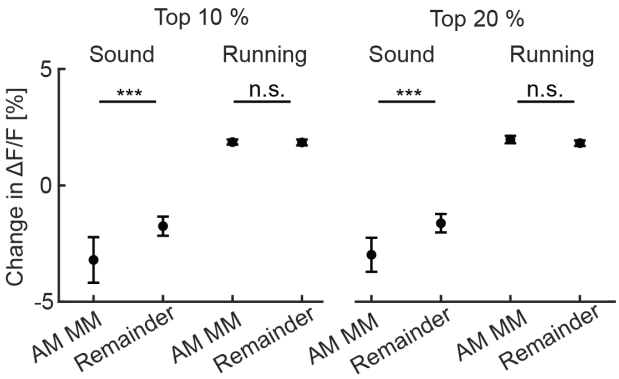
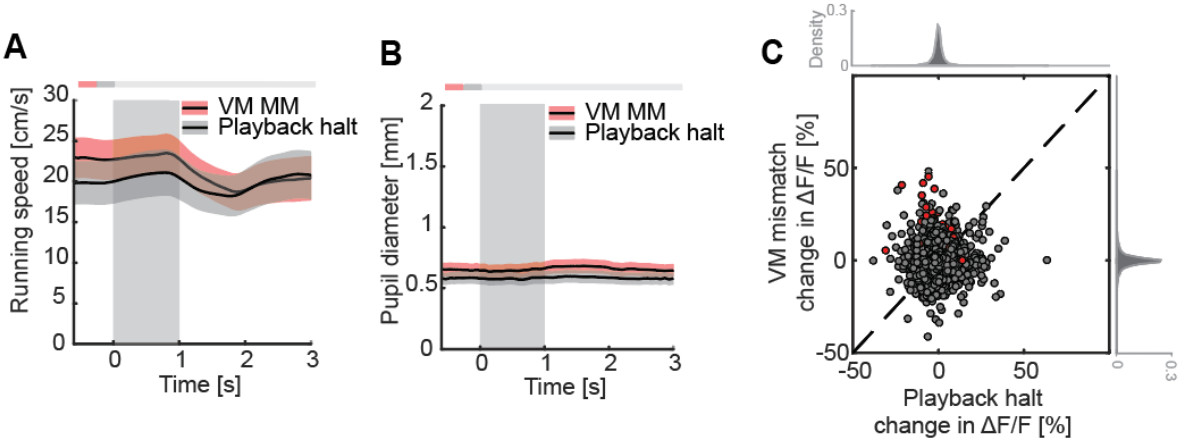


Figure S3. Opposing influence of sound and running on AM mismatch neurons. Related to Figure 2.

The same analysis presented as in **Figure 2F** but using the 10% of most AM mismatch responsive neurons (left), or the 20% of the most AM mismatch responsive neurons (right). Error bars indicate SEM. Here and elsewhere, n.s.: not significant; *: $p<0.05$; **: $p<0.01$; ***: $p<0.001$. See Table S1 for all statistical information.



450

451 **Figure S4. Controls for visuomotor mismatch responses. Related to Figure 3.**

452 (**A**) The average running speed of mice upon VM mismatches (red) and visual flow playback halts (black)
453 (47 sessions, 15 mice). VM mismatch duration was 1 s (gray shading).

454 (**B**) Changes in pupil diameter upon the VM mismatches (red) and sound playback halts (black)
455 presentation (32 sessions, 11 animals).

456 (**C**) Scatter plot of the responses to VM mismatch and visual flow playback halt for all neurons. The
457 percentage of neurons (6.6%; in red) that exhibited significant ($p < 0.05$) responses to VM mismatches, is
458 only barely above chance.

459

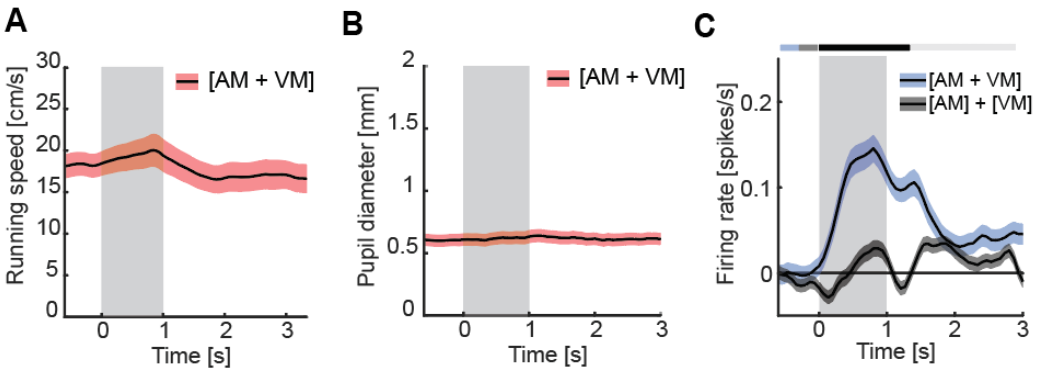


Figure S5. Controls for multimodal mismatch responses. Related to Figure 4.

(**A**) The average running speed of mice upon concurrent [AM + VM] mismatch (26 sessions, 10 mice). Gray shading marks the duration of the mismatch stimulus.

(**B**) Changes in pupil diameter upon the concurrent [AM + VM] mismatch (26 sessions, 10 mice).

(**C**) Non-linear combination of mismatch responses with spike estimation. As in **Figure 4E**, but using an estimate of firing rate calculated using CASCADE (Rupprecht et al., 2021). Gray shading marks the duration of the mismatch stimulus.

470 **METHODS**

471 **Mice and surgery.** All animal procedures were approved by and carried out in accordance with guidelines
472 of the Veterinary Department of the Canton Basel-Stadt, Switzerland. C57BL/6 female mice (Charles
473 River), between the ages of 7 and 12 weeks were used in this study. For cranial window implantation,
474 mice were anesthetized using a mixture of fentanyl (0.05 mg/kg), medetomidine (0.5 mg/kg), and
475 midazolam (5 mg/kg). Analgesics were applied perioperatively. Lidocaine was injected subcutaneously
476 into the scalp (10 mg/kg s.c.) prior to the surgery. Mice underwent a cranial window implantation surgery
477 at an age of between 7 and 8 weeks. First, a custom-made titanium head-plate was attached to the skull
478 (right hemisphere) with dental cement (Heraeus Kulzer). Next, a 3 mm craniotomy was made over left
479 ACx (4.2 mm to 4.4 mm lateral from the midline and 2.6 mm to 2.8 mm posterior from bregma) followed
480 by 4 to 6 injections of approximately 200 nl each of the AAV vector: AAV2/1-EF1α-GCaMP6f-WPRE (10^{13-}
481 ^{14} GC/ml). A circular glass cover slip was glued (Ultragel, Pattex) in place to seal the craniotomy. Metacam
482 (5 mg/kg, s.c.) and buprenorphine (0.1 mg/kg s.c.) were injected intraperitoneally for 2 days after
483 completion of the surgery. Mice were returned to their home cage and group housed for 10 days prior to
484 the first experiments.

485 **Virtual reality environment.** All recordings were done with mice head-fixed in a virtual reality system, as
486 described previously (Leinweber et al., 2014). Mice were free to run on an air-supported polystyrene ball.
487 Three types of closed loop conditions were used for the experiments. The rotation of the spherical
488 treadmill was either coupled 1: to the sound amplitude of an 8 kHz pure tone (audiomotor coupling), while
489 the animal was locomoting in near-darkness (low ambient light in the experimental rooms, primarily from
490 computer screens), 2: to the movement in a virtual corridor (visuomotor coupling), or 3: to both the sound
491 amplitude of an 8 kHz pure tone and the movement in a virtual corridor (audio-visuo-motor coupling). For
492 audiomotor coupling, we used the running speed of the mouse to control the SPL of an 8 kHz pure tone
493 presented to the mouse through a loudspeaker (see section auditory stimulation). This closed loop
494 coupling was not instantaneous but exhibited a delay of 260 ms ± 60 ms (mean ± STD). For visuomotor
495 coupling, the running speed of the mouse was coupled to the visual flow speed in the virtual environment
496 projected onto a toroidal screen surrounding the mouse using a Samsung SP-F10M projector synchronized
497 to the turnaround times of the resonant scanner of the two-photon microscope. The delay in the
498 visuomotor closed loop coupling was 90 ms ± 10 ms (mean ± STD). From the point of view of the mouse,
499 the screen covered a visual field of approximately 240 degrees horizontally and 100 degrees vertically.

500 The virtual environment presented on the screen was a corridor tunnel with walls consisting of vertical
501 sinusoidal gratings. In auditory experiments, the mouse generated additional sounds while running on an
502 air-supported treadmill (e.g., treadmill rotation, changes in airflow, footsteps), which correlated with
503 running speed intensity. Prior to the recording experiments, mice were habituated to the setup, without
504 any coupling, in 1 to 2-hour long sessions for up to 5 days, until they displayed regular locomotion. On the
505 first recording day, the mice experienced all three types of closed loop conditions (audiomotor,
506 visuomotor, or combined coupling) in a random order. Closed loop sessions were followed by open loop
507 sessions, in which rotation of the spherical treadmill was decoupled from both the sound amplitude and
508 the movement in the virtual corridor. During these open loop sessions, we replayed the amplitude
509 modulated sound or the visual flow recorded in the previous closed loop session.

510 **Auditory stimulation.** Sounds were generated with a 16-bit digital-to-analog converter (PCI6738, National
511 Instruments) using custom scripts written in LabVIEW (LabVIEW 2020, National Instruments) at 160 kHz
512 sampling rate, amplified (SA1, Tucker Davis Technologies, FL, USA) and played through an MF1 speaker
513 (Tucker Davis Technologies. FL, USA) positioned 10 cm from the mouse's right ear. Stimuli were calibrated
514 with a wide-band ultrasonic acoustic sensor (Model 378C01, PCB Piezotronics, NY, USA). To study sound-
515 evoked responses, we used 4 kHz, 8 kHz, 16 kHz, and 32 kHz pure tones played at 60 dB and 75 dB SPL (1
516 s duration, at a randomized inter-stimulus interval 4 s ± 1 s, 10 repetitions, 1 ms on and off-ramp, in a
517 randomized order). For audiomotor coupling experiments, we used an 8 kHz pure tone with a sound
518 amplitude that varied between 40 dB and 75 dB SPL.

519 **Visual stimulation.** For visual stimulation, we used full-field sinusoidal drifting grating (0 degrees, 45
520 degrees, 90 degrees, 270 degrees, moving in either direction) in a pseudo-random sequence, each
521 presented for a duration of 6 s ± 2 s, with between 2 and 7 repetitions, with a randomized inter-stimulus
522 interval of 4.5 s ± 1.5 s during which a gray screen was displayed.

523 **Running onsets.** Running onsets were defined as the running speed crossing a threshold of 3 cm/s, where
524 the average speed in the previous 3 s was below 1.8 cm/s. To separate trials with AM mismatch, VM
525 mismatch, auditory stimulus and grating stimulus based on locomotion state into those running and those
526 while sitting, we used threshold of 0.3 cm/s in a 1 s window preceding the stimulus onset.

527 **Widefield calcium imaging.** To establish a reference tonotopic map of A1 and AAF (**Figure 1E**), we
528 performed widefield fluorescence imaging experiments on a custom-built microscope consisting of

529 objectives mounted face-to-face (Nikon 85 mm/f1.8 sample side, Nikon 50 mm/f1.4 sensor side), as
530 previously described (Heindorf and Keller, 2023). Blue illumination was provided by a light-emitting diode
531 (470 nm, Thorlabs) and passed through an excitation filter (SP490, Thorlabs). Green fluorescence emission
532 was filtered with a 525/50 bandpass filter. Images were acquired at a frame rate of 100 Hz on a sCMOS
533 camera (PCO edge 4.2). The raw images were cropped on-sensor, and the resulting data was saved to disk
534 with custom-written software in LabVIEW (National Instruments).

535 **Two-photon imaging.** Calcium imaging of L2/3 neurons in A1 and AAF was performed using a modified
536 Thorlabs Bergamo II microscope with a 16x, 0.8 NA objective (Nikon N16XLWD-PF), as previously
537 described (Leinweber et al., 2014). To record in left ACx, the microscope was tilted 45 degrees to the left.
538 The excitation light source was a tunable, femtosecond-pulsed laser (Insight, Spectra Physics or
539 Chameleon, Coherent) tuned to 930 nm. The laser power was adjusted to 30 mW. A 12 kHz resonance
540 scanner (Cambridge Technology) was used for line scanning, and we acquired 400 lines per frame. This
541 resulted in a frame rate of 60 Hz at a resolution of 400 × 750 pixels. We used a piezo-electric linear actuator
542 (Physik Instrumente, P-726) to record from imaging planes at four different cortical depths, separated by
543 15 µm. This reduced the effective frame rate per layer to 15 Hz. The emission light was bandpass filtered
544 using a 525/50 nm filter (Semrock), and signals were detected with a photomultiplier (Hamamatsu,
545 H7422), amplified (Femto, DHCPCA-100), digitized at 800 MHz (National Instruments, NI5772), and
546 bandpass filtered at 80 MHz with a digital Fourier-transform filter on a field-programmable gate array
547 (National Instruments, PXIe-7965). Recording locations were visually registered against the reference
548 images acquired with widefield imaging previously using blood vessels patterns.

549 **Widefield image analysis.** Off-line data processing and data analysis were done with custom-written
550 MATLAB scripts. Slow drifts in the fluorescence signal were removed using 8th percentile filtering with a
551 62.5 s moving window, similar to what was used for two-photon imaging data (Dombeck et al., 2007).
552 Activity was calculated as the $\Delta F/F_0$, where F_0 was the median fluorescence over the entire recording
553 session. For stimulus responses, we use a response window of 0.2 s to 1.2 s following stimulus onset and
554 a baseline window of -1 s to 0 s before stimulus onset. The pixels with the strongest response (top 3% -
555 5% of response distribution), were used to mark the tonotopic areas corresponding to the different
556 stimuli.

557 **Two-photon image analysis.** Calcium imaging data were processed as described previously. In brief, raw
558 images were full-frame registered to correct for lateral brain motion. Neurons were selected manually

559 based on mean and maximum fluorescence images. Average fluorescence per neuron over time was
560 corrected for slow fluorescence drift using an 8th percentile filter and a 66 s (or 1000 frames) window
561 (Dombeck et al., 2007; Keller et al., 2012; Leinweber et al., 2014) and divided by the median value over
562 the entire trace to calculate $\Delta F/F_0$. All stimulus-response curves were baseline subtracted. The baseline
563 subtraction window was -0.5 s to 0 s before stimulus onset. For quantification of responses during
564 different onset types (auditory, visual, running, mismatch), $\Delta F/F$ was averaged over the response time
565 window (0.5 s to 2.5 s after stimulus onset) and baseline subtracted (mean activity in a window preceding
566 stimulus onset, -0.5 s to 0 s). Onsets which were not preceded by at least 2 s of baseline or not followed
567 by at least 3 s of recording time, were excluded from the analysis. Sessions with less than two onsets were
568 not included in the analysis. To quantify the difference in average calcium responses as a function of time,
569 we used a hierarchical bootstrap test for every 5 frames of the calcium trace (333 ms) and marked
570 comparisons where responses were different ($p < 0.05$). Mismatch responsive neurons were selected
571 based on the absolute response strength over the response time window (0.5 s to 2.5 s). To infer spikes
572 from calcium signals (**Figure S5C**), we used CASCADE (Rupprecht et al., 2021).

573 **Statistical tests.** All statistical information for the tests performed in this manuscript is provided in Table
574 S1. We used hierarchical bootstrapping (Saravanan et al., 2020) for statistical testing to account for the
575 nested structure of the data (multiple neurons from one imaging site). We first resampled the data with
576 replacement at the level of imaging sites, followed by resampling at the level of neurons. We then
577 computed the mean responses across the resampled population and repeated this process 10 000 times.
578 The probability of one group being different from the other was calculated as a fraction of bootstrap
579 sample means which violated the tested hypothesis.

580

Table S1. Statistics

All information on statistical tests used in this manuscript are shown in Table S1 below. We used hierarchical bootstrap (Saravanan et al., 2020) or a correlation coefficient for all comparisons.

Figure	Comparison	Test	N_1	N_2	Unit	P-value
Figure 1H	Sound evoked response vs 0	Hierarchical bootstrap	7637 (60 fields of view, 17 mice)	-	Neurons	gray: n.s, black: p<0.05).
Figure 1K	Grating evoked response vs 0	Hierarchical bootstrap	7637 (60 fields of view, 17 mice)	-	Neurons	gray: n.s, black: p<0.05).
Figure 1N	Running onset response vs 0	Hierarchical bootstrap	7637 (60 fields of view, 17 mice)	-	Neurons	gray: n.s, black: p<0.05).
Figure 2D	AM mismatch vs sound playback halt response	Hierarchical bootstrap	4755 (38 fields of view, 12 mice)	4755 (38 fields of view, 12 mice)	Neurons	gray: n.s, black: p<0.05).
Figure 2E	Sound response vs running onset response of AM mismatch responsive neurons	-	238 (33 fields of view, 12 mice)	238 (33 fields of view, 12 mice)	Neurons	-
Figure 2F	Sound response of AM mismatch responsive neurons vs that of the remainder of the population	Hierarchical bootstrap	238 (33 fields of view, 12 mice)	4517 (38 fields of view, 12 mice)	Neurons	0.0091
	Running onset response of AM mismatch responsive neurons vs that of the remainder of the population	Hierarchical bootstrap	238 (33 fields of view, 12 mice)	4517 (38 fields of view, 12 mice)	Neurons	0.32
Figure 2H	Running onset response in closed vs open loop session	Hierarchical bootstrap	2407 neurons (21 fields of view, 10 mice)	2407 neurons (21 fields of view, 10 mice)	Neurons	gray: n.s, black: p<0.05).
Figure 3D	VM mismatch vs playback halt response	Hierarchical bootstrap	5688 (47 fields of view, 15 mice)	5688 (47 fields of view, 15 mice)	Neurons	gray: n.s, black: p<0.05).
Figure 3E	Grating response vs running onset response	-	284 (43 fields of view, 15 mice)	284 (43 fields of view, 15 mice)	Neurons	-

	of VM mismatch responsive neurons					
Figure 3F	Grating response of VM mismatch responsive neurons vs that of the remainder of the population	Hierarchical bootstrap	284 (43 fields of view, 15 mice)	5404 (47 fields of view, 15 mice)	Neurons	0.23
	Running onset response of VM mismatch responsive neurons vs that of the remainder of the population	Hierarchical bootstrap	284 (43 fields of view, 15 mice)	5404 (47 fields of view, 15 mice)	Neurons	0.45
Figure 3H	Running onset response in closed vs open loop session	Hierarchical bootstrap	2316 (17 fields of view, 9 mice)	2316 (17 fields of view, 9 mice)	Neurons	gray: n.s, black: $p<0.05$).
Figure 4C	Response to concurrent [AM + VM] mismatches vs 0	Hierarchical bootstrap	3289 (26 fields of view, 10 mice)	3289 (26 fields of view, 10 mice)	Neurons	gray: n.s, black: $p<0.05$).
Figure 4D	Response to concurrent [AM + VM] mismatches vs response to AM mismatches	Hierarchical bootstrap	3289 (26 fields of view, 10 mice)	3289 (26 fields of view, 10 mice)	Neurons	gray: n.s, black: $p<0.05$).
	Response to concurrent [AM + VM] mismatches vs response to VM mismatches	Hierarchical bootstrap	3289 (26 fields of view, 10 mice)	3289 (26 fields of view, 10 mice)	Neurons	gray: n.s, black: $p<0.05$).
Figure 4E	Response to concurrent [AM + VM] mismatches vs the linear sum of the responses evoked by [AM] and [VM] mismatches presented in isolation	Hierarchical bootstrap	3289 (26 fields of view, 10 mice)	3289 (26 fields of view, 10 mice)	Neurons	gray: n.s, black: $p<0.05$).
Figure 4F	Correlation of the responses to the concurrent [AM+VM] mismatches against the linear sum of the responses evoked by [AM] and [VM] mismatches presented in isolation	Correlation	3289 (26 fields of view, 10 mice)	3289 (26 fields of view, 10 mice)	Neurons	$p < 0.001$ $r = 0.31$

Figure S1A	Sound evoked response during sitting vs running	Hierarchical Bootstrap	4390 (33 fields of view, 14 mice)	4390 (33 fields of view, 14 mice)	Neurons	gray: n.s, black: p<0.05).
Figure S1B	Grating evoked response during sitting vs running	Hierarchical Bootstrap	3701 (28 fields of view, 10 mice)	3701 (28 fields of view, 10 mice)	Neurons	gray: n.s, black: p<0.05).
Figure S2A	Running speed upon AM mismatch vs sound playback halt	Hierarchical Bootstrap	38 sessions (12 mice)	38 sessions (12 mice)	Sessions	gray: n.s, black: p<0.05).
Figure S2B	Pupil diameter upon AM mismatch vs sound playback halt	Hierarchical Bootstrap	32 sessions (11 mice)	32 sessions (11 mice)	Sessions	gray: n.s, black: p<0.05).
Figure S2C	AM mismatch on the 1st recording day vs 0	Hierarchical Bootstrap	1271 (12 fields of view, 12 mice)	-	Neurons	gray: n.s, black: p<0.05).
	AM mismatch on the 2nd recording day vs 0	Hierarchical Bootstrap	904 (10 fields of view, 10 mice)	-	Neurons	gray: n.s, black: p<0.05).
Figure S2D	Running speed in AM closed loop sessions vs that in open loop	Hierarchical Bootstrap	38 sessions (12 mice)	38 sessions (12 mice)	Sessions	0.47
Figure S2E	Sound playback halt response during sitting vs running	-	4017 (33 fields of view, 12 mice)	1878 (17 fields of view, 9 mice)	Neurons	-
Figure S2F	Correlation of the responses to AM mismatch and sound playback halt	Correlation	4755 (38 fields of view, 12 mice)	4755 (38 fields of view, 12 mice)	Neurons	$p < 0.001$ $r = 0.66$
Figure S3	Response to sound of the top 10% AM mismatch responsive neurons vs that of the remainder of the population	Hierarchical bootstrap	476 (36 fields of view, 12 mice)	4279 (38 fields of view, 12 mice)	Neurons	0.0004
	Response to running onsets of the top 10% AM mismatch responsive neurons vs that of the remainder of the population	Hierarchical bootstrap	476 (36 fields of view, 12 mice)	4279 (38 fields of view, 12 mice)	Neurons	0.48

	Response to sound of the top 20% AM mismatch responsive neurons vs that of the remainder of the population	Hierarchical bootstrap	951 (38 fields of view, 12 mice)	3804 (38 fields of view, 12 mice)	Neurons	0.0002
	Response to running onsets of the top 20% AM mismatch responsive neurons vs that of the remainder of the population	Hierarchical bootstrap	951 (38 fields of view, 12 mice)	3804 (38 fields of view, 12 mice)	Neurons	0.38
Figure S4A	Running speed upon VM mismatch vs playback halt	Hierarchical bootstrap	47 sessions (15 mice)	47 sessions (15 mice)	Sessions	gray: n.s, black: p<0.05).
Figure S4B	Pupil diameter upon VM mismatch vs playback halt	Hierarchical bootstrap	32 sessions (11 mice)	32 sessions (11 mice)	Sessions	gray: n.s, black: p<0.05).
Figure S4C	Correlation of the responses to VM mismatch and visual flow playback halt	Correlation	5688 (47 fields of view, 15 mice)	5688 (47 fields of view, 15 mice)	Neurons	p = 0.41 r = 0.01
Figure S5A	Running speed upon concurrent [AM + VM] mismatch	-	26 sessions (10 mice)	26 sessions (10 mice)	Sessions	-
Figure S5B	Pupil diameter upon concurrent [AM + VM] mismatch	-	26 sessions (10 mice)	26 sessions (10 mice)	Sessions	-
Figure S5C	Response to concurrent [AM + VM] mismatches vs the linear sum of the responses evoked by [AM] and [VM] mismatches presented in isolation	Hierarchical bootstrap	3289 (26 fields of view, 10 mice)	3289 (26 fields of view, 10 mice)	Neurons	gray: n.s, black: p<0.05).

588 **Key Resource Table**

REAGENT or RESOURCE	Source	Identifier
Bacterial and virus strains		
AAV2/1-EF1α-GCaMP6f-WPRE (10^{14} GC/ml)	FMI vector core	vector.fmi.ch
Chemicals, peptides, and recombinant proteins		
Fentanyl citrate	Actavis	CAS 990-73-8
Midazolam (Dormicum)	Roche	CAS 59467-96-8
Medetomidine (Domitor)	Orion Pharma	CAS 86347-14-0
Ropivacaine	Presenius Kabi	CAS 132112-35-7
Lidocaine	Bichsel	CAS 137-58-6
Buprenorphine	Reckitt Benckiser Healthcare	CAS 52485-79-7
Ophthalmic gel (Humigel)	Virbac	N/A
Flumazenil (Anexate)	Roche	CAS 78755-81-4
Atipamezole (Antisedan)	Orion Pharma	CAS 104054-27-5
N-Butyl-2-cyanoacrylate (Histoacryl)	Braun	CAS 6606-65-1
Dental cement (Paladur)	Heraeus Kulzer	CAS 9066-86-8
Meloxicam (Metacam)	Boehringer Ingelheim	CAS 71125-39-8
Deposited data		
All data and code used to generate manuscript figures	This paper	data.fmi.ch
Experimental models: Organisms/strains		
Mus musculus: C57BL/6	Charles River	N/A
Software and algorithms		
MATLAB (2021b)	The MathWorks	RRID: SCR_001622
LabVIEW	National Instruments	RRID: SCR_014325
Python	python.org	RRID: SCR_008394
Panda3D	panda3d.org	N/A

589

590 **ACKNOWLEDGEMENTS**

591 We thank Tingjia Lu for the production of viral vectors and all the members of the Keller lab for discussion
592 and support. This project has received funding from the Swiss National Science Foundation (GBK), the
593 Novartis Research Foundation (GBK), and the European Research Council (ERC) under the European
594 Union's Horizon 2020 research and innovation programme (grant agreement No 865617) (GBK).

595 **AUTHOR CONTRIBUTIONS**

596 MS designed and performed the experiments and analyzed the data. All authors wrote the manuscript.

597 **DECLARATION OF INTERESTS**

598 The authors declare no competing financial interests.

REFERENCES

Attinger, A., Wang, B., Keller, G.B., 2017. Visuomotor Coupling Shapes the Functional Development of Mouse Visual Cortex. Cell 169, 1291-1302.e14. https://doi.org/10.1016/j.cell.2017.05.023

Audette, N.J., Schneider, D.M., 2023. Stimulus-specific prediction error neurons in mouse auditory cortex. https://doi.org/10.1101/2023.01.06.523032

Audette, N.J., Zhou, W., La Chioma, A., Schneider, D.M., 2022. Precise movement-based predictions in the mouse auditory cortex. Curr Biol 32, 4925-4940.e6. https://doi.org/10.1016/j.cub.2022.09.064

Ayaz, A., Stäuble, A., Hamada, M., Wulf, M.A., Saleem, A.B., Helmchen, F., 2019. Layer-specific integration of locomotion and sensory information in mouse barrel cortex. Nature Communications 10, 1–14. https://doi.org/10.1038/s41467-019-10564-8

Bigelow, J., Morrill, R.J., Dekloe, J., Hasenstaub, A.R., 2019. Movement and VIP Interneuron Activation Differentially Modulate Encoding in Mouse Auditory Cortex. eNeuro 6. https://doi.org/10.1523/ENEURO.0164-19.2019

Bigelow, J., Morrill, R.J., Olsen, T., Hasenstaub, A.R., 2022. Visual modulation of firing and spectrotemporal receptive fields in mouse auditory cortex. Current Research in Neurobiology 3, 100040. https://doi.org/10.1016/j.crneur.2022.100040

Clavagnier, S., Falchier, A., Kennedy, H., 2004. Long-distance feedback projections to area V1: Implications for multisensory integration, spatial awareness, and visual consciousness. Cognitive, Affective, & Behavioral Neuroscience 4, 117–126. https://doi.org/10.3758/CABN.4.2.117

Dombeck, D.A., Khabbaz, A.N., Collman, F., Adelman, T.L., Tank, D.W., 2007. Imaging large-scale neural activity with cellular resolution in awake, mobile mice. Neuron 56, 43–57. https://doi.org/10.1016/j.neuron.2007.08.003

Eliades, S.J., Wang, X., 2008. Neural substrates of vocalization feedback monitoring in primate auditory cortex. Nature 453, 1102–6. https://doi.org/10.1038/nature06910

Falchier, A., Clavagnier, S., Barone, P., Kennedy, H., 2002. Anatomical evidence of multimodal integration in primate striate cortex. J Neurosci 22, 5749–5759. https://doi.org/10.1523/JNEUROSCI.22-13-05749.2002

Garner, A.R., Keller, G.B., 2022. A cortical circuit for audio-visual predictions. Nat Neurosci 25, 98–105. https://doi.org/10.1038/s41593-021-00974-7

Han, S., Helmchen, F., 2023. Behavior-relevant top-down cross-modal predictions in mouse neocortex. https://doi.org/10.1101/2023.04.03.535389

Heindorf, M., Arber, S., Keller, G.B., 2018. Mouse Motor Cortex Coordinates the Behavioral Response to Unpredicted Sensory Feedback. Neuron 99, 1040-1054.e5. https://doi.org/10.1016/j.neuron.2018.07.046

Heindorf, M., Keller, G.B., 2023. Antipsychotic drugs selectively decorrelate long-range interactions in deep cortical layers. https://doi.org/10.1101/2022.01.31.478462

Henschke, J.U., Price, A.T., Pakan, J.M.P., 2021. Enhanced modulation of cell-type specific neuronal responses in mouse dorsal auditory field during locomotion. Cell Calcium 96, 102390. https://doi.org/10.1016/j.ceca.2021.102390

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

Jordan, R., Keller, G.B., 2023. The locus coeruleus broadcasts prediction errors across the cortex to promote sensorimotor plasticity. eLife 12. https://doi.org/10.7554/eLife.85111

Jordan, R., Keller, G.B., 2020. Opposing Influence of Top-down and Bottom-up Input on Excitatory Layer 2/3 Neurons in Mouse Primary Visual Cortex. Neuron 108, 1194-1206.e5. https://doi.org/10.1016/j.neuron.2020.09.024

Keller, G.B., Bonhoeffer, T., Hübener, M., 2012. Sensorimotor Mismatch Signals in Primary Visual Cortex of the Behaving Mouse. Neuron 74, 809–815. https://doi.org/10.1016/j.neuron.2012.03.040

Keller, G.B., Hahnloser, R.H.R., 2009. Neural processing of auditory feedback during vocal practice in a songbird. Nature 457, 187–90. https://doi.org/10.1038/nature07467

Kim, J.-H., Jung, A.-H., Jeong, D., Choi, I., Kim, K., Shin, S., Kim, S.J., Lee, S.-H., 2016. Selectivity of Neuromodulatory Projections from the Basal Forebrain and Locus Ceruleus to Primary Sensory Cortices. J. Neurosci. 36, 5314–5327. https://doi.org/10.1523/JNEUROSCI.4333-15.2016

Leinweber, M., Ward, D.R., Sobczak, J.M., Attinger, A., Keller, G.B., 2017. A Sensorimotor Circuit in Mouse Cortex for Visual Flow Predictions. Neuron 95, 1420-1432.e5. https://doi.org/10.1016/j.neuron.2017.08.036

Leinweber, M., Zmarz, P., Buchmann, P., Argast, P., Hübener, M., Bonhoeffer, T., Keller, G.B., 2014. Two-photon calcium imaging in mice navigating a virtual reality environment. Journal of visualized experiments : JoVE e50885.

Liu, J., Kanold, P.O., 2022. Interactive auditory task reveals complex sensory-action integration in mouse primary auditory cortex. https://doi.org/10.1101/2022.12.12.520155

Markov, N.T., Ercsey-Ravasz, M., Van Essen, D.C., Knoblauch, K., Toroczkai, Z., Kennedy, H., 2013. Cortical high-density counterstream architectures, Science. American Association for the Advancement of Science. https://doi.org/10.1126/science.1238406

McGinley, M.J., David, S.V., McCormick, D.A., 2015. Cortical Membrane Potential Signature of Optimal States for Sensory Signal Detection. Neuron 87, 179–192. https://doi.org/10.1016/j.neuron.2015.05.038

Morandell, K., Yin, A., Del Rio, R.T., Schneider, D.M., 2023. Movement-related modulation in mouse auditory cortex is widespread yet locally diverse. bioRxiv 2023.07.03.547560. https://doi.org/10.1101/2023.07.03.547560

Morrill, R.J., Hasenstaub, A.R., 2018. Visual Information Present in Infragranular Layers of Mouse Auditory Cortex. J Neurosci 38, 2854–2862. https://doi.org/10.1523/JNEUROSCI.3102-17.2018

Niell, C.M., Stryker, M.P., 2010. Modulation of visual responses by behavioral state in mouse visual cortex. Neuron 65, 472–9. https://doi.org/10.1016/j.neuron.2010.01.033

O'Toole, S.M., Oyibo, H.K., Keller, G.B., 2023. Molecularly targetable cell types in mouse visual cortex have distinguishable prediction error responses. Neuron 111, 2918-2928.e8. https://doi.org/10.1016/j.neuron.2023.08.015

Rao, R.P.N., Ballard, D.H., 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nature Neuroscience 2, 79–87. https://doi.org/10.1038/4580

Rupprecht, P., Carta, S., Hoffmann, A., Echizen, M., Blot, A., Kwan, A.C., Dan, Y., Hofer, S.B., Kitamura, K., Helmchen, F., Friedrich, R.W., 2021. A database and deep learning toolbox for noise-optimized, generalized spike inference from calcium imaging. Nat Neurosci 24, 1324–1337. https://doi.org/10.1038/s41593-021-00895-5

Saravanan, V., Berman, G.J., Sober, S.J., 2020. Application of the hierarchical bootstrap to multi-level data in neuroscience. Neuron Behav Data Anal Theory 3, https://nbdt.scholasticahq.com/article/13927-application-of-the-hierarchical-bootstrap-to-multi-level-data-in-neuroscience.

Schneider, D.M., Nelson, A., Mooney, R., 2014. A synaptic and circuit basis for corollary discharge in the auditory cortex. Nature 513, 189–194. https://doi.org/10.1038/nature13724

Schneider, D.M., Sundararajan, J., Mooney, R., 2018. A cortical filter that learns to suppress the acoustic consequences of movement. Nature 561, 391–395. https://doi.org/10.1038/s41586-018-0520-5

Sharma, S., Srivastava, H.K., Bandyopadhyay, S., 2021. Modulation of auditory responses by visual inputs in the mouse auditory cortex. https://doi.org/10.1101/2021.01.22.427870

Shiramatsu, T.I., Mori, K., Ishizu, K., Takahashi, H., 2021. Auditory, Visual, and Cross-Modal Mismatch Negativities in the Rat Auditory and Visual Cortices. Front Hum Neurosci 15, 721476. https://doi.org/10.3389/fnhum.2021.721476

Stanley, J., Miall, R.C., 2007. Functional activation in parieto-premotor and visual areas dependent on congruency between hand movement and visual stimuli during motor-visual priming. NeuroImage 34, 290–9. https://doi.org/10.1016/j.neuroimage.2006.08.043

St-Yves, G., Allen, E.J., Wu, Y., Kay, K., Naselaris, T., 2023. Brain-optimized deep neural network models of human visual areas learn non-hierarchical representations. Nat Commun 14, 3329. https://doi.org/10.1038/s41467-023-38674-4

Suzuki, M., Pennartz, C.M.A., Aru, J., 2023. How deep is the brain? The shallow brain hypothesis. Nat. Rev. Neurosci. 1–14. https://doi.org/10.1038/s41583-023-00756-z

Vasilevskaya, A., Widmer, F.C., Keller, G.B., Jordan, R., 2023. Locomotion-induced gain of visual responses cannot explain visuomotor mismatch responses in layer 2/3 of primary visual cortex. Cell Rep 42, 112096. https://doi.org/10.1016/j.celrep.2023.112096

Vivaldo, C.A., Lee, J., Shorkey, M., Keerthy, A., Rothschild, G., 2023. Auditory cortex ensembles jointly encode sound and locomotion speed to support sound perception during movement. PLOS Biology 21, e3002277. https://doi.org/10.1371/journal.pbio.3002277

Widmer, F.C., O'Toole, S.M., Keller, G.B., 2022. NMDA receptors in visual cortex are necessary for normal visuomotor integration and skill learning. eLife 11, e71476. https://doi.org/10.7554/eLife.71476

Yavorska, I., Wehr, M., 2021. Effects of Locomotion in Auditory Cortex Are Not Mediated by the VIP Network. Front Neural Circuits 15, 618881. https://doi.org/10.3389/fncir.2021.618881

Yogesh, B., Keller, G.B., 2023. Cholinergic input to mouse visual cortex signals a movement state and acutely enhances layer 5 responsiveness. eLife 12. https://doi.org/10.7554/eLife.89986

Zhao, M., Ren, M., Jiang, T., Jia, X., Wang, X., Li, A., Li, X., Luo, Q., Gong, H., 2022. Whole-Brain Direct Inputs to and Axonal Projections from Excitatory and Inhibitory Neurons in the Mouse Primary Auditory Area. Neurosci Bull 38, 576–590. https://doi.org/10.1007/s12264-022-00838-5

Zhou, M., Liang, F., Xiong, X.R., Li, L., Li, H., Xiao, Z., Tao, H.W., Zhang, L.I., 2014. Scaling down of balanced excitation and inhibition by active behavioral states in auditory cortex. Nat Neurosci 17, 841–850. https://doi.org/10.1038/nn.3701

Zmarz, P., Keller, G.B., 2016. Mismatch Receptive Fields in Mouse Visual Cortex. Neuron 92, 766–772. https://doi.org/10.1016/j.neuron.2016.09.057