Seeing what you hear: sound intensity alters rat visual perception in a temporal frequency classification task

Mattia Zanzi*, Francesco Guido Rinaldi*, Silene Fornasaro, Eugenio Piasini# and Davide Zoccolan#

Neuroscience Area, International School for Advanced Studies (SISSA), Trieste, 34136, Italy

* *These authors contributed equally to this work*
# *Corresponding authors*

Abstract

While integration of multimodal sensory stimuli in cortical hierarchies is well studied, the functional impact of hetero-modal inputs on visual perception is less explored. Here we use a visual classification task in rats to investigate how task-irrelevant sounds modify visual processing. Sound intensity, but not temporal congruency with visual stimuli, effectively compresses the task's visual perceptual space, suggesting inhibition as a mediator of auditory-visual interactions in neural representations.

Main text

In the cerebral cortex, integration of information from different sensory modalities takes place at multiple stages along the sensory pathways. The process that is best understood is the way unisensory cues can be combined by high-order association cortices, often in a statistically optimal way, to increase the accuracy of perceptual judgements[1–7]. By comparison, the functional role played by hetero-modal modulation of primary sensory cortices is less well established.

For instance, several authors have found direct cortico-cortical connections from primary auditory (A1) to visual (V1) cortex in rodents[8–11], but studies disagree on the impact of these projections on local cortical dynamics and encoding of visual stimuli. Sound was originally reported to hyperpolarize supragranular pyramidal cells in mouse V1, by recruiting a translaminar GABAergic network activated by A1 projections to infragranular V1 neurons[9]. Later studies, however, found that most A1 projections terminate in superficial layers of V1 and engage local inhibitory and disinhibitory circuits underlying a much richer variety of auditory influences on V1: enhancement and suppression of visually evoked responses have both been documented, as well as sharpening of orientation tuning and improvement in the coding of visual

information[8,10,12,13]. This heterogeneous assortment of sound-mediated modulations in V1 seems to depend on factors such as the contrast of the visual stimuli, the luminosity of the environment, the spectral (e.g., pure tones vs. noise bursts) and envelope (e.g., loud vs. quiet onset) properties of the sounds and the temporal congruency between visual and auditory stimuli. Finally, it remains highly debated the extent to which, in awake mice, sound-driven responses in V1 reflect auditory inputs or, rather, behavioral modulation via sound-evoked orofacial movements[11,12,14].

Amid such contrasting findings at the cortical circuitry level, the functional impact of auditory signals on rodent visual perception is poorly explored. Does sound improve acuity in visual discrimination tasks? Can any signature of sound-induced suppression (or enhancement) of visual representations be found at the behavioral level? Is the temporal consistency between sounds and visual stimuli important in this regard?

To address these questions, we trained 10 rats in a visual temporal frequency (TF) classification task, where the visual (V) stimuli to categorize were paired with simultaneous but task-irrelevant auditory (A) stimuli (Fig. 1a; see also Extended Data Fig. 1). The visual stimuli were circular sinusoidal gratings moving outward with 9 possible, randomly selected TFs (Fig. 1b; x axis). Rats categorized these stimuli as "Low TF" or "High TF", relative to an ambiguous (randomly rewarded) boundary value (2.12 Hz) in the middle of the tested TF range (vertical dashed lines in Fig. 1c and d). The auditory stimuli consisted of either fixed amplitude or amplitude modulated (AM) white noise bursts (Extended Data Fig. 2a). The sounds were uninformative about the correct classification of the visual stimuli, to avoid high-level multisensory integration effects and focus instead on the direct impact of auditory signals on the representation and perception of visual stimuli.

During the training phase, the gratings were paired with the constant amplitude noise burst. Rats were gradually introduced to the TFs they had to classify, starting from the most discriminable ones (0.25 vs. 4 Hz), until they reached an average performance higher than 70% correct over the non-ambiguous TF values (Extended Data Fig. 3). The contrast of the gratings was initially set at 100%, but it was progressively reduced to 25%, given that the impact of sound on V1 responses has been found to be stronger when visual stimuli are shown at low/intermediate contrast[8,12]. Once the animals had become proficient in this task (Extended Data Fig. 3), they were moved to the testing phase, where they faced a rich variety of audiovisual stimuli (illustrated in Fig. 1b). The gratings were paired not only with the fixed amplitude burst used during training (Fig. 1b, green cells), but also with AM white noise stimuli. These sounds were obtained by modulating the amplitude of a white noise burst with sinusoids having 9 possible TFs (the same of the visual gratings). The rats were presented with all possible pairwise combinations of visual and auditory TFs (cells in the squared matrix of Fig. 1b). This yielded trials where the auditory and visual stimuli had either the same TF (blue cells) or different TFs (gray and red cells). The latter included a special pool of anti-congruent conditions (red cells), where progressively larger visual TFs were paired with progressively smaller auditory TFs. Additionally, we also tested unimodal, purely visual

trials (purple cells) and unimodal, purely auditory trials with AM sounds (cyan cells). Since the rats were never trained to classify the TFs of the auditory stimuli, with the latter trials we simply measured the spontaneous responses of the animals to the AM sounds, without providing any feedback about their choices (see Methods). The rats did not spontaneously classify the AM bursts according to their TF in the auditory only conditions, thus confirming the task-irrelevance of the sounds (Extended Data Fig. 4).

Overall, this design allowed probing the impact of task-irrelevant auditory stimuli on visual perception along two distinct dimensions: 1) whether sound alters the sensitivity of visual discrimination or, rather, introduces a bias in visual classification; and 2) whether the impact of sound depends on the temporal consistency of audio-visual stimuli or, rather, on the intensity of the auditory stimuli.

When the gratings were paired with the fixed amplitude noise, the group average psychometric curve reporting the proportion of "high TF" choices as a function of the grating frequency was sharp and symmetrical, with the point of subjective equality sitting squarely on the ambiguous TF (Fig. 1c, green). Interestingly, pairing the gratings with either the congruent (blue) or anti-congruent (red) AM sounds produced an identical increase in the proportion of "high TF" choices, which was particularly prominent for the highest TFs. This observation was confirmed by a 2-way ANOVA having as factors the audiovisual experimental condition and the grating TF (see Methods for details). Both factors significantly modulated rat choices and so did their interaction (Extended Data Table 1), thus confirming that the increase of "high TF" choices yielded by the modulated sounds was not homogeneous along the TF axis. A post-hoc Tukey test confirmed that the curves observed for the two conditions featuring AM sounds were statistically indistinguishable. This indicated that the congruency between the TF of the AM sounds and the TF of the gratings did not affect the impact of the auditory stimuli on the visual discrimination task.

This observation led us to group together all the trials where the visual stimuli were paired with an AM sound, regardless of the congruency of their TFs. The resulting psychometric curve (Fig. 1d, black) featured an asymmetrical vertical shift, with respect to the reference curve measured with the fixed amplitude sounds (green dots/line), that was equivalent to the one previously observed with the congruent and anti-congruent AM sounds (compare to the blue and red curves in Fig. 1c). Strikingly, a similar but even more prominent shift of the psychometric curve was observed for the unimodal visual conditions, when the gratings were presented without any concomitant sound (purple dots/line). Again, a two-way ANOVA showed a significant main effect of experimental condition, TF of the gratings and their interaction, and a post-hoc Tukey test revealed a significant difference between the curve obtained with the unimodal visual stimuli and the curves measured with the paired sounds, either AM or fixed amplitude (Extended Data Table 2).

These results showed that rat perception of visual temporal frequencies was systematically shifted towards reporting the "high-TF" category if the noise bursts

became amplitude modulated and, even more so, when the sounds disappeared altogether. Wondering about the possible cause of these progressively larger shifts, we realized that one key attribute of the sound did decrease from fixed amplitude to AM bursts and then further to the purely visual conditions: its intensity. In fact, because of the sinusoidal modulation, there were moments during the AM bursts where the instantaneous sound intensity dropped to zero. As a result, the average sound intensity (as computed over the inferred reaction time of the animals) was lower for the AM bursts than for the fixed amplitude burst (Extended Data Fig. 2), besides being minimal for the purely visual stimuli (corresponding, in this case, to the environmental background noise). In a previous study[9], noise bursts were reported to hyperpolarize V1 neurons, with the magnitude of this inhibition increasing monotonically with their intensity. Given this observation, we hypothesized that the perceptual effect of sound is to compress visual cortical representations in a way that is proportional to its intensity.

To test this hypothesis, we developed a Bayesian ideal observer model of rat perceptual choices, resting on three assumptions: 1) that V1 neurons respond to progressively larger TFs with approximately linearly increasing firing rates, at least within the TF range tested in our study (cartoon in Fig. 2a; cyan vs. red dots); 2) that sound dampens neuronal responses, compressing them towards the origin of the V1 representational space, with the compression being stronger for higher intensity sounds (compare the green to the purple distributions in Fig. 2a); and 3) that downstream decision neurons learn an optimal discrimination boundary in the representational space under the sensory conditions they were repeatedly exposed to during training (visual gratings paired with the fixed amplitude sound; green distributions in Fig. 2a) and keep it fixed when new conditions (e.g., purely visual gratings; purple distributions) are interleaved with those used for training.

In our model, on each trial, the combination of a visual and an auditory stimulus generates an internal representation sampled from a Normal distribution with fixed standard deviation σ, centered on some average representation of the TF value of the visual stimulus s. Following from the assumptions above, the internal representation is then scaled by a factor γ (see Methods), and the decision rule is the one that would be Bayes-optimal in the training condition. We can thus derive the probability for a rat of classifying a given stimulus s as belonging to the H (i.e., “High TFs”) category as

$$p(\text{report } H|s) = \phi\left(\frac{\gamma s - s_0}{\sigma}\right),$$

where ϕ is the standard Normal cumulative function and s_0 is the boundary 2.12 Hz visual stimulus that separates the “Low TFs” class from the “High TFs” class (see Methods). In the model, the σ parameter controls the slope of the psychometric function and measures the internal noise of the representation (i.e., rat sensitivity over the visual TF axis), while the γ parameter controls the gain of the internal perceptual representation.

This model allowed testing alternative hypotheses about the impact of sounds on visual representations, depending on the number of distinct values that the σ and γ parameters were free to take. At one extreme, allowing a different value of σ for each sound condition would represent the hypothesis that sound affects the sensitivity of the visual representation (i.e., the sharpness of the psychometric curves). At the other extreme, keeping a constant σ but allowing a different value of γ for each level of sound intensity would represent the hypothesis that sound compresses the visual representation according to its intensity, but leaves visual perceptual sensitivity unaltered.

Using the model to test for the impact of sound intensity on visual perception required estimating the sound intensity experienced by the rats in the time interval between stimulus onset and each animal's estimated reaction time (see Methods and Extended Data Fig. 2b). Analyzing the resulting, trial-averaged intensity levels revealed a finer-grain grouping of stimulus conditions, compared to the one shown in Fig. 1d. In addition to the conditions with maximal and minimal intensity (i.e., 74.8 dB for the fixed amplitude sounds and 32 dB for the purely visual stimuli, respectively), we found that the AM conditions could be grouped into three main clusters with increasingly larger intensity, based on the TF at which the noise bursts were modulated: TF = 0.25 Hz (48.5 $\pm$ 2.4 dB; mean $\pm$ s.e.m. across rats); TF = 0.82 Hz (64.2 $\pm$ 2.3 dB); and 1.32 Hz $\leq$ TF $\leq$ 4 Hz (70.7 $\pm$ 0.5 dB; these conditions are shown with progressively darker shades of gray in the stimulus matrix of Fig. 1a and in Fig. 2b-c; see Extended Data Fig. 2c for details).

We inferred the parameters of the ideal observer using a hierarchical Bayesian approach (see Methods and Extended Data Fig. 5), which yielded a posterior probability distribution over the values σ and γ that characterized each specific rat (referred to as *rat-level* parameters), as well as the values that characterized all rats taken together (referred to as *group-level* parameters). The group-level estimates obtained for a model with a single σ and 5 different γ per rat (i.e., one γ for each of the 5 levels of sound power in our stimulus set) are summarized by the psychometric curves shown in Fig. 2b.

The curves fitted tightly the measured classification accuracies and captured well the way the proportion of "high TF" choices progressively increased, while sound intensity dropped from maximal (green curve; fixed amplitude sound) to minimal (purple curve; no sound), passing through the three intermediate levels in which the AM sounds were clustered (gray curves). This means that an ideal observer with constant sensitivity (i.e., a single σ) could describe well the impact of sound on rat visual perceptual choices, provided that the internal representation of visual TFs was appropriately scaled by a gain factor (γ). Importantly, the magnitude of this gain decreased monotonically across the 5 levels of increasing sound intensity in which our stimulus conditions could be grouped (Fig. 2c).

To verify that this was the best description of the data, we compared the predictive power of this "reference" model to that of other variants, with a different number of values that σ and γ were allowed to take. The predictive power of each model was measured using the expected log predictive density (ELPD), that is, the average,

expected log-likelihood of the model on new data. This metric is estimated by leave-one-out cross-validation using a standard approach and automatically penalizes overly complex models that tend to overfit (see Methods). In our comparison, we computed the difference in ELPD between the reference model with 1 σ and 5 γ and each other model. We refer to this metric as ΔELPD.

The reference model with a fixed σ and 5 γ (one for each of the 5 levels of sound intensity) was the best (lowest ELPD) of a broad range of models we tested (Fig. 2b, inset). In particular, allowing a separate γ for each of the 9 AM conditions, thus obtaining a total of 11 possible γ values (while still keeping σ fixed to a single value), yielded a worse model (an indication of overfitting). This confirmed that noise bursts modulated within the [1.32 Hz, 4 Hz] range were all equivalent in terms of their impact on visual perceptual choices, as expected given their very similar intensity (Extended Data Fig. 2c). A model with 5 σ and 1 γ also performed much worse than the reference model, thus confirming that the different sound stimuli did not impact rat sensitivity to the visual TFs, but, rather, the dynamic range of the internal representation of the TFs. This conclusion was further reinforced by noticing that not even a model with 5 σ and 5 γ outperformed the reference model with only 1 σ and 5 γ.

In summary, our Bayesian ideal observer analysis supports the hypothesis that the effect of sound on the representation of temporally modulated visual gratings is to compress the visual perceptual space rather than increase or decrease perceptual uncertainty. Mechanistically, this finding suggests that sound acts as a powerful modulator of the responses of visual cortical neurons (see Fig. 2a) but leaves their tuning for temporal frequency unaltered.

This conclusion implies that, among the variety of auditory influences that have been reported in V1[8–13], inhibition may play a dominant role at the functional level[9]. At the same time, our results appear inconsistent with the sound-induced enhancement of cortical responses and the sharpening of orientation and direction tuning reported by some authors[8,12]. One possibility to reconcile these findings is that different kinds of perceptual tasks may be differently affected by auditory signals. In tasks requiring the discrimination of visual features that are related to the energy of the stimulus (such as its TF) and that are likely encoded by locally linear tuning curves (as postulated in our model; see Fig. 2a and Methods), sound-mediated engagement of inhibitory circuits may lead to a dominant, strong suppression of visually-evoked responses. In tasks where spatiotemporal features (e.g., orientation and direction) are encoded by neuronal populations with bell-shaped, unimodal tuning curves (e.g., over the orientation axis), the interplay between local inhibitory and disinhibitory circuits activated by sound could perhaps lead to a decrease of perceptual uncertainty and an improvement of decoding accuracy.

In conclusion, our study establishes that task-irrelevant auditory signals have a strong influence on the way task-relevant visual information is processed in the rodent brain.

Future work on behavior and visual cortical codes will be necessary to understand the nature of this modulation and its interplay with task requirements.

Acknowledgments

This work was funded by the European Union – NextGenerationEU – PNRRM4C2-I.1.1, in the framework of the PRIN Project no. 2022WX3FM5, CUP:G53D23003220006 (D.Z.) and PRIN Project no. 2022XE8X9E, CUP:G53D23004590001 (EP). The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union, nor can the European Union be held responsible for them. The study was also sup-ported by the Italian Ministry of University and Research under the call PRO3, project NEMESI (D.Z.). We thank Giuliano Iurilli, Maximiliano Jose Nigro, Matteo Marsili and Carlo Fantoni for valuable discussions and insightful feedback on the manuscript.

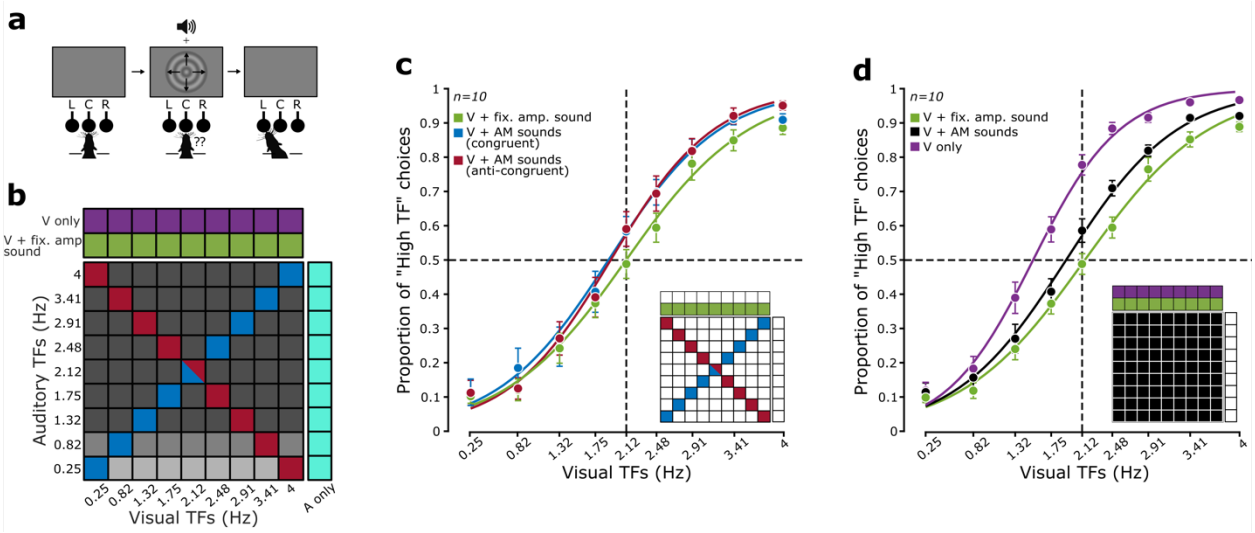

Figure 1 | Impact of sound on rat classification of visual temporal frequencies. a, Schematic representation of a behavioral trial. Rats were trained in sound-isolated operant boxes equipped with a display and a speaker for presentation of visual and auditory stimuli (see also Extended Data Fig. 1). As in previous studies[15,16], the rats learned to protrude their head through an opening in the box, so as to face the display and the speaker and trigger stimulus presentation by touching the central nose-poke port in a three-port array. The visual stimuli were circular, outward-moving (see arrows) sinusoidal gratings that rats learned to classify according to their TF. Gratings with TF < 2.12 HZ had to be classified as “Low TF”, while those with TF > 2.12Hz as “High TF”. Each class was associated with one of the two lateral nose-poke ports, and correct responses were rewarded with a drop of stevia-water solution. Auditory stimuli were task irrelevant. **b**, Stimulus conditions presented to the rats during the test phase of the experiment. These included: 1) visual gratings paired with fixed amplitude sounds (green cells); 2) visual gratings paired with AM sounds (all possible combinations of TFs in the two modalities; cells in the squared matrix; blue, gray and red cells refer, respectively, to congruent, incongruent and anti-congruent audiovisual conditions; the shades of gray, from light to dark, indicate AM sounds with progressively larger intensity; see main text for details); 3) unimodal purely visual gratings (purple cells); and 4) unimodal purely auditory AM sounds (cyan cells). **c**, Group average proportion of “high TF” choices (*n*=10 rats) as a function of the TF of the visual gratings, when the latter were paired with: 1) fixed amplitude sounds (green); 2) congruent AM sounds (blue); and 3) anti-congruent AM sounds (red). The error bars denote s.e.m. over the group of 10 rats. The curves are the result of logistic regressions to the data points. The inset illustrates which of the stimulus conditions shown in b contributed to the data points/curves. Extended Data Table 1 reports the results of a 2-way ANOVA having as factors the experimental condition ("V + fixed amplitude sound”, "V + AM congruent" and "V + AM anti-congruent") and the grating TF, as well as the results of a post-hoc Tukey test between each pair of experimental conditions. **d**, Same as in c, but for stimulus conditions where the gratings were paired with: 1) fixed amplitude sounds (green; same data as in c); 2) AM sounds (black); and 3) no sounds (purple). Extended Data Table 2 reports the results of a 2-way ANOVA having as factors the experimental condition and the grating TF, and the results of a post-hoc Tukey test between each pair of experimental conditions.

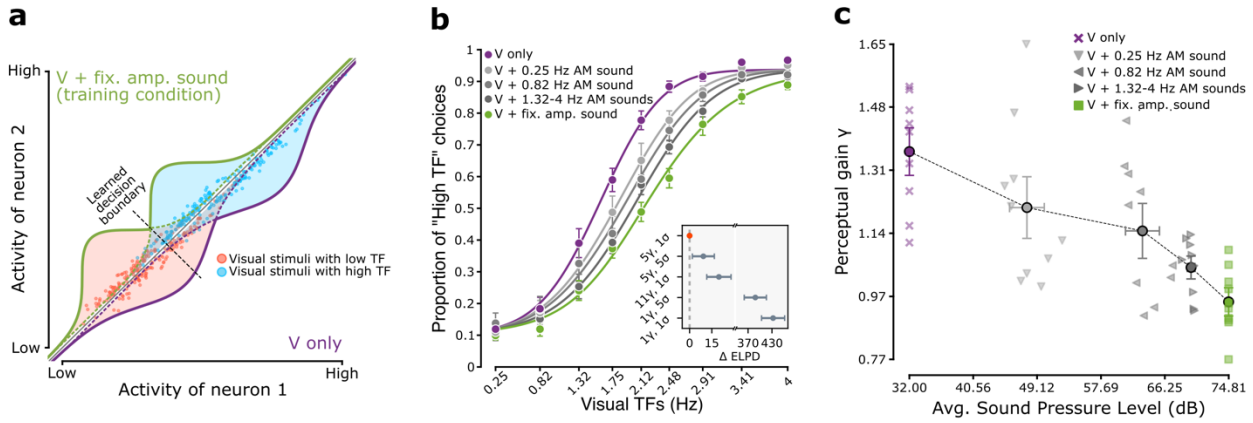

Figure 2 | An ideal observer model where neuronal representations encode linearly visual temporal frequencies and are suppressed by sound intensity accurately accounts for rat perceptual choices. a, Schematic of the assumptions at the base of the ideal observer model. The drawing illustrates a hypothetical neuronal representational space made of two units. Each neuron responds to visual gratings with an average firing rate that grows linearly as a function of the grating TF. Due to neuronal noise, the responses to repeated presentations of the same grating are distributed according to a Gaussian function centered on the average firing rate elicited by the grating. In the example, two TFs yields two distributions of population vectors (red and cyan dots). When the gratings are paired with a sound with strong intensity (e.g., the fixed amplitude noise burst), all responses are inhibited and, as a result, the distributions are compressed toward the origin of the representational space (green curves). Since rats are trained to discriminate the gratings under these conditions, they learn a decision boundary (dashed line) that optimally separates such compressed distributions. When the animals are presented with unimodal, purely visual gratings, the sound-induced inhibition is released and the response distributions shift towards higher firing rate values (purple curves). The animals, however, still rely on the previously learned decision boundary to discriminate these stimulus conditions. As a result, the overall proportion of "high TF" choices increases (i.e., more red dots fall on the "false alarm" side of the decision boundary). **b**, Group average proportion of "high TF" choices (n=10 rats) as a function of the TF of the visual gratings, when the latter were paired with: 1) fixed amplitude sounds (green); 2) AM sounds with increasingly larger TFs (progressively darker shades of gray, matching those in the stimulus matrix shown in Fig. 1b); and 3) no sounds (purple). The error bars denote s.e.m. over the group of 10 rats. The curves are group-level predictions of the proportions of "high TF" choices according to the Bayesian ideal observer model defined in eq. 1, with a single value of the sensitivity parameter σ and 5 different values of the scaling factor γ, one for each level of sound power (more precisely, each curve is the mean of the distribution of psychometric curves induced by the posterior distribution of the σ and γ parameters obtained for a given sound power level). The inset shows the difference between the expected log predictive density (ELPD) of the reference model with 1 σ and 5 γ values (i.e., the one yielding the psychometric curves) and variants of the model having a given number of free σ and γ parameters (as indicated on the ordinate axis). In all comparisons, the reference model was the one with lowest ELPD, i.e., with the higher predictive power. **c**, Relationship between the magnitude of the scaling factor γ in the ideal observer model and the intensity of the sounds that were paired with the visual gratings. Triangle, square and cross symbols refer to individual rats, while circles are group averages. Sound conditions are labeled according to the same color code as in b and Fig. 1b. The error bars denote s.e.m. over the group of 10 rats.

References

1. Nikbakht, N., Tafreshiha, A., Zoccolan, D. & Diamond, M. E. Supralinear and Supramodal Integration of Visual and Tactile Signals in Rats: Psychophysics and Neuronal Mechanisms. *Neuron* **97**, 626-639.e8 (2018).
2. Raposo, D., Sheppard, J. P., Schrater, P. R. & Churchland, A. K. Multisensory Decision-Making in Rats and Humans. *J. Neurosci.* **32**, 3726–3735 (2012).
3. Raposo, D., Kaufman, M. T. & Churchland, A. K. A category-free neural population supports evolving demands during decision-making. *Nat. Neurosci.* **17**, 1784–1792 (2014).
4. Fetsch, C. R., Turner, A. H., DeAngelis, G. C. & Angelaki, D. E. Dynamic Reweighting of Visual and Vestibular Cues during Self-Motion Perception. *J. Neurosci.* **29**, 15601–15612 (2009).
5. Seilheimer, R. L., Rosenberg, A. & Angelaki, D. E. Models and processes of multisensory cue combination. *Curr. Opin. Neurobiol.* **25**, 38–46 (2014).
6. Ernst, M. O. & Banks, M. S. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **415**, 429–433 (2002).
7. Ernst, M. O. & Bülthoff, H. H. Merging the senses into a robust percept. *Trends Cogn. Sci.* **8**, 162–169 (2004).
8. Ibrahim, L. A. *et al.* Cross-Modality Sharpening of Visual Cortical Processing through Layer-1-Mediated Inhibition and Disinhibition. *Neuron* **89**, 1031–1045 (2016).
9. Iurilli, G. *et al.* Sound-Driven Synaptic Inhibition in Primary Visual Cortex. *Neuron* **73**, 814–828 (2012).

10. Deneux, T. *et al.* Context-dependent signaling of coincident auditory and visual events in primary visual cortex. *eLife* **8**, e44006 (2019).

11. Oude Lohuis, M. N., Marchesi, P., Olcese, U. & Pennartz, C. M. A. Triple dissociation of visual, auditory and motor processing in mouse primary visual cortex. *Nat. Neurosci.* 1–14 (2024) doi:10.1038/s41593-023-01564-5.

12. Williams, A. M., Angeloni, C. F. & Geffen, M. N. Sound Improves Neuronal Encoding of Visual Stimuli in Mouse Primary Visual Cortex. *J. Neurosci.* **43**, 2885–2906 (2023).

13. Meijer, G. T., Montijn, J. S., Pennartz, C. M. A. & Lansink, C. S. Audiovisual Modulation in Mouse Primary Visual Cortex Depends on Cross-Modal Stimulus Configuration and Congruency. *J. Neurosci.* **37**, 8783–8796 (2017).

14. Bimbard, C. *et al.* Behavioral origin of sound-evoked activity in mouse visual cortex. *Nat. Neurosci.* **26**, 251–258 (2023).

15. Caramellino, R. *et al.* Rat sensitivity to multipoint statistics is predicted by efficient coding of natural scenes. *eLife* **10**, e72081 (2021).

16. Matteucci, G., Zattera, B., Marotti, R. B. & Zoccolan, D. Rats spontaneously perceive global motion direction of drifting plaids. *PLOS Comput. Biol.* **17**, e1009415 (2021).