Neural Correlates of Multisensory Detection Behavior: Comparison of Primary and Higher-Order Visual Cortex

Meijer, G.T.; Marchesi, P.; Mejias, J.F.; Montijn, J.S.; Lansink, C.S.; Pennartz, C.M.A.

DOI
10.1016/j.celrep.2020.107636

Publication date
2020

Document Version
Final published version

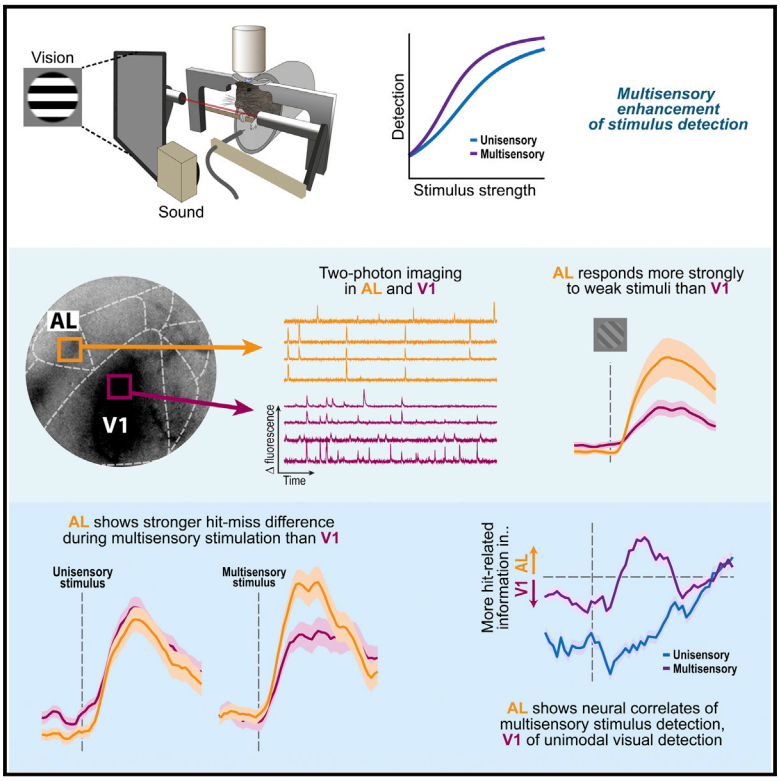
Published in
Cell Reports

Cell Reports

Article

Neural Correlates of Multisensory Detection Behavior: Comparison of Primary and Higher-Order Visual Cortex

Graphical Abstract



Highlights

- Mice show a multisensory enhancement of stimulus detection behavior
- Area AL neurons respond more strongly to weak visual and multisensory stimuli than V1
- Responses of both V1 and AL neurons show a contrast-dependent cross-modal modulation
- AL shows a stronger neural correlate of multisensory detection behavior than V1

Authors

Guido T. Meijer, Pietro Marchesi, Jorge F. Mejias, Jorrit S. Montijn, Carien S. Lansink, Cyriel M.A. Pennartz

Correspondence

c.s.lansink@uva.nl (C.S.L.), c.m.a.pennartz@uva.nl (C.M.A.P.)

In Brief

Meijer et al. show that mice combine vision and audition to improve their ability to detect faint stimuli in their surroundings. They reveal that a brain region, which lies in between the visual and auditory cortex (the anterolateral [AL] area), shows a neural correlate of this multisensory detection behavior.

Meijer et al., 2020, Cell Reports *31*, 107636
May 12, 2020
https://doi.org/10.1016/j.celrep.2020.107636

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Neural Correlates of Multisensory Detection Behavior: Comparison of Primary and Higher-Order Visual Cortex

Guido T. Meijer,[1] Pietro Marchesi,[1] Jorge F. Mejias,[1] Jorrit S. Montijn,[1] Carien S. Lansink,[1,2,3,*] and Cyriel M.A. Pennartz[1,2,3,4,*]

[1]Swammerdam Institute for Life Sciences, Center for Neuroscience, Faculty of Science, University of Amsterdam, 1098 XH Amsterdam, the Netherlands

[2]Research Priority Program Brain and Cognition, University of Amsterdam, 1098 XH Amsterdam, the Netherlands

[3]These authors contributed equally

[4]Lead Contact

*Correspondence: c.s.lansink@uva.nl (C.S.L.), c.m.a.pennartz@uva.nl (C.M.A.P.)

https://doi.org/10.1016/j.celrep.2020.107636

SUMMARY

We act upon stimuli in our surrounding environment by gathering the multisensory information they convey and by integrating this information to decide on a behavioral action. We hypothesized that the anterolateral secondary visual cortex (area AL) of the mouse brain may serve as a hub for sensorimotor transformation of audiovisual information. We imaged neuronal activity in primary visual cortex (V1) and AL of the mouse during a detection task using visual, auditory, and audiovisual stimuli. We found that AL neurons were more sensitive to weak uni- and multisensory stimuli compared to V1. Depending on contrast, different subsets of AL and V1 neurons showed cross-modal modulation of visual responses. During audiovisual stimulation, AL neurons showed stronger differentiation of behaviorally reported versus unreported stimuli compared to V1, whereas V1 showed this distinction during unisensory visual stimulation. Thus, neural population activity in area AL correlates more closely with multisensory detection behavior than V1.

INTRODUCTION

Our experience of the external world is generally defined by objects and features that are not limited to one sensory modality; it is indeed almost exclusively the result of multisensory processing (Bizley et al., 2012; Meijer et al., 2019; Pennartz, 2015). Combining information from multiple senses can increase the accuracy of our perceptual judgment when cues are composed of features originating from multiple modalities compared to a situation where solely unisensory information is available (Ernst and Banks, 2002; Gu et al., 2008; Nikbakht et al., 2018; Raposo et al., 2012). Behavioral studies in cats (Gingras et al., 2009) and mice (Meijer et al., 2018) showed that, indeed, the increased detection accuracy of audio-visual compared to unimodal stimuli is most often accounted for by the integration, rather than redundancy, of sensory cues.

Sensory cortices not only process information from their primary modality but also from other ones (Driver and Noesselt, 2008; Ghazanfar and Schroeder, 2006). For example, in the mouse primary visual cortex, neuronal activity is modulated by auditory input (Ibrahim et al., 2016; Meijer et al., 2017; Olcese et al., 2013) and neuronal activity in the auditory cortex is modulated by visual input, as shown in ferrets and macaques (Atilgan et al., 2018; Kayser et al., 2008, 2010). However, the areas in the cortical hierarchy and the neuronal mechanisms mediating multisensory integration, culminating in behavioral decisions, are still largely unknown. Because V1 does not project directly to premotor cortex (Wang et al., 2012), one or more additional cortical areas are expected to be involved in instantiating a behavioral response upon stimulus detection. A key candidate that may serve as hub for sensorimotor transformation of audio-visual information is the anterolateral (AL) area (Wang and Burkhalter, 2007), sometimes referred to as lateral secondary visual area (V2L; Banks et al., 2011; Hirokawa et al., 2008). Area AL receives input from both V1 and A1 (Laramée et al., 2011) and does project to premotor cortex (Wang et al., 2011). In addition, multisensory convergence was shown to occur most prominently at the interface of primary sensory areas (Hirokawa et al., 2008; Nikbakht et al., 2018; Olcese et al., 2013; Raposo et al., 2014; Wallace et al., 2004) and area AL is anatomically located in between V1 and A1 (Wang and Burkhalter, 2007). Thus, it is timely to compare how V1 and AL activity correlates with behavior in the context of multisensory integration.

In our previous work we showed that mice integrate sensory information from visual and auditory stimuli to improve their behavioral detection performance (Meijer et al., 2018). To investigate neural mechanisms that may underlie this improvement, we now used two-photon calcium imaging to record neuronal activity of ensembles of single neurons in V1 and AL of the mouse during the same audio-visual stimulus detection task. With this

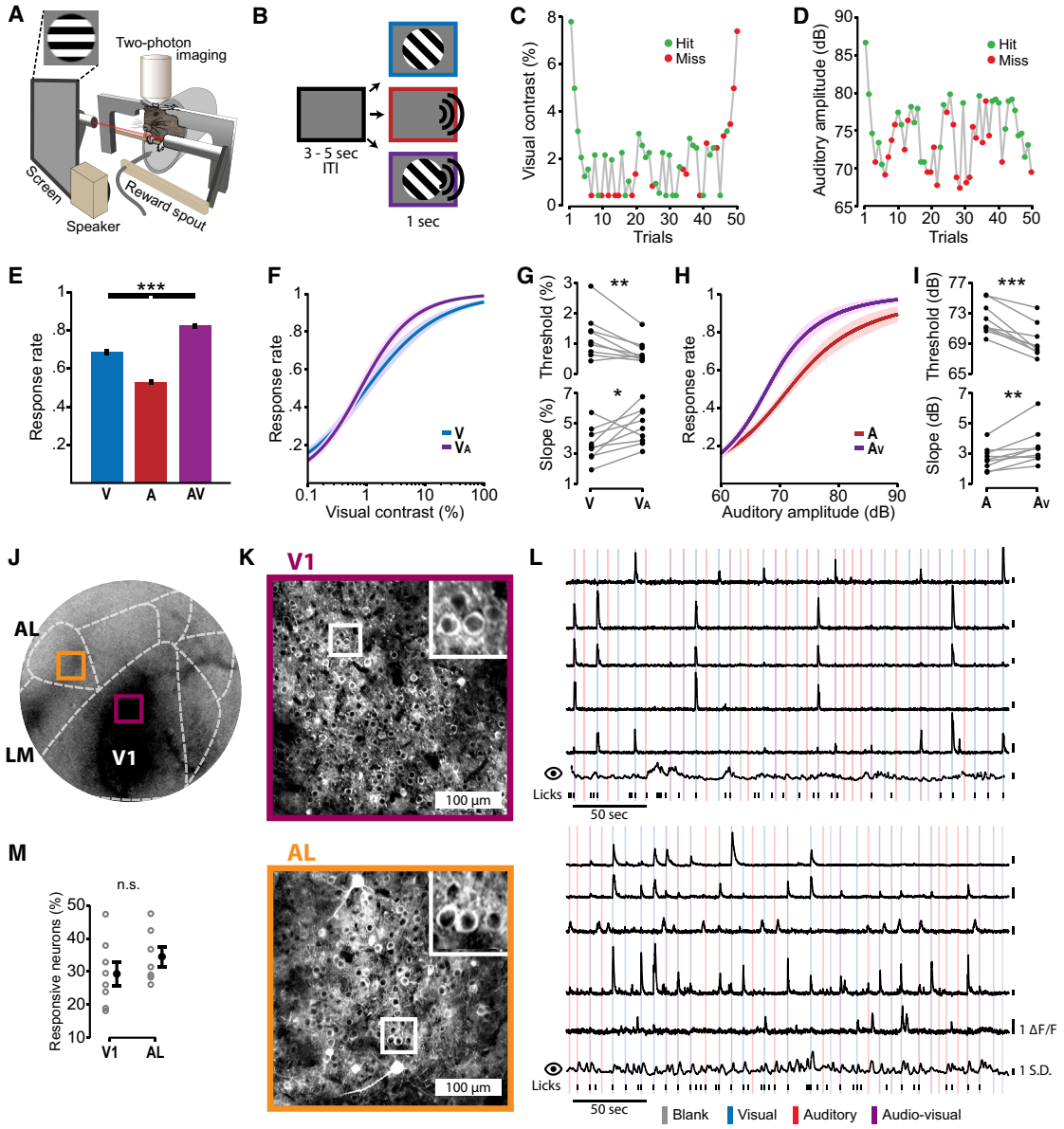


Figure 1. Two-Photon Calcium Imaging of V1 and AL during Audiovisual Stimulus Detection

(A) Two-photon and behavioral setup for head-fixed mice.

(B) Mice reported detection of visual (V), auditory (A) and audiovisual (AV) stimuli by performing a licking response during a 1-s stimulus presentation. Stimulus presentations were interleaved with a 3–5-s random inter-trial interval (ITI).

(C) Visual contrast was calibrated around the perceptual threshold of individual mice using an adaptive staircase procedure. Green circles represent trials in which the animal made a correct response (hit trials); the red circles indicate trials in which the animal did not detect the stimulus (miss trials).

(D) Same as (C), but for auditory amplitude.

(E) Mice showed enhanced detection performance in the multi- versus unisensory conditions as indicated by significantly higher response rates for all audio-visual-staircase trials compared to all visual- and auditory-staircase trials (one-way ANOVA).

(F) Psychometric function for visual-only (blue) trials and visual supported by subthreshold audio (V_A; purple) trials. Shading indicates SEM over mice.

(G) Audiovisual enhancement of detection performance is indicated by a significantly lower detection threshold and a steeper slope of the psychometric function for V_A compared with visual stimuli (paired t test over mice).

(H) Psychometric function for auditory-only (red) and auditory combined with sub-threshold visual trials (A_V; purple).

(I) The mean perceptual threshold and mean slope of the psychometric functions were significantly lower and steeper, respectively, in the A_V compared with the audiovisual condition (paired t test over mice).

(J) Imaging locations of example recordings overlaid with the intrinsic optical signal imaging map.

(K) Example two-photon imaging planes from V1 and AL, inset depicts zoom-in of the white square in the plane. Scale bar in the lower right corner indicates 100 µm.

(legend continued on next page)

experimental setup we specifically investigated three possible neural correlates of multisensory detection behavior: (1) neuronal populations in V1 and AL may be more sensitive to weak multisensory versus unisensory stimuli, (2) neurons in V1 and AL may show strong multisensory modulation in response to behaviorally relevant stimuli, or (3) V1/AL neurons may show a stronger correlate of behaviorally reporting multisensory stimuli compared to unisensory stimuli by responding differentially to reported versus non-reported stimuli.

RESULTS

We investigated the neuronal mechanisms in V1 and AL associated with the detection of visual and audiovisual stimuli by recording the activity of neuronal populations using two-photon calcium imaging (Figures 1A and 1B). Mice were first trained head-fixed in an audiovisual stimulus detection paradigm, which we previously used to show that the enhanced detection of multisensory compared to unisensory stimuli is dependent on cue integration (Meijer et al., 2018). Mice were presented with visual (V), auditory (A), or audiovisual (AV) stimuli of which the intensities were either near the threshold for detection or at asymptotic detection performance (for example, visual: 100% contrast/auditory: 90 dB). They reported the detection of a stimulus ("hit") with a lick response, triggering subsequent reward delivery. If a stimulus presentation was not followed by a lick response ("miss"), no reward was dispensed. The intensity of the visual and auditory stimuli was calibrated to the perceptual thresholds of each mouse by using two parallel adaptive staircase procedures (Figures 1C and 1D). Audiovisual trials were presented intermingled with the visual and auditory trials and consisted of the latest presented visual and auditory intensity from the respective staircases. Mice (n = 9) performed many trials per day across multiple recording days, adding up to a large number of trials per mouse, allowing robust psychophysics and statistics (mean ± SEM per mouse: 260 ± 11 trials per day, 1902 ± 153 trials in total). In general, mice responded specifically to the stimuli, because the rate of licking in blank trials, in which no stimulus was shown, was relatively low (false-alarm rate: 21.0% ± 2.4%, mean ± SEM., n = 9; Figure S1A). Mice were highly motivated to perform the task, as revealed by the low rates of omitted responses to uni- and multisensory stimuli (lapse rates: visual: 4.0% ± 0.8%, auditory: 8.0% ± 0.3%, audiovisual: 2.1% ± 0.7%; Figure S1A).

Mice showed a robust multisensory detection benefit because they responded significantly more often to audiovisual compared to visual and auditory stimuli across stimulus contrasts (16.8% ± 1.9% increase of response rate between visual and audiovisual stimuli; one-way ANOVA with post hoc Tukey-Kramer, $p < 10^{-5}$, n = 9; Figure 1E). The perceptual performance of the mice was determined by fitting a psychometric function for both visual and auditory trials using logistic regression. The detection thresholds, indicating the sensitivity of the perceptual system, were defined as the midpoints of the behavioral range between the minimum- and the maximum-attained values on the y axis of the psychometric curves. The mean visual perceptual threshold across mice was 1.3% ± 0.3% contrast and the mean auditory threshold was 72.9 dB ± 0.7 dB. The presentation of a subthreshold auditory stimulus together with the visual stimulus (V_A) facilitated visual stimulus detection performance, which was demonstrated by a significantly lower visual detection threshold and a steepening of the slope of the psychometric curve compared with the visual-only condition (Threshold: paired t test, p = 0.009; Slope: paired t test, p = 0.035, n = 9; Figures 1F and 1G). In the same vein, the cross-modal facilitation of auditory detection performance by a subthreshold visual stimulus (A_V) resulted in a lowering of the detection threshold and a steepening of the slope of the psychometric curve (threshold: paired t test, p < 0.001; slope: paired t test, p = 0.008, n = 9; Figures 1H and 1I). Thus, our results indicate a robust behavioral gain in the detection performance of multisensory, compared with unisensory, stimuli. This gain resulted from two factors: a shift of the psychometric curves toward lower contrasts, and a steepening of the psychometric function.

We recorded the activity of a large number of single neurons in V1 or AL over multiple days with two-photon Ca^{2+} imaging while mice performed the behavioral task (V1: 1,149 ± 95 trials, n = 8 mice, total of 32 recording sessions; AL: 1,175 ± 54 trials, n = 7 mice, total of 28 recording sessions). Six mice contributed to both V1 and AL recordings. Two-photon imaging was targeted to either V1 or AL using intrinsic optical imaging (Figures 1J–1L and S1C–S1F). In V1, 29.3% ± 3.5% (mean ± SEM over mice) of neurons were responsive to at least one stimulus modality, meaning that the fluorescence response of these neurons in high intensity uni- or multisensory trials exceeded the average fluorescence during blank trials by at least one SD. In area AL, a similar percentage of neurons was stimulus responsive (34.5% ± 3.1%; two sample t test of V1 versus AL, p = 0.29; Figure 1M). All subsequent analyses were performed on these responsive neurons unless otherwise specified.

The visual stimulus was a moving grating oriented in one of three possible directions (90°, 210°, or 330°). The visual response of neurons included in the analysis was always the response to the preferred direction. The preferred direction of neurons was stable over days (Figures S1G and S1H). The proportions of responsive neurons in V1 and AL were similar for each stimulus condition (visual, audiovisual, and auditory). The two areas did not differ significantly in basic response properties to visual stimuli such as response reliability and population sparseness (data not shown). Notably, a small percentage of neurons was responsive to auditory stimulation without any visual stimulation (V1: 2.7% ± 0.9%, AL: 2.5% ± 0.8%); approximately half of these neurons also responded to a visual stimulus (Figures S2A–S2G).

(L) Fluorescence activity in ΔF/F of four example neurons from the recording sites in (K), Z-scored pupil diameter and licking timestamps. Height of the line on the right-hand side indicates value of 1 ΔF/F for fluorescence traces and 1 SD for pupil size. Colored vertical bars indicate stimulus presentations of blank (gray), visual-only (blue), auditory-only (red) and audiovisual (purple) stimuli.
(M) Percentage of responsive neurons to visual, auditory, and/or audiovisual stimuli per mouse in V1 and AL.
*p < 0.05, **p < 0.01, ***p < 0.001. Error bars and shading indicates SEM. See also Figure S1.

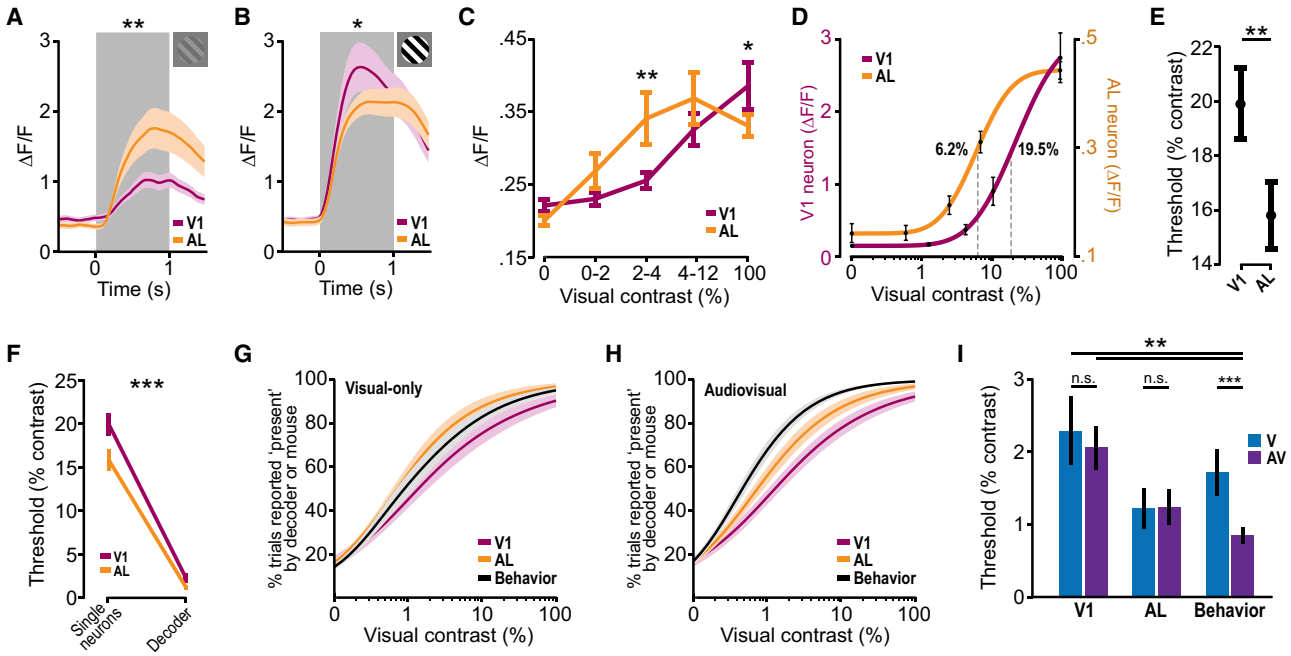


Figure 2. AL Is More Sensitive to Weak Visual Stimuli Compared to V1

(A) During low-contrast (2%–4%) visual stimulation, AL neurons responded significantly stronger compared to V1 neurons (significance tested for maximum in 0–500-ms time window, Wilcoxon rank-sum test). Gray box indicates time of stimulus presentation; shading indicates SEM over neurons.

(B) During 100% contrast visual-only stimuli, the fluorescence responses of V1 neurons were significantly stronger compared with AL neurons. Wilcoxon rank-sum test as in (A).

(C) AL population response is stronger than V1 for visual stimuli of low contrast, but weaker than V1 for high contrast.

(D) Neurometric curve of example V1- and AL neurons fitted to mean fluorescence response for increasing visual contrast. Dotted line indicates the neurometric thresholds defined as the midpoint between the upper and lower boundary of the functions.

(E) Neurons in AL had, on average, lower neurometric visual thresholds compared with V1 neurons (Wilcoxon rank-sum test).

(F) Visual thresholds of decoders trained on V1- and AL-population data are much lower than thresholds from single neurons in these areas (Wilcoxon rank-sum test).

(G) Neurometric curves fitted to the performance of a Bayesian classifier that was trained to predict whether a visual-only stimulus was present based on population activity from V1 or AL. In black, the behavioral psychometric function during visual-only stimulation is plotted.

(H) Same as (G) for audiovisual trials.

(I) Thresholds derived from neurometric and psychometric curves in (G) and (H) show that V1 thresholds are significantly higher compared to audiovisual behavior whereas AL decoder thresholds are statistically similar to audiovisual behavior (Kruskal-Wallis Test). Mice show a multisensory improvement of detection behavior whereas both V1 and AL decoders do not (paired Wilcoxon signed-rank test). Thus, neither V1 nor AL neurometric functions could explain the multisensory enhancement in behavior.

*p < 0.05, **p < 0.01, ***p < 0.001. Error bars and shading indicates SEM. See also Figure S2.

AL Is More Sensitive to Weak Stimuli Compared to V1, but This Difference Does Not Correspond to Behavioral Multisensory Enhancement

We first asked whether the multisensory improvement in stimulus detection could be due to an increase in neuronal sensitivity in AL. Neurons in AL have larger receptive field sizes compared to V1 (Wang and Burkhalter, 2007), suggesting that single neurons in AL receive input from multiple V1 cells. Such a convergent processing scheme may increase the responsivity of AL neurons to lower-contrast stimuli, which may apply to both visual and multisensory stimuli. We first assessed the responsivity of neurons in V1 and AL to low- and high-contrast visual-only stimuli comparing the mean fluorescence responses between areas (pooled over mice, V1: n = 264 neurons; AL: n = 253 neurons). Neurons in area AL showed a significantly stronger mean response to low-contrast (2%–4% contrast) visual stimuli compared to neurons in V1 (Wilcoxon rank-sum test over maximum fluorescence in 0-500-ms window, p = 0.007; Figure 2A). Contrary to low contrasts, high-contrast stimulation (100%) elicited significantly stronger V1 responses to visual stimuli than AL responses (Wilcoxon rank-sum test, p = 0.01; Figure 2B). This was confirmed when binning trials into contrast bins and plotting the V1 and AL population response (Figure 2C).

The stronger response to low-contrast visual stimuli in AL indicated that neurons in this area are more sensitive to low-intensity stimuli compared to V1 neurons. We tested this by fitting a neurometric function to each neuron's fluorescence response to visual stimuli of progressively increasing contrasts. The neurometric threshold for visual stimuli was determined as the midpoint between the upper and lower bound of the neurometric function (Parker and Newsome, 1998; Stüttgen et al., 2011; Figure 2D). Neurons in AL showed significantly lower neurometric thresholds compared to V1 neurons (Wilcoxon rank-sum test, p = 0.001; Figure 2E). This higher sensitivity could not be explained by

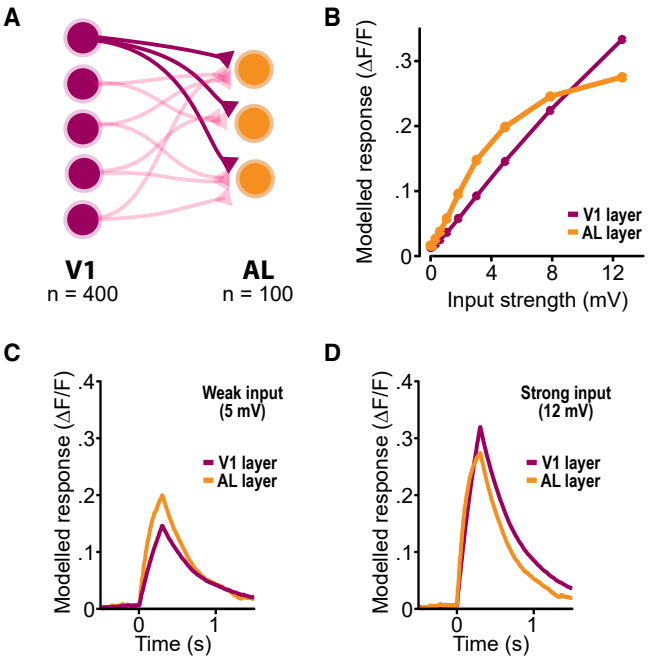


Figure 3. Computational Model of V1-AL Interaction

(A) V1 and AL were modeled as networks of leaky integrate-and-fire neurons. Feedforward input from V1 (n = 400 neurons) converges onto AL (n = 100 neurons). Neurons were modeled with an activity-dependent short-term synaptic depression mechanism.

(B) The model was provided with input of increasing strength and its output qualitatively matched the observed neural data in Figure 2C.

(C) Similar to the experimental data from Figure 2A, providing weak input (5 mV) to the model resulted in a stronger modeled fluorescence response in the AL layer compared to V1.

(D) Modeled fluorescence response of V1 and AL. When providing a strong input (12 mV) to V1, this area shows a stronger response than AL.

better behavioral performance during AL compared to V1 recording sessions (visual-only hit rates; V1: 0.69 ± 0.04, AL: 0.68 ± 0.02; two-sample t test, p = 0.90). Furthermore, the order of imaging V1 and AL was counterbalanced across days, precluding the possibility that the increased sensitivity to visual stimuli is due to within-session experience.

While the presence of larger receptive fields in AL than in V1 provides a plausible explanation for the larger responses in AL to low-contrast stimuli, the weaker response at high contrast is not explained by this argument. One possibility is that a biophysical regulatory mechanism operating between V1 and AL contributes to response saturation, which would reduce the response strength for high-contrast stimuli in AL relative to V1. We built a computational model (Figure 3) to test the hypothesis that short-term synaptic depression could account for the stronger V1 versus AL response during high-contrast stimulation, and the weaker V1 versus AL response during low-contrast stimulation. The model involves a population of 400 V1 neurons whose synaptic outputs converge onto a population of 100 AL neurons (see Methods S1). Providing the model with input of increasing strength resulted in a modeled fluorescence response in V1 and AL that resembled the experimentally observed neural activity (Figures 3B and 2C). When V1 neurons receive weak input, simulating low-contrast visual stimuli, activity in the network is low and the larger AL-receptive fields, relying on the synaptic convergence of V1 outputs onto AL neurons, lead to a modeled fluorescence response that is higher in AL than V1 (Figure 3C). However, with stronger visual input the high firing rates cause a faster depletion of synaptic resources, effectively decreasing the strength of V1 projections to AL and reducing the AL response, as observed experimentally (Figure 3D). Therefore, a relatively simple mechanism such as short-term depression may account for the responses being lower in AL compared to V1 during high-intensity visual stimuli.

Does the high sensitivity of AL neurons to low-contrast stimuli correlate with the enhanced multisensory detection performance observed in behavior? The neural response of AL neurons during audiovisual stimulation was similar to their visual-only response; AL neurons responded more strongly to low-contrast audiovisual stimuli and more weakly to high-contrast audiovisual stimuli compared to V1 neurons (Figures S2H and S2I). This suggests that, although the sensitivity of AL is higher than V1, its multi- and unisensory sensitivities are similar. To test this further, we investigated the link between the behavioral psychometric performance of the mouse and the neurometric performance of V1 and AL population activity. The combined response of a population of neurons may greatly outperform single-neuron responses in reporting stimulus presence, because, rather than relying on a single value, it can integrate information across neurons. The population-level sensitivity to both visual and audiovisual stimuli was assessed using a Bayesian decoder that was trained to distinguish blank trials from trials in which a stimulus was presented. For every trial, the decoder then predicted whether a stimulus had been presented or not, based solely on the neural population activity. Only those trials in which the mouse made a behavioral response (hit trials) were included in this analysis, so that trials with and without a behavioral response were not mixed (but see Figures S2J–S2M for hit-and-miss decoding results). Subsequently, a neurometric function was fitted to the output of the trained decoder. The thresholds of the decoders were indeed much lower compared to the average threshold of single neurons (Wilcoxon rank-sum test, $p < 10^{-3}$; Figure 2F). Furthermore, decoding stimulus presence from AL population activity resulted in significantly lower detection thresholds compared to V1 populations (Wilcoxon rank-sum test, V1: n = 8 mice; AL: n = 7 mice, p = 0.037). The population neurometric functions obtained from decoding the presence of a visual-only stimulus from V1 and AL population activity partially overlapped with the behavioral visual detection performance of the mouse (Figure 2G). During audiovisual stimulation, the psychometric function was shifted toward lower-contrast stimuli, whereas both V1 and AL neurometric functions stayed largely the same (Figure 2H).

If the increased behavioral sensitivity to multisensory stimuli would be explained by an increase in neuronal sensitivity to multisensory stimuli, one would expect that the neurometric functions derived from audiovisual trials would be shifted toward lower contrasts compared to the visual-only neurometric functions. However, there was no significant shift in the neurometric threshold in the audiovisual compared to visual-only stimuli for both V1 (paired

Wilcoxon signed-rank test, p = 0.39) and AL (paired Wilcoxon signed-rank test, p = 0.75; Figure 2I). Mouse behavior did show a shift in psychometric threshold in the audiovisual compared to the visual condition (paired Wilcoxon signed-rank test, $p < 10^{-3}$; Figure 2I). An overall comparison showed that the neurometric thresholds from V1 for both visual and audiovisual stimuli were significantly higher than the audiovisual behavioral thresholds whereas the AL neurometric thresholds were not statistically different from audiovisual behavioral thresholds (Kruskal-Wallis with post hoc Tukey-Kramer, p = 0.0018). Taken together, these results show that AL neurons are more sensitive to low-contrast visual and audiovisual stimuli compared to V1 neurons, but this increased sensitivity does not correspond to the multisensory detection enhancement observed in behavior.

Different Subsets of Neurons Show Cross-Modal Modulation Depending on Stimulus Intensity

Another neural correlate of the multisensory enhancement in detection behavior may concern cross-modal response modulation. Although areas V1 and AL are considered to be visual areas, they are both innervated by efferents from auditory cortex, indicating that they may support multisensory processing (Deneux et al., 2019; Leinweber et al., 2017; Oh et al., 2014). We first examined the processing of multisensory stimuli in both areas and then addressed their relation with audiovisual detection behavior. Only trials in which the mouse responded to the stimulus were used in this analysis. During performance of the multisensory detection task, the responses of single neurons in V1 were either suppressed or enhanced when a tone was presented concurrently with a visual stimulus (Figure S3). Similarly, the response of AL neurons was either suppressed (Figure 4A) or enhanced (Figure 4B) when a tone was presented concurrently with a visual stimulus. There was no difference in pupil diameter in the time window used to gauge neuronal activity (0–500 ms after stimulus onset), indicating that any difference in neuronal activity between modalities was not likely due to differences in arousal (Figures S4A–S4G). In both V1 and AL, the neurometric functions for the responses to visual and audiovisual stimuli indicated a surprising feature of stimulus processing: neurons that showed auditory modulation of visual responses to low-contrast stimuli generally did not show modulation for high-contrast stimuli (example AL neuron in Figure 4C; example V1 neuron in Figure S3C). Conversely, neurons that showed auditory modulation at high-visual contrast did not do so during threshold-contrast stimulation (example AL neuron in Figure 4D; example V1 neuron in Figure S3D). This indicates that the modulation of high- and low-contrast visual and audiovisual stimuli may be mediated by different subsets of neurons in both low- and high-level areas of visual cortex.

We quantified this difference by computing a modulation index per neuron that reflected whether a neuron's visual response was enhanced (positive index value) or suppressed (negative index value) by the presence of a tone. We plotted the modulation index of each significantly modulated neuron for high and threshold visual contrasts and color-coded them according to stimulus contrast modulation (Figure 4E for AL; Figure S3E for V1). AL and V1 neurons were either modulated at high or low stimulus intensities but rarely at both. Indeed, the percentage of neurons modulated in both intensity ranges (1.6% ± 0.5%) was not significantly different from chance (paired t test versus joint probability, p = 0.90). These data indicate that neurons in V1 and AL showing cross-modal modulation at threshold visual contrast intensities and neurons that were modulated at high stimulus contrast formed largely non-overlapping populations. In addition, we found no difference between the modulation strength of V1 and AL neurons during both high- and low-contrast stimulation (Wilcoxon signed-rank test, high contrast: p = 0.43, low contrast: p = 0.29).

We further investigated this effect on the population level by calculating the mean absolute modulation index of the two populations for stimuli of increasing visual contrast. Both high- and low-contrast modulated subpopulations of neurons varied their cross-modal modulation with contrast (one-way ANOVA, low-contrast modulated: p = 0.032; high-contrast modulated: $p < 10^{-8}$; Figure 4F). Moreover, the analysis confirmed that the population modulated at high contrast did not show modulation when the visual contrast was low, and vice versa (t test high versus low, 2%–4% contrast bin: p = 0.033, 4%–12% contrast bin: p = 0.0023, 100% contrast bin: $p < 10^{-9}$; Figure 4F). A ceiling effect possibly occurring in cells that were modulated at low, but not high, contrast was deemed unlikely because weakly responsive neurons would then be predicted to show a higher prevalence of cross-modal modulation at high contrast as compared to strongly responsive cells, which was not found in V1 or AL (data not shown). Additionally, we investigated population-level coding of stimulation type by using a Bayesian decoder to classify the stimulus (visual or audiovisual) based on the population activity of either the threshold-modulated or the high-contrast-modulated neuronal population. Decoding was performed on either low-contrast (0%–12%) or high-contrast (100%) visual and audiovisual stimuli. For low-contrast trials, the stimulus type (visual or audiovisual) could only be decoded above chance level from the activity of AL populations composed of neurons that were modulated at low visual contrast (t test versus 50%, false discovery rate (FDR)-corrected p = 0.0028, n = 8 mice; Figure 4G; Figure S3G for V1). Likewise, decoding stimulus type during high-contrast stimuli only exceeded chance level when using the activity of neuronal populations modulated at high contrast (t test versus 50%, FDR-corrected p = 0.0009, n = 8 mice; Figure 4G). Together, these results indicate that the difference between visual and audiovisual stimuli is coded in V1 and AL, both at high- and low-visual-contrast intensities, albeit by different neuronal subpopulations.

Could the cross-modal modulation of neural responses constitute a neural correlate of the enhanced ability of mice to detect weak multisensory stimuli? The absolute modulation strength was significantly weaker for low- compared to high-contrast stimuli (Wilcoxon signed-rank test, $p < 10^{-18}$; Figure 4H), which does not correlate with the behavioral observation that multisensory enhancement is stronger for weak- versus high-contrast stimuli. Despite this mismatch, it cannot be ruled out that a ceiling effect in behavioral multisensory enhancement may play a role. Therefore, the present data do not permit us to conclude that the cross-modal modulation of neural responses is behaviorally meaningful.

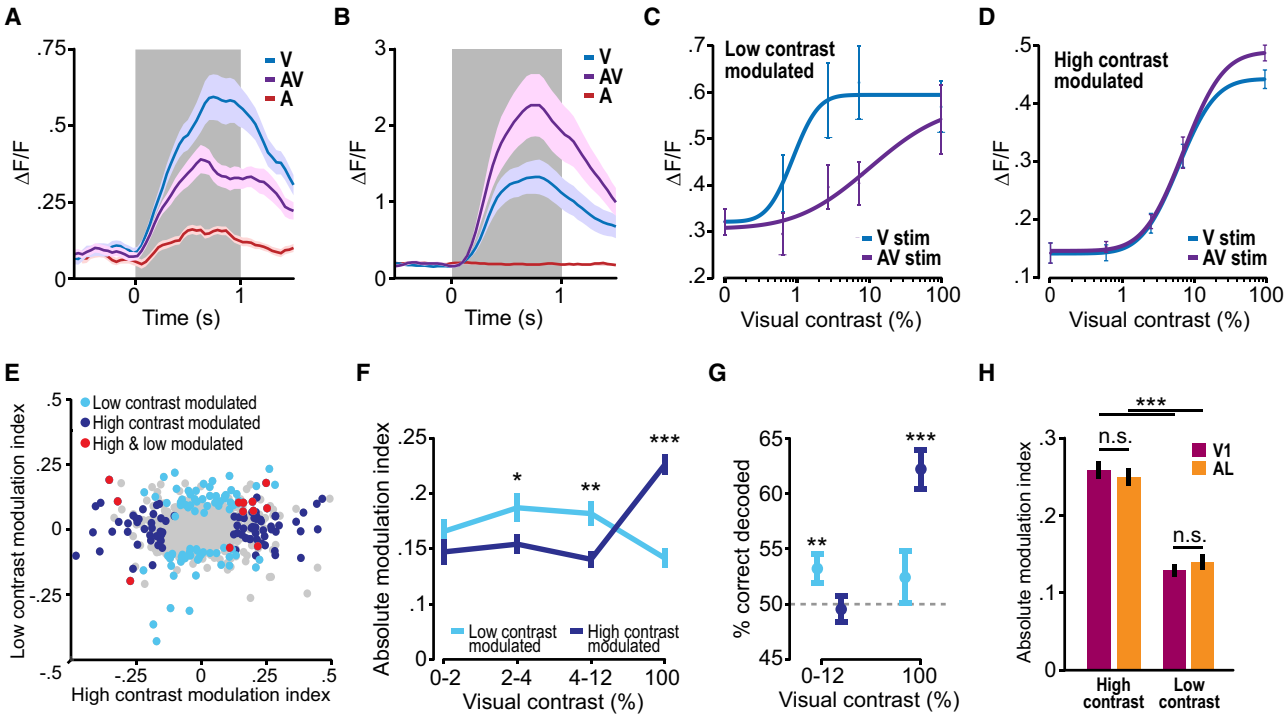


Figure 4. Different Subsets of AL Neurons Code Stimulus Modality Depending on Visual Contrast

(A–D) Different neurons are represented.

(A) Average fluorescence response of a single AL neuron to high-contrast visual-only (blue line), high-amplitude auditory-only (red line) and combined audiovisual (purple line) trials (synchronized to stimulus onset; t = 0 s). The visual response of this example AL neuron is suppressed by the concurrent presentation of a tone.

(B) The visual response of this example neuron is enhanced in the audiovisual compared to the visual-only condition.

(C) Neurometric functions of a single-example AL neuron for visual-only (blue line) and audiovisual (purple line) stimuli. This neuron shows a significant differential response to visual and audiovisual stimuli only during low-contrast visual stimulation ($p < 0.05$, significance determined by shuffling; modulation index for low contrasts [0%–12%] = −0.29; high contrasts [100%] = −0.05).

(D) Neurometric curve of an example AL neuron showing cross-modal modulation when the visual contrast was high but not during low-contrast stimulation (modulation index for low contrasts = −0.01; high contrasts = 0.17).

(E) The modulation index during high- and low-contrast visual stimulation for all AL neurons; every dot represents a neuron. Neurons were significantly modulated in their response at low-contrast intensity ("low contrast modulated," example in C) or during high-contrast stimulation ("high contrast modulated," example in D), but rarely at both high and low contrast ("high & low modulated"). Neurons that were not modulated for any contrast are shown in gray.

(F) Mean absolute modulation index of the low-contrast-modulated AL neuronal population (light blue) and the high-contrast-modulated population (dark blue) for bins of increasing visual contrast (two-sample t tests; *P*-values were FDR corrected).

(G) A Bayesian decoder classifying stimulus modality (visual or audiovisual) was trained on either the AL population activity of the low-contrast-modulated population or the high-contrast-modulated population. For low-visual contrast (0%–12% contrast), stimulus modality could only be decoded above chance (50%; dotted line) when training the decoder on the low-contrast-modulated population and not when trained on the high-contrast-modulated population. The opposite was true when modality was decoded from high-contrast trials (t test versus 50% chance level; p values FDR corrected for multiple comparisons).

(H) The strength of the modulation was not significantly different between V1 and AL for either high- or low-contrast stimuli.

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. Error bars and shading indicates SEM. See Figure S3 for V1 results.

AL Neuronal Activity Is Associated with the Behavioral Report of Audiovisual Stimulus Detection

Above, we showed that neurons in the visual cortical system differentiate between uni- and multisensory stimuli, but this cross-modal modulation does not correlate with the multisensory gain observed in behavior. Therefore, we investigated a third possible neural correlate of multisensory detection behavior: an increased association between neuronal responses and behavioral reportability, specifically in the multisensory condition. Whereas the previous analyses pertained to the sensory sensitivity of neuronal responses to a stimulus that was, in fact, present or not, the analyses in this section focus on how well neurons encode the behavioral report of the mouse (hit-versus-miss response to presented stimuli). A first indication of how strongly the neuronal populations differentiate between reported and unreported stimuli is provided by the difference between the mean Z-scored activity of all responsive neurons between hit-and-miss trials. The mean hit-miss difference for visual-only threshold intensity trials was positive for both V1 and AL, and of similar strength, indicating that neurons generally responded more strongly in hit-versus-miss trials (Montijn et al., 2015; two-sample t test, $p = 0.72$, V1: $n = 264$, AL: $n = 253$; Figure 5A). During multisensory stimulation, however, neurons in area AL showed a stronger mean hit-miss response difference than V1

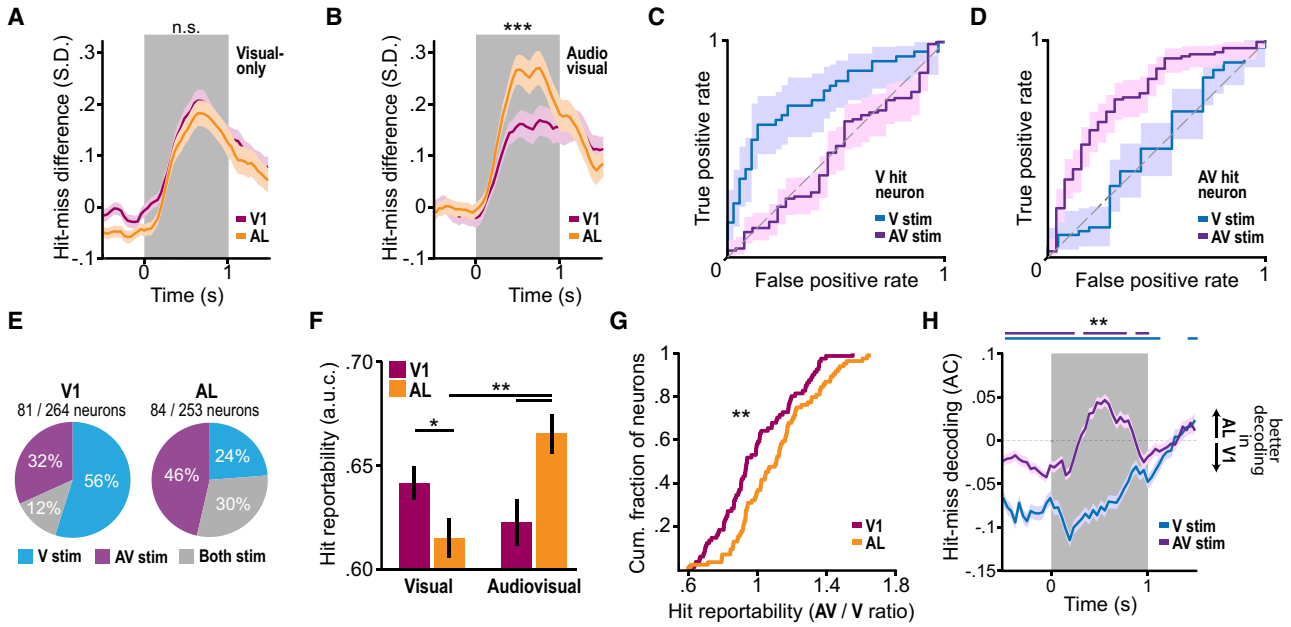


Figure 5. AL Shows a Neuronal Correlate of Audiovisual Hit-Miss Behavior

(A) For visual-only stimuli, V1 and AL show similar hit-miss differences in Z-scored fluorescence in degrees of freedom of fluorescence (df/F) values of all responsive neurons (significance tested between the maximum fluorescence response in a 0–500-ms window after stimulus onset using a two-sample t test). Gray box indicates stimulus presentation.

(B) Hit-miss difference for audiovisual trials shows stronger hit-related activity in AL compared with V1 (two-sample t test).

(C) ROC curve of hit-versus-miss differentiation of the activity of a single example V1 neuron during visual-only (blue) and audiovisual (purple) stimulation. This neuron differentiates hit from miss trials during visual but not during audiovisual trials. Shading indicates 95% confidence intervals as determined by bootstrapping.

(D) Example AL neuron as in (C), which shows a separation of hit-versus-miss trials in its neuronal responses only for audiovisual but not visual stimuli.

(E) Total number of neurons showing significant hit-miss differentiation (significance determined by bootstrapping of ROC curves), separated for visual and/or audiovisual stimuli in V1 and AL. Pie charts show the percentage of neurons that were significantly hit-miss modulated in visual trials (blue), audiovisual trials (purple), or both (gray). There is a significantly larger fraction of visual-hit neurons in V1 compared to AL- and audiovisual-hit neurons in AL compared to V1 ($p < 0.05$, binominal test of proportions).

(F) Hit-miss reportability, defined as the area under the ROC curve (AUC) that differentiates between hits and misses, for V1 (burgundy) and AL (orange) calculated for visual and audiovisual trials. V1 showed stronger hit-miss differentiation for visual-only compared with audiovisual stimuli. Area AL showed stronger hit-miss differentiation during audiovisual- compared to visual-only trials. During audiovisual stimulation, AL showed stronger hit-related activity compared with V1 (FDR-corrected t tests).

(G) Cumulative distributions of the ratio of audiovisual- and visual-hit probabilities of all significant hit neurons pooled over mice. Hit-miss reportability for audiovisual stimuli was per-ratio higher in area AL than in V1 (Kolmogorov-Smirnov test).

(H) Frame-by-frame hit-miss classification performance was quantified as the area coefficient (AC) using a mixed effects model. Positive AC values indicate better decoding of behavioral outcome in AL versus V1 and negative AC values indicate that hit-miss decoding was better in V1 versus AL. Before stimulus onset, both visual-only and audiovisual trials were significantly better decoded by V1. During audiovisual stimulation, decoding of behavioral outcome was better in AL whereas during visual-only stimulation, hit-miss decoding was better in V1. During reward consumption, after stimulus offset, decoding of response type (hit or miss) was better in AL compared to V1. Blue lines above plot indicate periods in which visual trials are decoded significantly better in either V1 or AL; purple lines for audiovisual trials (Z-test, Bonferroni corrected).

*$p < 0.05$, **$p < 0.01$, ***$p < 0.01$. Error bars and shading indicates SEM. See also Figures S4 and S5.

neurons (two-sample t test, $p < 10^{-3}$; Figure 5B). This indicates that neuronal activity in V1 and AL correlates with behavioral detection, and that AL is particularly associated with the detection of multisensory stimuli.

To determine whether neurons showed a stronger correlate of behavioral reportability specifically during uni- or multisensory stimulation, we computed the probability that an ideal observer can categorize behavioral choice (hit-versus-miss) based upon the trial-by-trial activity fluctuations of the neuron. We constructed receiver operating characteristic (ROC) curves per neuron for visual - and audiovisual trials separately. Hit-miss reportability was defined as the area under the ROC curve per neuron; a neuron was labeled "hit neuron" when its hit-miss reportability was > 0.5 and the bootstrapped 95% confidence interval did not include 0.5. In this case, the neuron significantly signaled detection-related information. We found a subset of neurons in both V1 and AL that coded hit-miss differences only when the stimulus was a unisensory visual stimulus, but not when the visual stimulus was accompanied by a tone (visual hit neuron; example in Figure 5C). Conversely, another subset of neurons coded multisensory, but not visual, detection performance (audiovisual hit neuron; example in Figure 5D). Figure 5E

indicates the proportions of hit neurons in V1 and AL for the different stimulus types. In V1, the majority of hit neurons was associated to visual stimuli, whereas, in AL, the largest proportion of hit neurons pertained to audiovisual stimuli. The group of audiovisual hit neurons in AL was significantly larger than in V1 (V1: 32%, AL: 46%; binominal test of proportions, $p = 0.011$, $n = 84$). The same held for the group of AL neurons coding both visual and audiovisual detection as compared to V1 (V1: 12%, AL: 30%; binominal test of proportions, $p = 0.0014$, $n = 84$). By consequence, AL's proportion of visual-only hit neurons was significantly smaller compared with V1 (V1: 56%, AL: 24%; binominal test of proportions, $p < 10^{-5}$, $n = 84$). Therefore, although the number of neurons that show a neural correlate of reportability was similar in V1 and AL, most neurons in V1 coded visual-only hits, while most AL neurons coded audiovisual hits.

Given this contrast between fractions of V1 and AL hit neurons, we next asked whether the magnitude of neural correlates of reportability is in line with these properties. For this analysis, we selected all neurons that showed significant hit-miss reportability for either visual or audiovisual stimuli. During visual-only trials, V1 neurons showed stronger hit-miss reportability compared to AL neurons (two-sample t test, FDR-corrected $p = 0.048$; Figure 5F). AL neurons, on the other hand, showed a stronger correlate of reportability compared to V1 during audiovisual stimulation (two-sample t test, FDR-corrected $p = 0.008$, V1: $n = 81$, AL: $n = 84$; Figure 5F). Moreover, AL neurons showed stronger hit-miss reportability for audiovisual compared with visual stimuli (paired t test, FDR-corrected $p = 0.002$, $n = 84$; Figure 5F). In short, we found that V1 neurons showed a strong neural correlate of visual-only reportability, whereas AL neurons strongly correlated with the reportability of multisensory stimuli. These V1-AL differences were mostly driven by neurons that showed a cross-modal enhancement, whereas cross-modally suppressed neurons did not show this pattern (Figures S5A–S5D).

We quantified on a neuron-by-neuron level if AL neurons signaled a consistently higher audiovisual versus visual reportability compared with V1 neurons by pooling all significant hit neurons over mice and plotting the cumulative histogram of the audiovisual/visual ratio. The ratio of audiovisual to visual was larger in area AL compared to V1, consistent with the stronger audiovisual hit-related modulation of AL neurons (Kolmogorov-Smirnov test, $p = 0.006$, V1: $n = 81$, AL: $n = 84$; Figure 5G). In conclusion, single neurons in area AL show strong detection-related modulation of activity during audiovisual stimulation.

We next investigated how detection-related modulation of population activity evolved over the time course of stimulus presentation. We trained a Bayesian decoder to classify hit from miss trials and fitted a linear mixed effects model on the performance of this decoder to investigate at which point in time population activity in V1 or AL is a better predictor of behavioral outcome. Whether V1 or AL was better at coding for behavioral outcome at every point in time was quantified as the area coefficient, which is the regression coefficient of the fitted linear regression. If the area coefficient was positive, the decoding performance derived from AL was higher compared to V1 and when it was negative, the decoding performance was higher in V1 versus AL. During visual trials, decoding performance of behavioral outcome was significantly better in V1 compared to AL, even before stimulus onset (Z-test with Bonferroni correction, $p < 0.01$; Figure 5H). The higher decoding performance before stimulus onset suggests that non-stimulus-induced fluctuations in V1 activity are informative as to whether incoming stimulus information will be relayed to motor decision areas (Montijn et al., 2015, but see Figures S5E–S5H). Such fluctuations may cohere, for instance, with the level of arousal or attention. In audiovisual trials, the pre-stimulus dominance of V1 over AL persisted, but during stimulus presentation this shifted to a significantly higher decoding of behavioral outcome in AL. This increased decoding performance in AL largely continued after stimulus offset, presumably because the mouse was consuming the reward during this time period and neural activity in AL may be more strongly linked to motor output compared to V1 population activity because of the strong AL-M2 projection (Wang et al., 2011). The V1-AL differences in decoding during and after stimulus presentation were robust when using different decoders, metrics, and group sizes (Figures S5E–S5H). In conclusion, neuronal populations in AL correlate with behavioral detection of audiovisual stimuli, whereas population activity in V1 correlates with detection of visual-only stimuli.

DISCUSSION

How does neuronal activity in the sensory cortex reflect the processing of visual and multisensory information, and how are these neural processes translated into behavioral detection performance of the organism? To address these questions, we performed two-photon calcium imaging of neuronal populations in V1 and AL during (multi)sensory stimulus detection. Previously, AL neurons were shown to respond to visual stimuli of high temporal and low spatial frequency (Andermann et al., 2011; Marshel et al., 2011), which results from functionally selective input from V1 (Glickfeld et al., 2013). We circumvented this issue by presenting a visual stimulus with a temporal and spatial frequency to which both V1 and AL neurons are responsive. To our knowledge, area AL has not been scrutinized with single-cell precision in the context of multisensory behavior, even though several studies pointed toward the involvement of AL in multisensory behavior (Hirokawa et al., 2008; Wallace et al., 2004).

Increased Stimulus Sensitivity of AL Neurons Does Not Correspond to the Multisensory Enhancement in Detection Behavior

Area AL was more responsive to weak visual stimuli compared to V1, but during strong sensory stimulation, V1 showed a stronger response. The increased sensitivity of AL compared to V1 may be explained by at least two possible mechanisms. First, V1 input is pooled across multiple AL neurons, thereby increasing their sensitivity (Felleman and Van Essen, 1991; Hubel and Wiesel, 1962; Wang and Burkhalter, 2007). Second, highly sensitive V1 neurons may selectively provide input to AL, tantamount to the emergence of tuning to high temporal frequency in AL (Glickfeld et al., 2013). Our computational model of V1-AL interaction (Figure 3) suggests that both mechanisms would be sufficient to explain the observed results, while not excluding alternative mechanisms. Moreover, our computational model provides the testable hypothesis that V1 respond more strongly to a high-contrast visual

stimulus because of saturation due to short-term depression in V1-to-AL synapses. However, the high sensitivity of AL neurons could not explain the multisensory enhancement of detection behavior because, although AL neurons were more sensitive than V1 neurons for low-contrast stimuli, their sensitivity was the same for uni- versus multisensory stimuli.

Cross-Modal Modulation of Neural Responses to High- and Low-Contrast Stimuli

Our results show that cross-modal modulation of visual responses of LII/III neurons is mediated through different subsets of neurons depending on whether the stimulus intensity is high or low (Figures 4 and S3). This is an important extension to previous results, which showed in anesthetized or passively observing mice that the neural coding of multisensory integration depends on stimulus intensity (Ibrahim et al., 2016; Meijer et al., 2017). However, these studies did not reveal that this cross-modal modulation is expressed by different subsets of neurons when stimulus intensity is either high or low. Depending on task requirements, multisensory information may be integrated to subserve different behavioral functions (reviewed in Meijer et al., 2019). When stimulus intensity is high, the need to perform multisensory integration to detect the stimulus is low, because the unisensory stimulus is easily perceivable; however, multisensory integration is still required to differentiate uni- from multisensory stimuli. During low-intensity stimulation, detection of sensory stimuli is more difficult, and integrating information from different modalities can help this process (Gleiss and Kayser, 2012; Hollensteiner et al., 2015). In line with this, we previously showed that the behavioral gain during multisensory detection is highest when the unisensory stimulus constituents are presented around their perceptual thresholds, but is largely absent when the stimulus intensities are higher (Meijer et al., 2017). Therefore, if cross-modal modulation would underlie the animal's enhanced multisensory behavior, one would expect that cross-modal modulation would be strongest for threshold-level stimulus intensities. This was not the case, which argues against the hypothesis that cross-modal modulation can explain the multisensory enhancement in behavioral detection.

AL Neurons Correlate More Closely to the Behavioral Report of Multisensory Stimuli Compared to V1 Neurons

How does neuronal activity in the sensory cortex correlate with detection performance of the organism? To answer this question, we compared the detection performance of the mouse with the neural representation of that sensory information (Parker and Newsome, 1998; Stüttgen et al., 2011). We found that neurometrics of single neurons did not resemble the psychometric performance of the mouse (Figure 2F), whereas population-level neurometrics did, to a higher degree (Figures 2G–2I). We report a seemingly contradictory finding: AL showed a higher sensitivity to visual stimuli compared to V1 but did not predict the behavioral response better in the visual-only condition (Figure 5F). Our computational model showed that the high sensitivity of AL may well be due to the smaller size of area AL as compared to V1 and to the convergence of input that AL receives from V1 (Figure 3). Therefore, stimulus-related information is likely compressed in AL, resulting in a higher sensitivity per neuron, but the population-level sensitivity of both areas in their totality may be similar. Furthermore, a higher stimulus sensitivity does not necessarily imply a stronger correlate to behavioral (hit-miss) reportability: to generate a stronger correlate, a neuron must not only be sensitive to the stimulus but also differentiate between hit-and-miss responses. Thus, even though AL layer-II to -III neurons are highly sensitive to visual-only stimuli, AL output may become relatively decoupled from behavioral decisions (as compared to V1 or to audiovisual configurations), because this output may be more prone to other, ambient factors or task-irrelevant intracortical influences.

Out of the three possible mechanisms examined, AL neurons showed a stronger correlate of multisensory detection behavior than V1 neurons for only one mechanism, namely in the neural correlates of the animal's behavioral report. We found a double dissociation between the correlates of reportability of visual and audiovisual stimuli in V1 and AL. First, neurons in AL, but not in V1, showed a strong modulation in their response between reported and non-reported stimuli when these stimuli were audiovisual. Second, V1 showed a stronger hit-miss differentiation of visual-only stimuli compared to AL. These results suggest that V1 mainly codes unisensory information used for behavioral decision-making, whereas AL predominantly codes multisensory information before being relayed to motor areas to initiate behavioral action. Why would AL be specifically recruited for multisensory detection behavior? Although V1 does not project to premotor areas, several higher-order areas do (namely RL, AM, A, and AL; Wang et al., 2012). This suggests that, depending on the behavioral constraints and stimulus configuration, different higher-order visual areas may be recruited to guide behavior. Our data suggest that the particular case of multisensory detection behavior is contingent on AL, whereas visual-only detection coheres more directly with V1 information, which may influence motor areas via other routes than AL. In other words, depending on task requirements, different sensory cortical regions may be read out for behavioral decision-making, with AL being central in audiovisual trials and V1 in visual-only trials. Importantly, interventional experiments need to be performed to elucidate the causal involvement of AL in multisensory detection behavior (cf. Hirokawa et al., 2008). Sensory information may affect motor areas through different anatomical pathways than those involving AL layer II/III. The role of these candidate pathways, such as those involving the posterior parietal cortex (Hishida et al., 2014; Nikbakht et al., 2018), the superior colliculus (Meredith and Stein, 1986; Wallace et al., 1998), and direct projections of auditory areas to motor cortex (Budinger and Scheich, 2009), awaits further examination.

Potentially, our results could be confounded by licking-related neural activity, especially when hit-and-miss trials were compared. We defined neural activity in each trial as the maximum fluorescence in a 0–500-ms window after stimulus onset; during hit trials it is possible that the animal made a licking response within this time window (Figure S1B). We chose this approach because the rise time of fluorescence signals is relatively slow (~150 ms for GCaMP6f; Chen et al., 2013), which poses a delay for licking activity to influence the fluorescence readout in this time window. However, as a control we replicated the main results using a 0–300-ms time window, which

precludes almost all licking activity (Figures S4H–S4K). All main results held up except for the difference in hit-miss reportability between V1 and AL during visual-only stimulation.

In conclusion, our results show that area AL strongly represents low-contrast stimuli and shows a stronger correlate of audiovisual reportability compared to V1, making it an important candidate area for mediating multisensory stimulus detection behavior. These results can be used to guide future research on multisensory information processing along sensory cortical hierarchies and on cortical-subcortical interactions underlying multisensory integration during active behavior.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Surgical procedures
  - Behavioral task
  - Intrinsic optical signal imaging
  - Two-photon calcium imaging
  - Behavioral analysis
  - Neurometric psychophysics
  - Hit-miss reportability of neurons
  - Decoding stimulus presence
  - Computational model of V1 and AL
  - Hit-miss decoding
  - Neuronal d'
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2020.107636.

ACKNOWLEDGMENTS

We thank Daphnee Chabal, Jesper Vernooij, and Thomas Dolman for help with training animals and Stephan Grzelkowski for developing software for eye tracking. For the use of GCaMP6, we thank Vivek Jayaraman, Rex Kerr, Douglas Kim, Loren Looger, and Karel Svoboda from the GENIE Project, Janelia Research Campus, Howard Hughes Medical Institute. This study was supported by the European Union's Horizon 2020 Framework Programme for Research and Innovation under the specific grant agreement no. 785907 (Human Brain Project SGA2 and SGA3) to C.M.A.P.

AUTHOR CONTRIBUTIONS

Conceptualization, G.T.M., C.S.L., and C.M.A.P.; Methodology, G.T.M. and J.F.M.; Software, P.M. and J.S.M.; Formal Analysis, G.T.M. and P.M.; Investigation, G.T.M.; Resources, J.S.M.; Writing – Original Draft, G.T.M. and C.S.L.; Writing – Review & Editing, C.S.L. and C.M.A.P.; Visualization, G.T.M.; Supervision, C.S.L. and C.M.A.P.; Funding acquisition, C.S.L. and C.M.A.P.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 20, 2019
Revised: January 10, 2020
Accepted: April 21, 2020
Published: May 12, 2020

REFERENCES

Aarts, E., Verhage, M., Veenvliet, J.V., Dolan, C.V., and van der Sluis, S. (2014). A solution to dependency: using multilevel analysis to accommodate nested data. Nat. Neurosci. *17*, 491–496.

Andermann, M.L., Kerlin, A.M., Roumis, D.K., Glickfeld, L.L., and Reid, R.C. (2011). Functional specialization of mouse higher visual cortical areas. Neuron *72*, 1025–1039.

Atilgan, H., Town, S.M., Wood, K.C., Jones, G.P., Maddox, R.K., Lee, A.K.C., and Bizley, J.K. (2018). Integration of visual information in auditory cortex promotes auditory scene analysis through multisensory binding. Neuron *97*, 640–655.e4.

Banks, M.I., Uhlrich, D.J., Smith, P.H., Krause, B.M., and Manning, K.A. (2011). Descending projections from extrastriate visual cortex modulate responses of cells in primary auditory cortex. Cereb. Cortex *21*, 2620–2638.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. J. Stat. Softw. *67*, 1–48.

Bizley, J.K., Shinn-Cunningham, B.G., and Lee, A.K.C. (2012). Nothing is irrelevant in a noisy world: sensory illusions reveal obligatory within-and across-modality integration. J. Neurosci. *32*, 13402–13410.

Brainard, D.H. (1997). The psychophysics toolbox. Spat. Vis. *10*, 433–436.

Britten, K.H., Newsome, W.T., Shadlen, M.N., Celebrini, S., and Movshon, J.A. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque MT. Vis. Neurosci. *13*, 87–100.

Budinger, E., and Scheich, H. (2009). Anatomical connections suitable for the direct processing of neuronal information of different modalities via the rodent primary auditory cortex. Hear. Res. *258*, 16–27.

Chen, T.-W., Wardill, T.J., Sun, Y., Pulver, S.R., Renninger, S.L., Baohan, A., Schreiter, E.R., Kerr, R.A., Orger, M.B., Jayaraman, V., et al. (2013). Ultrasensitive fluorescent proteins for imaging neuronal activity. Nature *499*, 295–300.

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

Driver, J., and Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments. Neuron *57*, 11–23.

Ernst, M.O., and Banks, M.S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. Nature *415*, 429–433.

Faisal, A.A., Selen, L.P.J., and Wolpert, D.M. (2008). Noise in the nervous system. Nat. Rev. Neurosci. *9*, 292–303.

Felleman, D.J., and Van Essen, D.C. (1991). Distributed hierarchical processing in the primate cerebral cortex. Cereb. Cortex *1*, 1–47.

Garrett, M.E., Nauhaus, I., Marshel, J.H., and Callaway, E.M. (2014). Topography and areal organization of mouse visual cortex. J. Neurosci. *34*, 12587–12600.

Ghazanfar, A.A., and Schroeder, C.E. (2006). Is neocortex essentially multisensory? Trends Cogn. Sci. *10*, 278–285.

Gingras, G., Rowland, B.A., and Stein, B.E. (2009). The differing impact of multisensory and unisensory integration on behavior. J. Neurosci. *29*, 4897–4902.

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

Glickfeld, L.L., Andermann, M.L., Bonin, V., and Reid, R.C. (2013). Cortico-cortical projections in mouse visual cortex are functionally target specific. Nat. Neurosci. *16*, 219–226.

Goldey, G.J., Roumis, D.K., Glickfeld, L.L., Kerlin, A.M., Reid, R.C., Bonin, V., Schafer, D.P., and Andermann, M.L. (2014). Removable cranial windows for long-term imaging in awake mice. Nat. Protoc. *9*, 2515–2538.

Goltstein, P.M., Coffey, E.B.J., Roelfsema, P.R., and Pennartz, C.M.A. (2013). In vivo two-photon Ca2+ imaging reveals selective reward effects on stimulus-specific assemblies in mouse visual cortex. J. Neurosci. *33*, 11540–11555.

Gu, Y., Angelaki, D.E., and Deangelis, G.C. (2008). Neural correlates of multisensory cue integration in macaque MSTd. Nat. Neurosci. *11*, 1201–1210.

Guizar-Sicairos, M., Thurman, S.T., and Fienup, J.R. (2008). Efficient subpixel image registration algorithms. Opt. Lett. *33*, 156–158.

Hirokawa, J., Bosch, M., Sakata, S., Sakurai, Y., and Yamamori, T. (2008). Functional role of the secondary visual cortex in multisensory facilitation in rats. Neuroscience *153*, 1402–1417.

Hishida, R., Kudoh, M., and Shibuki, K. (2014). Multimodal cortical sensory pathways revealed by sequential transcranial electrical stimulation in mice. Neurosci. Res. *87*, 49–55.

Hollensteiner, K.J., Pieper, F., Engler, G., König, P., and Engel, A.K. (2015). Crossmodal integration improves sensory detection thresholds in the ferret. PLoS ONE *10*, e0124952.

Hubel, D.H., and Wiesel, T.N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol. *160*, 106–154.

Ibrahim, L.A., Mesik, L., Ji, X.Y., Fang, Q., Li, H.F., Li, Y.T., Zingg, B., Zhang, L.I., and Tao, H.W. (2016). Cross-modality sharpening of visual cortical processing through layer-1-mediated inhibition and disinhibition. Neuron *89*, 1031–1045.

Juavinett, A.L., Nauhaus, I., Garrett, M.E., Zhuang, J., and Callaway, E.M. (2017). Automated identification of mouse visual areas with intrinsic signal imaging. Nat. Protoc. *12*, 32–43.

Kayser, C., Petkov, C.I., and Logothetis, N.K. (2008). Visual modulation of neurons in auditory cortex. Cereb. Cortex *18*, 1560–1574.

Kayser, C., Logothetis, N.K., and Panzeri, S. (2010). Visual enhancement of the information representation in auditory cortex. Curr. Biol. *20*, 19–24.

Kontsevich, L.L., and Tyler, C.W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. Vision Res. *39*, 2729–2737.

Kwon, S.E., Yang, H., Minamisawa, G., and O'Connor, D.H. (2016). Sensory and decision-related activity propagate in a cortical feedback loop during touch perception. Nat. Neurosci. *19*, 1243–1249.

Laramée, M.E., Kurotani, T., Rockland, K.S., Bronchti, G., and Boire, D. (2011). Indirect pathway between the primary auditory and visual cortices through layer V pyramidal neurons in V2L in mouse and the effects of bilateral enucleation. Eur. J. Neurosci. *34*, 65–78.

Leinweber, M., Ward, D.R., Sobczak, J.M., Attinger, A., and Keller, G.B. (2017). A sensorimotor circuit in mouse cortex for visual flow predictions. Neuron *95*, 1420–1432.e5.

Marshel, J.H., Garrett, M.E., Nauhaus, I., and Callaway, E.M. (2011). Functional specialization of seven mouse visual cortical areas. Neuron *72*, 1040–1054.

Meijer, G.T., Montijn, J.S., Pennartz, C.M.A., and Lansink, C.S. (2017). Audiovisual modulation in mouse primary visual cortex depends on cross-modal stimulus configuration and congruency. J. Neurosci. *37*, 8783–8796.

Meijer, G.T., Pie, J.L., Dolman, T.L., Pennartz, C.M.A., and Lansink, C.S. (2018). Audiovisual integration enhances stimulus detection performance in mice. Front. Behav. Neurosci. *12*, 231.

Meijer, G.T., Mertens, P.E.C., Pennartz, C.M.A., Olcese, U., and Lansink, C.S. (2019). The circuit architecture of cortical multisensory processing: distinct functions jointly operating within a common anatomical network. Prog. Neurobiol. *174*, 1–15.

Mejías, J.F., and Torres, J.J. (2008). The role of synaptic facilitation in spike coincidence detection. J. Comput. Neurosci. *24*, 222–234.

Mejias, J.F., and Torres, J.J. (2011). Emergence of resonances in neural systems: the interplay between adaptive threshold and short-term synaptic plasticity. PLoS ONE *6*, e17255.

Meredith, M.A., and Stein, B.E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. J. Neurophysiol. *56*, 640–662.

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

Montijn, J.S., Goltstein, P.M., and Pennartz, C.M. (2015). Mouse V1 population correlates of visual detection rely on heterogeneity within neuronal response patterns. eLife *4*, e10163.

Montijn, J.S., Meijer, G.T., Lansink, C.S., and Pennartz, C.M.A. (2016). Population-level neural codes are robust to single-neuron variability from a multidimensional coding perspective. Cell Rep. *16*, 2486–2498.

Nikbakht, N., Tafreshiha, A., Zoccolan, D., and Diamond, M.E. (2018). Supralinear and supramodal integration of visual and tactile signals in rats: psychophysics and neuronal mechanisms. Neuron *97*, 626–639.e8.

Oh, S.W., Harris, J.A., Ng, L., Winslow, B., Cain, N., Mihalas, S., Wang, Q., Lau, C., Kuan, L., Henry, A.M., et al. (2014). A mesoscale connectome of the mouse brain. Nature *508*, 207–214.

Olcese, U., Iurilli, G., and Medini, P. (2013). Cellular and synaptic architecture of multisensory integration in the mouse neocortex. Neuron *79*, 579–593.

Parker, A.J., and Newsome, W.T. (1998). Sense and the single neuron: probing the physiology of perception. Annu. Rev. Neurosci. *21*, 227–277.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: machine learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.

Pennartz, C.M.A. (2015). The Brain's Representational Power (MIT Press).

Raposo, D., Sheppard, J.P., Schrater, P.R., and Churchland, A.K. (2012). Multisensory decision-making in rats and humans. J. Neurosci. *32*, 3726–3735.

Raposo, D., Kaufman, M.T., and Churchland, A.K. (2014). A category-free neural population supports evolving demands during decision-making. Nat. Neurosci. *17*, 1784–1792.

Stüttgen, M.C., Schwarz, C., and Jäkel, F. (2011). Mapping spikes to sensations. Front. Neurosci. *5*, 125.

Tsodyks, M.V., and Markram, H. (1997). The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. Proc. Natl. Acad. Sci. USA *94*, 719–723.

Wallace, M.T., Meredith, M.A., and Stein, B.E. (1998). Multisensory integration in the superior colliculus of the alert cat. J. Neurophysiol. *80*, 1006–1010.

Wallace, M.T., Ramachandran, R., and Stein, B.E. (2004). A revised view of sensory cortical parcellation. Proc. Natl. Acad. Sci. USA *101*, 2167–2172.

Wang, Q., and Burkhalter, A. (2007). Area map of mouse visual cortex. J. Comp. Neurol. *502*, 339–357.

Wang, Q., Gao, E., and Burkhalter, A. (2011). Gateways of ventral and dorsal streams in mouse visual cortex. J. Neurosci. *31*, 1905–1918.

Wang, Q., Sporns, O., and Burkhalter, A. (2012). Network analysis of corticocortical connections reveals ventral and dorsal processing streams in mouse visual cortex. J. Neurosci. *32*, 4386–4399.

Wichmann, F.A., and Hill, N.J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. Percept. Psychophys. *63*, 1293–1313.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
AAV1.Syn.GCaMP6f.WPRE.SV40	Penn Vector Core	p2822; RRID: Addgene_100837
Experimental Models: Organisms/Strains		
Male C57Bl6 mice	Harlan Sprague Dawley Inc.	N/A
Software and Algorithms		
Code for computational model	Authors (J.F.M.)	https://modeldb.yale.edu/263992
Psychophysics toolbox	Brainard, 1997	http://psychtoolbox.org
Scikit-learn	Pedregosa et al., 2011	http://scikit-learn.org
MATLAB custom code	Authors	N/A
Python custom code	Authors	N/A
Other		
Custom experimental setup	Technology Center, University of Amsterdam	N/A
SP5 Resonant mirror microscope	Leica	N/A
Mai Tai mode-locked Ti:sapphire laser	Spectra-Physics	N/A
Imager 3001 intrinsic optical imaging setup	Optical Imaging Ltd	N/A

RESOURCE AVAILABILITY

Lead Contact

The lead contact for this study is prof. dr. Cyriel M.A. Pennartz (c.m.a.pennartz@uva.nl).

Materials Availability

This study did not generate new unique reagents. Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact.

Data and Code Availability

The code for the computational model is publicly available on ModelDB (https://modeldb.yale.edu/263992). Other code and the dataset supporting the current study are available from the Lead Contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All experiments described in this paper were conducted with approval of the Dutch central committee for animal testing (CCD). Eight weeks old male C57BL/6 mice were obtained from Harlan Sprague Dawley Inc. and housed socially (groups of four). The lights in the animal facility were on a reversed day night cycle (8 AM off, 8 PM on) so that the mice were trained in the behavioral task during their active phase. Mice were kept on a water restriction regime in which they could earn their daily ration of fluid by performing the behavioral task. Their weight was not actively adjusted to a specific percentage of their initial body weight. If a mouse received below 1 mL of fluid on a training day due to poor task performance it received supplemental fluid. Mice were taken off water rationing if their weight dropped below 90% of their average weight of the week before. The age of the mice on the last recording day ranged between 122 – 212 days (n = 9). Both V1 and AL were imaged in most mice (n = 6), but due to insufficient viral expression or occlusion by blood vessels; only V1 (n = 2) or only AL (n = 1) was imaged in some mice.

METHOD DETAILS

Surgical procedures

Mice were implanted with a custom-built titanium headbar to allow head-fixation during training and recording. For analgesic purposes, mice received a subcutaneous injection of 0.05-0.1 mg/kg of buprenorphine prior to the surgery. Anesthesia was induced using 3% isoflurane in 100% oxygen which was lowered during surgery to 1%–2%. The headbar, which contained a circular opening (diameter 7 mm), was positioned over the left hemisphere such that both V1 (A-P: −3.16, M-L: 2.5) and AL (A-P: −2.46, M-L 3.5) were

accessible. The headbar was firmly attached to the skull with C&B Superbond (Sun Medical). The skull was protected from infections by a layer of cyanoacrylate glue (Locktite 401, Henkel) covering the circular opening.

In a second surgery with similar analgesic and anesthetic procedures, the viral vector AAV1.Syn.GCaMP6f.WPRE.SV40 (150-200 nL undiluted; Penn Vector Core) was injected in V1 and AL, of which the locations were determined using Intrinsic Optical Signal imaging (IOS; see below). A circular craniotomy (diameter 3 mm) was made exposing V1 and AL, which were chronically covered by a double-layered coverglass of which the bottom part fitted snugly into the craniotomy, applying a limited pressure on the brain and preventing skull regrowth (Goldey et al., 2014; Montijn et al., 2016).

Behavioral task

Mice were trained in a detection task in which they made a licking response upon detection of a visual (V) or auditory (A) stimulus while they were head-fixed and their body was positioned in a tube to limit body movements. During initial training, mice were presented with 100% contrast square wave drifting gratings (spatial frequency: 0.05 cycles/°, temporal frequency: 1.5 Hz, movement direction: 90°, 210° or 330° randomly chosen each trial) presented in a circular window with a soft gradient boundary of 60 retinal degrees diameter on an isoluminant gray background. Visual stimuli were displayed on a 15-inch TFT screen (refresh rate: 60 Hz) positioned in front of the right eye at a distance of 16 cm and a 45° angle from the midline of the animal. Auditory stimuli were 90 dB (background noise 63.5 dB) pure tones with a center frequency of 15 kHz and being subject to frequency modulation between 14 kHz and 16 kHz. The rate of frequency modulation was 1.5 Hz, which matched the temporal frequency of the visual stimulus (Meijer et al., 2017). Auditory stimuli were amplified (TA1630, Sony) and presented by a tweeter (NEO CD 3.0, Audaphon) positioned 22 cm straight in front of the animal.

Licking responses were registered by an infra-red beam positioned in front of the mouth. Breaking this beam during stimulus presentation triggered reward delivery (6-10 μL of instant formula baby milk). The reward was not delivered immediately but only after the trial had ended. It was made available from a spout that was moved close to the mouth but was positioned out of reach during the trial (i.e., period of stimulus onset to motor response). Stimulus duration decreased over a 3-4 week training period from 5 s at the start of training to 1 s in the final task. Trials were separated by a 3-5 s inter-trial interval. The false alarm rate was assessed using blank trials in which no stimulus was presented and was defined as the percentage of blank trials in which the mouse responded. Sessions showing a false alarm rate of > 45% were excluded from analysis. To reduce the amount of spontaneous licking, a new trial only started when no licks were registered for 1-3 s, which was chosen randomly during each ITI. Mice progressed to the final task when they showed a hit rate for both V and A trials of > 90% and a false alarm rate of < 40% for three consecutive days.

The final task design was the same as described before (Meijer et al., 2018): visual and auditory stimuli were presented around their corresponding detection thresholds using an adaptive Bayesian staircase method (Kontsevich and Tyler, 1999) for each stimulus modality independently (Figures 1C and 1D). This staircase method aims to estimate both the threshold and the slope of the psychometric function. To achieve this, a model estimate of the psychometric function is constructed during the recording session. For each trial, the stimulus amplitude is selected such that it provides the largest expected information gain about threshold and slope values (Kontsevich and Tyler, 1999), and the estimate of the psychometric function is updated with the behavioral information available at the end of the trial. Because the adaptive procedure is built to estimate both threshold and slope, stimulus amplitudes do not converge on the threshold because if only the threshold value is sampled there is no information regarding the slope. Instead stimulus amplitudes *around* the inferred threshold are selected. However, we still consider all these trials to be 'weak' stimulus trials since their intensity is very close to threshold intensity. 'Strong' stimulus trials are those which are far above threshold intensity (e.g., 100% contrast). The staircase method was implemented using PsychStairCase of the Psychophysics Toolbox for MATLAB (Brainard, 1997). Audiovisual (AV) trials were composed of visual and auditory components presented at the last used stimulus intensities in preceding unisensory trials. High intensity uni- and bimodal trials were added to assess the lapse rate of the mouse as an estimate of its motivation to generate lick responses. Recording sessions were composed of visual staircase stimuli (25%), auditory staircase stimuli (25%), audiovisual staircase stimuli (25%), high contrast visual stimuli (100% contrast; 6.25%), high-amplitude auditory stimuli (90 dB; 6.25%), high intensity audiovisual stimuli (6.25%) and blank trials (6.25%). Visual staircase stimulus intensities were constrained to be in the range of 0.5 – 15% visual contrast; auditory staircase amplitudes were restricted to the range of 65 – 80 dB. The order of stimulus presentation was semi-random such that no more than three stimuli of the same type were presented sequentially.

Intrinsic optical signal imaging

Locations for viral injections and two-photon calcium imaging were determined using two separate IOS measurements through the intact skull and cranial window, respectively. The skull or coverglass of a lightly anaesthetized mouse (~1% isoflurane) was illuminated by 810 nm light and recorded by a CCD camera (100 m, Adimec) connected to an Imager 3001 recording setup (Optical Imaging Ltd.; 1 Hz sampling rate). Visual cortex was activated by presenting square wave drifting gratings moving in eight directions sequentially. Each direction was presented for 1 s (total stimulus duration: 8 s) with a 17 s inter-stimulus interval. A mark was put on the skull at the location of V1 which served to target the viral injections. After implantation of the cranial window, a second IOS measurement was performed which provided a more fine-grained map of the visual areas (Figures S1C–S1F). Area AL was determined to be the visually activated area 0.8-1.5 mm lateral and 0.5-1 mm anterior from the center of V1 (Garrett et al., 2014; Juavinett et al., 2017).

Two-photon calcium imaging

Fluorescent GCaMP6f proteins were excited using a Spectra-Physics Mai Tai mode-locked Ti:sapphire laser with a wavelength of 900-910 nm and a laser power exiting the objective of 18 – 22 mW. A plane of 365 × 365 μm at a depth of 130 – 195 μm was imaged using a Leica SP5 resonant mirror scanner operating at a sampling frequency of 25.4 Hz. The same population of neurons was found on each recording day by aligning the imaging plane relative to the vascular pattern (3-6 recording days per imaging location). The scanner produced 61 dB of background noise as measured at the location of the mouse.

Calcium imaging data processing was performed as described previously (Goltstein et al., 2013; Meijer et al., 2017; Montijn et al., 2014, 2016). In short, imaging planes were corrected for X-Y movement (Guizar-Sicairos et al., 2008), cell bodies were semi-manually detected and neuropil correction was performed (Chen et al., 2013). The fluorescence increase over baseline $\Delta F/F_0$ was calculated per imaging frame for every cell body as follows:

$$\Delta F/F_0 = \frac{F^i - F_0}{F_0} \quad \text{Equation 1}$$

Where F^i is the neuropil-corrected fluorescence of the cell body in frame i and F_0 is the baseline fluorescence, defined as the lower half of all fluorescence values in a 30 s sliding window preceding frame i. The first trial of a session always started > 30 s after the start of the calcium recording to allow enough time for the baseline period to elapse. The relative fluorescence activity per trial was defined as the maximum $\Delta F/F_0$ in a 500 ms window after stimulus onset, including blank trials, such that the activity in every trial was mostly related to the stimulus presentation and not the licking response. Although licks could occur before 500 ms, the relatively slow temporal dynamics of the fluorescence response (GCaMP6f rise time: ~150 ms; Chen et al., 2013) imposes a delay on the time for motor-related activity to be observable in visual areas. Therefore, a window of 500 ms after stimulus onset was deemed to consist mostly of sensory evoked activity and minimal motor-related activity (but see Figures S4H–S4K for a replication of the main results with a window of 300 ms after stimulus onset, which was nearly devoid of licking responses). A neuron was characterized as responsive when the fluorescence response in high intensity uni- or bimodal trials exceeded the average fluorescence during blank trials by one SD.

Behavioral analysis

Sensory sensitivity was assessed by fitting psychometric functions to the behavioral data for both visual-only and auditory-only trials using logistic regression (MATLAB function *glmfit* with binomial distribution). To get reliable fits, behavioral data from all recording days was merged. Visual contrast values were subjected to a $\log_{10}$ transform. The detection threshold was determined as the midpoint between the lower and the upper boundary of the fitted function, projected on the abscissa. Multisensory psychometric functions were determined for two conditions: visual supported by auditory (V_A), which included all bimodal trials in which the auditory component was presented below threshold (65 dB - 72.7 dB ± 0.7 dB), and auditory supported by visual (A_V), consisting of all bimodal trials in which the visual component was presented below threshold (0.5% - 1.2% ± 0.2% contrast).

Neurometric psychophysics

Neurometric functions were constructed for the fluorescence response of each neuron to visual stimuli of progressively increasing contrast. Visual-only trials were binned according to the visual contrast in three equally populated bins. For each contrast bin, the mean evoked fluorescence response across all trials was computed and plotted against the $\log_{10}$ transformed mean contrast of those trials. The mean evoked fluorescence response during blank trials was inserted as 0% contrast and the mean fluorescence response during high contrast trials was plotted as 100%. A cumulative Gaussian with four free parameters (Wichmann and Hill, 2001) was fitted to these five points:

$$f(x) = \gamma + (1 - \gamma - \lambda)\left(\frac{1}{2}\left[1 + erf\left(\frac{x - \mu}{\sigma\sqrt{2}}\right)\right]\right) \quad \text{Equation 2}$$

Here, the four free parameters were the lower boundary γ, the upper boundary λ, the mean μ, and the standard deviation σ. The *erf* stands for the error function as implemented by the MATLAB function erf().

Hit-miss reportability of neurons

Ideal observer analysis is instrumental for quantifying how well single neurons are able to differentiate between reported versus non-reported stimulus presentations (Britten et al., 1996; Kwon et al., 2016; Parker and Newsome, 1998) since it incorporates neuronal variability (Faisal et al., 2008). During ROC analysis, discrimination thresholds are placed between two distributions (i.e., neural activity during hit and miss trials) and each threshold position holds that all the points to the left are considered to belong to one of the two distributions (e.g., hits) and the point to the right to the other (e.g., misses). The actual true positive and true negative rate is calculated for every threshold setting and plotted against one another. If the ROC curve deviates from the diagonal, indicating that true positives outweigh false negatives, a neuron carries hit related information. If the ROC curve follows the diagonal, meaning that true positives and false negatives are balanced, the neuron does not code hit-miss differences.

Single-neuron correlates of behavioral report were determined using ROC analysis. Trials were first grouped into hit and miss trials. Each neuron's ability to classify these two classes based on its fluorescence activity was assessed by systematically varying a discrimination threshold over the entire range of fluorescence activity that the neuron exhibited. For every discrimination threshold

the ratio of false positives was plotted against the true positive ratio yielding an ROC curve (MATLAB function *perfcurve*). Hit-miss reportability was calculated as the area under the ROC curve. Significance was determined by first creating a distribution of possible ROC curves by a bootstrapping procedure in which the trial labels (hit/miss) were shuffled (500 iterations). Subsequently, neurons were classified as significantly discriminating if the lower bound of the 95th percentile of the bootstrapped distribution was > 0.5. A hit-miss reportability of 0.5 indicates that the ROC curve lies on the diagonal and the neuron does not signal any detection-related information.

Decoding stimulus presence

Bayesian maximum-likelihood decoding was used to infer from the population activity whether a stimulus was presented (Montijn et al., 2015). Per neuron, a likelihood function was computed for blank trials (stimulus absent) and staircase trials (stimulus present) using leave-one-out cross-validation. For every trial, the posterior probability for both classes was read out from the likelihood functions of all neurons in the population.

$$P(A_{pop}) \propto \prod_{i=1}^{n} P(A_i) \quad \text{Equation 3}$$

The population posterior probability $P(A_{pop})$ was calculated as the product of the posterior probabilities per neuron $P(A_i)$ for all n neurons. Subsequently, a trial was classified as either 'stimulus absent' or 'stimulus present' by selecting the class with the highest population posterior probability. Logistic regression was used to fit a neurometric function to the output of the decoder to assess its sensitivity and to compare its performance to the perceptual behavior of the mouse. We excluded false alarm trials (stimulus reported present in blank trials) in computation of the psychometric function for the mice because these trials cannot be classified as 'stimulus present' by the decoder since there was no neural signal that a stimulus was presented during these trials. When false alarm trials were included in this analysis, the neurometric and psychometric functions strongly deviated from one another indicating that the decoder distilled a sensory signal from the neuronal response instead of a motor signal.

Computational model of V1 and AL

The computational model used to explain the differences in response strength of V1 and AL consisted of a network of 400 excitatory V1 neurons sending feedforward projections to 100 excitatory AL neurons. Each AL neuron received projections from all neurons in the V1 subpopulation considered (some level of sparseness did not alter our results). Each neuron was modeled as a leaky integrate-and-fire unit (see Supplemental Information). The network was provided with a range of input strengths up to 12 mV per neuron to simulate 100% contrast visual stimulation. Following previous work on the computational impact of fast synaptic plasticity on signal detection (Mejías and Torres, 2008, 2011), synapses were assumed to display short-term depression, and were modeled using the Tsodyks-Markram model (Tsodyks and Markram, 1997). See also Methods S1.

Hit-miss decoding

Decoding of hit-miss behavior was performed by selecting neural responses for each trial at a given imaging frame. At every frame, 300 groups of 30 neurons were randomly drawn from the population without replacement. Trial outcome (hit or miss) was classified using a Naive Bayes decoder with Gaussian likelihood, as implemented in Scikit-learn (Pedregosa et al., 2011), averaged over a 5-fold cross-validation. Decoding performance was assessed by the F_1 score:

$$F_1 = \frac{2}{1/recall + 1/precision} \quad \text{Equation 4}$$

Here, recall is the fraction of correctly classified miss trials out of all miss trials and precision is the fraction of actual miss trials out of all trials which were classified as miss trials. The difference between V1 and AL decoding performance at each imaging frame was determined by fitting a linear mixed effects model (Bates et al., 2015) with fixed effect (V1-AL) and random intercept to accommodate for the multiple observations (randomly drawn groups of neurons) nested within each animal (Aarts et al., 2014).

Neuronal d'

Whether a neuron exhibited cross-modal response enhancement or suppression was determined by computing a neuronal *d'* metric per neuron. This was calculated as follows:

$$d' = \frac{\mu_V - \mu_{AV}}{\sqrt{(\sigma_V^2 + \sigma_{AV}^2)/2}} \quad \text{Equation 5}$$

Where μ_V is the mean response of a neuron during visual-only trials, μ_{AV} is the mean response during audiovisual trials and $\sigma^2{}_V$ and $\sigma^2{}_{AV}$ are the respective variances. The significance of response modulation was determined by permutation testing of the absolute d' (500 shuffles of V and AV trial labels). If the absolute *d'* of a neuron was > 95th percentile of the shuffled distribution, the neuron's *d'* was considered significantly higher than expected by chance (5% significance).

QUANTIFICATION AND STATISTICAL ANALYSIS

For every statistical test we first checked whether the data were normally distributed using a Jarque-Bera test for normality (H = 0: normal distribution). By default, we used parametric statistical testing, if however, the Jarque-Bera test was significant ($p < 0.05$), we used equivalent non-parametric statistical tests (Wilcoxon's Rank Sum test for t test, Kruskal-Wallis for ANOVA). By default, values are reported as mean ± s.e.m. and error bars and shading indicate s.e.m. in all figures.