Manuscript Title:

Pure tones modulate the representation of orientation and direction in the primary visual cortex

Abbreviated Title: Sound modulation of the representation of orientation in V1

Authors and Affiliations:

John P. McClure Jr.[1], Pierre-Olivier Polack[1]

1. Center for Molecular and Behavioral Neuroscience. Rutgers University Newark. 197 University Avenue. Newark, NJ 07102. USA.

Author Contributions:

J.P.M. and P-O.P. designed the study; J.P.M. performed the experiments; J.P.M. and P-O.P. analyzed the data; P-O.P. and J.P.M wrote the manuscript

Correspondence should be addressed to:

Correspondence should be addressed to Pierre-Olivier Polack, 197 University Avenue. Newark, NJ 07102. USA. E-mail: polack.po@rutgers.edu

Number of figures: 6
Number of tables: 0
Number of multimedia: 0
Number of words for Abstract: 249
Number of words for Introduction: 561
Number of words for Discussion: 1,384

Acknowledgements:

The authors thank Jacob Duijnhouwer and Bart Krekelberg for assistance and advice on data analysis; Denis Paré and Bart Krekelberg for critical reading of this manuscript; Drew Headley for his assistance when building the EEG experimental setup. The authors also thank Vivek Jayaraman, Ph.D., Douglas S. Kim, Ph.D., Loren L. Looger, Ph.D., and Karel Svoboda, Ph.D. from the GENIE Project, Janelia Research Campus, Howard Hughes Medical Institute for making the AAV1.eSyn.GCaMP6f.WPRE.SV40 available.

Conflict of Interest: The authors report no conflict of interest.

Funding Sources: This work was supported by a Rutgers University-Newark's startup fund and a grant from the Whitehall Foundation (2015-08-69).

Abstract

Multimodal sensory integration facilitates the generation of a unified and coherent perception of the environment. It is now well established that unimodal sensory perceptions, such as vision, are improved in multisensory contexts. While multimodal integration is primarily performed by dedicated multisensory brain regions such as the association cortices or the superior colliculus, recent studies have shown that multisensory interactions also occur in primary sensory cortices. In particular, sounds were shown to modulate the responses of neurons located in layers 2/3 (L2/3) of the mouse primary visual cortex (V1). Yet, the net effect of sound modulation at the V1 population level remained unclear. Here, we performed two-photon calcium imaging in awake mice to compare the representation of the orientation and the direction of drifting gratings by V1 L2/3 neurons in unimodal (visual only) or multi-modal (audiovisual) conditions. We found that sound modulation depended on the tuning properties (orientation and direction selectivity) and response amplitudes of V1 L2/3 neurons. Sounds potentiated the responses of neurons that were highly tuned to the cue's orientation and direction but weakly active in the unimodal context, following the principle of inverse effectiveness of multimodal integration. Moreover, sound suppressed the responses of neurons untuned for the orientation and/or the direction of the visual cue. Altogether, sound modulation improved the representation of the orientation and direction of the visual stimulus in V1 L2/3. Namely, visual stimuli presented with auditory stimuli recruited a neuronal population better tuned to the visual stimulus orientation and direction than when presented alone.

Introduction

Animals are continuously bombarded with multisensory information that must be identified and integrated before the selection, planning, and execution of actions adapted to the environment. Combining sensory inputs from different modalities was shown to improve detection (Odgaard et al., 2004; Lippert et al., 2007; Gleiss and Kayser, 2014) and discrimination thresholds (Vroomen and de Gelder, 2000), as well as decrease reaction times for object perception in humans (Hershenson, 1962; Posner et al., 1976; Gielen et al., 1983). Initially, these cross-modal interactions were thought to take place solely in higher-order multisensory cortices such as the posterior parietal cortex, a region that receives converging inputs from multiple primary sensory areas and plays an important role in cross-modal integration (Molholm et al., 2006; Song et al., 2017). Yet, the existence of direct long-range connections between primary sensory areas in primates (Falchier et al., 2002; Rockland and Ojima, 2003; Cappe and Barone, 2005) and in mice (Iurilli et al., 2012; Ibrahim et al., 2016) provided an anatomical substrate for potential cross-modal interactions at an early stage of sensory processing. Many studies in rodents, as well as human and non-human primates, have now provided compelling evidence for multimodal interactions between primary sensory cortices (Ghazanfar and Schroeder, 2006; Driver and Noesselt, 2008; Petro et al., 2017). Such evidence include the modulation of visually-evoked event-related potentials (ERPs) in V1 by sound (Giard and Peronnet, 1999), auditory cortical neurons modulated by visual and somatosensory stimuli (Brosch et al., 2005), and visual information integrated in A1 (Atilgan et al., 2018).

To determine the nature of cross-modal sensory interactions at the earliest stages of cortical processing, several laboratories examined how sounds influence the responses of V1 neurons to the presentation of oriented visual stimuli (Iurilli et al., 2012; Ibrahim et al., 2016; Meijer et al., 2017). However, these studies yielded contradictory results. In one study, whole-cell recordings performed in anesthetized mice showed that short 50-millisecond broadband noise bursts hyperpolarized V1 neurons, reducing their responses to the presentation of oriented bars (Iurilli et al., 2012). In another

study, white noise bursts were found to significantly enhance the responses of V1 neurons to their preferred orientation while decreasing their responses to orthogonal stimuli (Ibrahim et al., 2016), suggesting that sounds improve the representation of orientation in V1. In contrast with the latter study, the responses of neurons to the presentation of their preferred orientation were then found to be either not modulated or suppressed depending on the nature of the sound presented simultaneously with the visual stimulus (Meijer et al., 2017).

Given these divergent conclusions, the impact of sounds on the representation of orientation in V1 remained unclear. To address this question, we measured the response evoked by oriented stimuli presented alone (visual-only) or paired to a pure tone (audiovisual context) in thousands of V1 L2/3 neurons using calcium imaging in awake mice. This approach allowed us to test a possible solution for the contradictory results obtained so far, namely that sounds differentially alter visual responsiveness depending on the cells' tuning properties and unimodal response amplitudes. Moreover, the possibility that sounds affect the direction selectivity of V1 neurons was so far never assessed. Here, we show that sounds improve the representation of the orientation and the direction of the visual stimulus in V1 L2/3 by differentially modulating the neurons' responses as a function of their orientation and direction tuning properties and response amplitudes.

Materials & Methods

All the procedures described below have been approved by the Institutional Animal Care and Use Committee of Rutgers University, in agreement with the Guide for the Care and Use of Laboratory Animals (DHHS).

Surgery.

Head-bar implants: 10 minutes after systemic injection of an analgesic (carprofen, 5 mg per kg of body weight), adult (3-6 months old) male and female Gad2-IRES-Cre (Jackson stock #019022) x

Ai9 (Jackson stock #007909) mice were anesthetized with isoflurane (5% induction, 1.2% maintenance) and placed in a stereotaxic frame. Body temperature was kept at 37°C using a feedback-controlled heating pad. Pressure points and incision sites were injected with lidocaine (2%). Eyes were protected from desiccation with artificial tear ointment (Dechra). Next, the skin covering the skull was incised and a custom-made lightweight metal head-bar was glued to the skull using Vetbond (3M). In addition, a large recording chamber capable of retaining the water necessary for using a water-immersion objective was built using dental cement (Ortho-Jet, Lang). Mice recovered from surgery for 5 days, during which amoxicillin was administered in drinking water (0.25 mg/mL).

AAV virus injection: After recovery, mice were anesthetized using isoflurane as described above. A circular craniotomy (diameter = 3 mm) was performed above V1. The AAV vector AAV1.eSyn.GCaMP6f.WPRE.SV40 (UPenn Vector Core) carrying the gene of the fluorescent calcium sensor GCaMP6f was injected at three sites 500 µm apart around the center of V1 (stereotaxic coordinates: -4.0 mm AP, +2.2 mm ML from bregma) using a MicroSyringe Pump Controller Micro 4 (World Precision Instruments, WPI) at a rate of 30 nl/min. Injections started at a depth of 550 µm below the pial surface and the tip of the pipette was raised in steps of 100 µm during the injection, up to a depth of 200 µm below the dura surface. The total volume injected across all depths was 0.7 µl. After removal of the injection pipette, a 3-mm-diameter coverslip was placed over the dura, such that the coverslip fits entirely in the craniotomy and was flush with the skull surface. The coverslip was kept in place using Vetbond and dental cement. Mice were left to recover from the surgery for at least 3 weeks to obtain a satisfactory gene expression.

EEG electrode implants: Adult (3-6 months old) male and female C57BL/6 mice were implanted with head-bars (see procedure description above). After recovery from the head-bar implant surgery, mice were anesthetized using isoflurane as described above. Small craniotomies were performed above V1 (AP: -4.0 mm, ML: +2.2 mm), M1 (AP: +1.7 mm, ML: +2.0 mm), and A1 (AP: -2.5 mm, ML: +4.5 mm). Electrodes made of stainless-steel wire, isolated by polyester (diameter, 0.125 mm; FE245840; Goodfellow), and already soldered to a connector were implanted between the bone and

the dura. A reference was implanted above the cerebellum. Finally, the skull and wires were covered with dental cement.

Imaging

During the last week of recovery, mice were trained to stay on a spherical treadmill consisting of a ball floating on a small cushion of air that allowed for full 2D movement (Dombeck et al., 2007). During three daily twenty-minute sessions, the mouse head-bar was fixed to a post holding the mouse on the apex of the spherical treadmill (Figure 1A). Ball motion was tracked by an IR camera taking pictures of the ball at 30 Hz. Eye motion was monitored at 15 Hz using a second IR camera imaging the reflection of the eye on an infrared dichroic mirror. Functional imaging was performed at 15 frames per second using a resonant scanning two-photon microscope (Neurolabware) powered by a Ti-Sapphire Ultra-2 laser (Coherent) set at 910 nm. The laser beam was focused 200 microns below the cortical surface using a 16×, 0.8 NA Nikon water-immersion objective. The objective was tilted 30 degrees such that the objective lens was parallel to the dura surface. Laser power was kept below 70 mW. Frames (512x796 pixels) were acquired using the software Scanbox developed by Neurolabware.

EEG recordings

EEG electrode signals were pre-amplified (Analog Devices AD4177-4) at the head of the animal, fed into a four channel EEG amplifier (Model 1700 Differential AC Amplifier, A-M System), filtered between 1.0 and 5 kHz, and digitized at 1 kHz (NiDAQ, National Instrument) along with signals allowing synchronization with visual stimuli, locomotion, and an eye tracking system.

Audiovisual stimuli.

A gamma-corrected 40-cm diagonal LCD monitor was placed 30 cm from the eye contralateral to the craniotomy such that it covered the entire monocular visual field. Sounds were produced by a speaker located immediately below the center of the screen. Auditory or visual-only stimuli, as well as

audiovisual stimuli, were presented alternatively in blocks of at least 30 trials. Visual and auditory stimuli were generated in MATLAB (MathWorks) using the Psychtoolbox (Brainard, 1997). At the beginning of the recording session, the modality of the first block was randomly selected between visual and auditory. Visual stimuli consisted of the presentation of one of two vertical sinewave gratings that drifted toward the right and were rotated clockwise by 45° and 135° (temporal frequency = 2 Hz, spatial frequency = 0.04 cycle per degree, contrast = 75%; duration: 3 seconds; intertrial interval: 3 seconds). Auditory stimuli consisted of the presentation of one of two sine wave pure tones (10 kHz and 5 kHz; 78 dB, background noise 69 dB). Each audiovisual trial resulted from the random combination of one of the two pure tones with one of the two drifting gratings (four possibilities: 5 kHz tone + 45° drifting grating, 10 kHz tone + 45° drifting grating, 5 kHz tone + 135° drifting grating, and 10 kHz tone + 135° drifting grating; Figure 1B). This design of random pairings between two auditory and two visual stimuli was adopted to minimize the possibility of unwanted learned associations between the visual and auditory stimuli. At the end of the imaging session, we assessed the orientation tuning of the imaged neurons. The orientation tuning block consisted of the presentation of a series of drifting sine wave gratings (12 orientations evenly spaced by 30 degrees and randomly permuted; Figure 1D). The spatiotemporal parameters of the orientation tuning stimuli were identical to those for the visual only, auditory only, and audiovisual stimuli except for their duration (temporal frequency = 2 Hz, spatial frequency = 0.04 cycle per degree, contrast = 75%; duration: 1.5 seconds; intertrial interval: 3 seconds). A photodiode located at the top left corner of the screen was used to detect the exact timing of the visual stimuli onset and offset. This signal was acquired along with the following signals: 1) a signal provided by the two-photon microscope, which indicated the onset of each frame, and 2) two analog signals encoding the orientation of the drifting grating and the frequency of the auditory stimulus. These signals were digitized (NiDAQ, National Instruments) and recorded with the software WinEDR (John Dempster, University of Strathclyde). Imaging sessions started by recording one thousand frames with the green and red channels. The red channel was used to exclude GABAergic neurons from analysis.

Data analysis

All the analyses detailed below were performed using custom MATLAB routines.

Imaging data pre-processing: Calcium imaging frames were realigned offline to remove movement artifacts using the Scanbox algorithm (Neurolabware). A region of interest (ROI) was determined for each neuron using a semi-automatic segmentation routine. For every frame, the fluorescence level was averaged across the pixel of the ROI. Potential contamination of the soma fluorescence by the local neuropil was removed by subtracting the mean fluorescence of a 2-5-micron ring surrounding the neuron's ROI, excluding the soma of neighboring neurons, and then adding the median value across time of the subtracted background. We then computed the fractional fluorescence from the background subtracted fluorescence data. The fractional fluorescence ($dF/F = (F - F_0) / F_0$), was calculated with F_0 defined as the median of the raw fluorescence measured during every inter-trial interval. Trials were then sorted by stimulus. The mean dF/F measured during a 1.5 second intertrial period immediately preceding each visual stimulation was subtracted from the dF/F measured during the trial. Then we calculated the median across trials for each time point of the stimulus presentation. The median was preferred over the mean because the trial-to-trial variability of neuronal activity make the mean prone to follow outliers. We defined the amplitude of the neuronal response as the mean of the median response across the duration of the stimulus. To account for the intertrial activity of the neurons and avoid that constantly active neurons can be considered as responding neurons, the response amplitude was expressed in z-score of the intertrial activity by dividing the amplitude value expressed in dF/F by the standard deviation of the dF/F measured after combining all the intertrial intervals of the experiment.

Orientation tuning: For each trial, we computed the summed dF/F measured during the 1.5 second presentation of the 12 different drifting gratings used to construct the tuning curve (Figure 1D). Using a resampling-based Bayesian method on the summed dF/F of individual trials (Cronin et al., 2010), we estimated the best orientation tuning curve out of four models (constant, circular Gaussian

180, circular Gaussian 360, direction selective circular Gaussian). The preferred orientation of the neuron was defined as the orientation for which the value of the estimated tuning curve (TC) was at its maximum. The orientation selectivity index was defined as OSI = ($TC_{Preferred}$ - $TC_{Orthogonal}$) / ($TC_{Preferred}$ + $TC_{Orthogonal}$). The Direction Selectivity Index was calculated as DSI = ($TC_{Preferred}$ – $TC_{Opposite}$) / ($TC_{Preferred}$ + $TC_{Opposite}$) (Niell and Stryker, 2008). Neurons best fitted by a constant fit were excluded from analysis as they did not carry information on the orientation of the visual stimulus presented.

Sound modulation: A sound modulation index (SMI) was used to quantify the changes in V1 neuronal responses to pure tones in the audiovisual condition. We computed SMI from the visually evoked response measured in the visual only and audiovisual contexts (R_V and R_{AV}, respectively) using the following equation (Meijer et al., 2017):

$$SMI = \frac{R_{AV} - R_V}{R_{AV} + R_V}$$

Values between 0 and 1 indicate a positive modulation or a potentiation of V1 neuronal activity; a SMI of 0.2 corresponds to a potentiation of 50%. Values between 0 and -1 indicate a negative modulation or suppression of activity; a SMI of -0.2 corresponds to a suppression of 33%.

Frequency analysis.

Power spectra were computed as the absolute value of the Fast Fourier transform signal (obtained using a Hanning window) divided by N / (2* 0.375) to satisfy Parseval's Theorem (N represents the number of points of the ECoG signal segment). Spectra were then normalized by applying a 1/f correction. For each frequency band, the normalized power was calculated as the mean of the power spectrum curve.

Statistics

Statistical significance was calculated using ANOVAs or Kruskal-Wallis one-way ANOVAs on ranks using MATLAB. A Normality Test (Shapiro-Wilk) was always performed before each test to assess the

normality of the sample. A Pairwise Multiple Comparison was performed using the Tukey test (ANOVA) or Dunn-Sidak Methods (Kruskal-Wallis) with a significance threshold of $p < 0.05$. Circular statistics were computed with the Circular Statistics Toolbox for MATLAB (Berens, 2009).

Bootstrapping: Bootstrapping (1,000 samples) was performed using MATLAB's bootstrap sampling function (bootstrap) and two-tailed confidence intervals at the alpha level 0.05 were defined as the 2.5 and 97.5 percentile of the bootstrapped population. For each iteration of the bootstrapping (n = 1,000 iterations), we computed the direction of the circular mean vector from a resampled population of neurons active in the V_{only}, AV_{10kHz}, and AV_{5kHz} conditions. Then, we calculated the difference between the results obtained for each of the three conditions (Direction AV_{10kHz} - Direction V_{only}; Direction AV_{5kHz} - Direction V_{only}; Direction AV_{10kHz} - Direction AV_{5kHz}). Finally, we calculated the mean and 95% confidence interval (CI) from those 1,000 differences (AV_{10kHz} - V_{only}; AV_{5kHz} - V_{only}; AV_{10kHz} - AV_{5kHz}). The difference between groups was considered significant if the values of the CI boundaries were of same sign.

Results

We imaged the activity of 3,355 neurons (22 recording sessions in 14 animals) located in layers 2 and 3 (L2/3) of V1 using the genetically encoded calcium sensor GCaMP6f (Chen et al., 2013). During recording sessions, mice were placed head-fixed on a spherical treadmill in front of a speaker and a screen that covered the visual field of the eye contralateral to the imaged V1 (Figure 1A). Calcium imaging was performed while alternating the presentation of unimodal (visual only) and bimodal (audiovisual) blocks of 30 or more trials (Figure 1B). During unimodal and audiovisual blocks, visual stimuli consisted of one of two orthogonal drifting gratings, hereafter termed test stimuli (orientations: 45° or 135°, duration: 3 seconds, temporal frequency: 2 Hz, spatial frequency: 0.04 cycle per degree, contrast: 75%; Figure 1B). Orthogonal orientations were chosen to activate two very distinct neuronal populations in V1 with minimal response overlap (Figure 1C). During audiovisual blocks, the test stimuli were paired with one of two pure tones (low pitch: 5 kHz, or high

pitch: 10 kHz; Figure 1B, C). At the end of each recording session, we assessed the tuning for orientation and direction of all the imaged neurons by presenting series of drifting gratings of twelve evenly spaced orientations (gratings of six distinct orientations traveling in two opposite directions, Figure 1D). The orientations used to generate the tuning curves did not overlap with the orientations of the test stimuli, but had identical temporal frequency (2 Hz), spatial frequency (0.04 cycle per degree), and contrast (75%). Hence, we were able to compare neuronal responses to the presentation of the same visual stimulus in the absence or presence of pure tones and relate any modulation of the visually evoked activity to the neuron's orientation tuning properties (Figure 1C, D). Our experimental setup also allowed us to distinguish excitatory from inhibitory neurons by identifying GAD2 positive (GABAergic) neurons expressing the red fluorescent protein td-tomato. We found that 807 of the 3,355 recorded neurons expressed td-tomato (24% of the imaged neurons), matching earlier estimates of the proportion of GABAergic neurons in the cerebral cortex (Markram et al., 2004). As interneurons account solely for intra-cortical modulation of activity within V1 and do not transfer information to the next stage of visual processing, they were not included in the estimate of the representation of orientation in V1. Similarly, we excluded neurons that were not tuned for orientation (n = 172).

First, we investigated how neurons imaged in V1 L2/3 responded to the presentation of the test stimulus as a function of the angular distance (δ orientation) between their preferred orientation and the orientation of the test stimulus (Figure 1E, 'visual only', test stimulus orientation: 45°). The response amplitude of each neuron was defined as the mean, over the duration of the stimulus, of the median response across trials and was expressed in standard deviations of the neuron's own intertrial spontaneous activity. As expected, the representation of a full screen drifting grating was very sparse (Vinje and Gallant, 2000; Wohrer et al., 2013). The presentation of the test stimulus caused a negative response in most of the imaged neurons (-0.119 ± 0.005 z-scores; t-test: $p < 0.001$; Figure 1E, inset). A minute fraction of neurons displayed large positive responses to the presentation of the test stimulus (Figure 1E; 1.5% of cells with responses greater than 2 z-scores).

As the distribution of responses across the imaged population was fitted by a Gaussian curve (r^2 = 0.99; Figure 1E inset), we could determine an activity threshold above which the response of a neuron was significantly greater than the rest of the population (threshold = 0.4; one-tailed test with α = 0.05; inset Figure 1E, vertical dotted line). We focused the following analysis on neurons with responses above this activity threshold (hereafter termed "active neurons"). The populations of neurons active during the presentation of one of the two test stimuli in the visual only (test stimulus: 135°; Figure 1F) or the two audiovisual contexts (135° test stim + 5 KHz or 10 KHz tone; Figure 1G) shared similarities. In the three conditions (one unimodal and two audiovisual), a nearly identical proportion of neurons had responses that exceeded the activity threshold (visual only context: 5.85%, both audiovisual contexts: 5.68%; χ^2 test: p = 0.85). However, the distribution of the orientations of active neurons in the audiovisual context seemed more concentrated around the orientation and direction of the test stimulus. Therefore, we sought to test the hypothesis that tones increase the orientation selectivity of the V1 population response.

Representation of orientation and direction in the unimodal and audiovisual contexts.

The distributions of preferred orientations in the unimodal and audiovisual contexts (Figure 2A) suggested a higher probability of recruiting neurons with preferred orientations close (±30°) to the orientation of the test stimulus in the audiovisual than in the unimodal context. Indeed, while the circular means of the three distributions pointed similarly toward the orientation of the test stimulus (Figure 2A, inset), the resultant vectors were longer in the audiovisual contexts, again suggesting a higher specificity of orientation representation in the presence of sounds. Using bootstrapping, we determined whether the length and orientation of the circular mean vectors computed in the three conditions differed statistically (Figure 2B). We found that while having similar orientations (Figure 2C, top panel green and red), the circular mean vectors were significantly longer in the audiovisual contexts than in the unimodal context (Figure 2C, middle panel; $p<0.05$). Moreover, the

concentration of the circular distribution (Kappa parameter of the von Mises distribution for circular data) was significantly greater in the audiovisual than in the unimodal context (Figure 2C, bottom panel; $p<0.05$). This confirmed that active neurons were better tuned to the orientation of the test stimulus in the audiovisual than in the unimodal context. On the other hand, we did not find such differences between the mean vectors of the populations recruited by the two tones (Figure 2C, yellow).

Sound modulation as a function of the neurons' orientation preferences.

Tones could enhance the resultant length of the circular mean vector by potentiating the responses of V1 neurons with preferred orientations near that of the test stimulus and/or suppressing neurons with preferred orientations orthogonal or opposite to that of the test stimulus. To test these possibilities, we computed for each tone the sound modulation index (SMI) of every neuron whose response amplitude for one of the two test stimuli (45° or 135° visual cue orientation) was greater than the activity threshold in at least one of the three contexts (V_{only}, AV_{5KHz}, and AV_{10KHz}). Neurons were separated into three groups: neurons tuned to the orientation and direction of the test stimulus (matching neurons: preferred orientation <30° from the test stimulus; n=346), neurons tuned to orientations orthogonal to the test stimulus (orthogonal neurons: preferred orientation ranging between 30° and 150° from the test stimulus; n=302), and neurons tuned to the drifting grating's orientation, but opposite direction (opposite neurons: preferred orientation >150° from the test stimulus orientation; n=156; Figure 3A). The sound modulation of matching neurons was significantly greater than that of opposite and orthogonal neurons (Figure 3A: Kruskal-Wallis test: $\chi^2 = 21.3$, $p = 2.4\text{x}10^{-5}$; Dunn-Sidak post-hoc comparison: $p_{matching\text{-}ortho} = 0.0015$, $p_{matching\text{-}opposite} = 0.0001$ and $p_{ortho\text{-}opposite} = 0.50$). In the audiovisual context, the response of a majority of matching neurons was enhanced two-sided sign test: $p = 0.03$), while a majority of opposite neurons were suppressed (two-sided sign test: $p = 0.0003$).

As the question to whether sound modulation in V1 follows the principle of inverse effectiveness (stating that multisensory enhancement is most prominent when individual unimodal inputs are weak; Serino et al., 2007; Gleiss and Kayser, 2012) was debated (Ibrahim et al., 2016; Meijer et al., 2017), we repeated our analysis, instead focusing on the most highly responsive neurons (i.e. with response amplitudes significantly greater [> 2 z-scores] than spontaneous activity in at least 50% of the trials; Figure 3B). Here too, sound modulation of visual responses differed across the three groups of neurons (Kruskal-Wallis test: $\chi^2 = 17.9$, $p = 1.3x10^{-4}$). However, in this case, matching neurons were not significantly modulated (two-sided sign test: $p = 0.30$) while orthogonal and opposite neurons were both strongly negatively modulated (two-sided sign test: $p = 0.03$ for both groups; Figure 3B). To further understand how sound modulation of visual responses depends on the neuron's response amplitude and preferred orientation, we compared the SMI of matching neurons that were either inactive (response < 0.4 z-score), moderately active (0.4 < response < 2.0 z-score), or highly active (response > 2.0 z-score) when the test stimulus was presented in the unimodal context (Figure 3C). A majority of matching neurons inactive in the unimodal context were potentiated by sound (two-sided sign test: $p = 0.03$), while matching neurons moderately active or highly active in the unimodal context were either not modulated or suppressed (two-sided sign test: $p = 0.12$ and two-sided sign test: $p = 0.03$ respectively; Figure 3C, left panel). Moderately active and highly active orthogonal (Kruskal-Wallis test: $\chi^2 = 14.9$, $p = 0.006$; Dunn-Sidak post-hoc comparison: $p_{inactive\text{-}active} = 0.06$, $p_{inactive\text{-}high} = 0.003$, $p_{active\text{-}high} = 0.47$; Figure 3C, center panel) and opposite neurons (Kruskal-Wallis test: $\chi^2 = 39.8$, $p = 2.3x10^{-9}$; Dunn-Sidak post-hoc comparison: $p_{inactive\text{-}active} = 1.3x10^{-6}$, $p_{inactive\text{-}high} = 0.0002$, $p_{active\text{-}high} = 0.68$; Figure 3C, right panel) were similarly suppressed. Hence, our data showed that sound suppressed the response of neurons tuned for orthogonal orientations or the opposite direction to the test stimulus, and that the enhanced response in the audiovisual context of neurons whose preferred orientation matched the orientation of the test stimulus followed the principle of inverse effectiveness.

Sound modulation as a function of the orientation and direction selectivity of V1 neurons.

We then investigated whether sound modulation was expressed differentially depending on the neurons' orientation selectivity (Figure 4). We found a positive linear correlation between sound modulation and the orientation selectivity index (OSI) in matching neurons (Figure 4A, left panel; $p = 0.03$). Sounds tended to enhance the responses of matching neurons that had a high orientation selectivity (OSI > 0.75; two-sided sign test: $p< 0.001$), but not those of less selective neurons (OSI < 0.75; two-sided sign test: $p= 0.6$). The relationship between OSI and sound modulation was only seen in matching neurons (Figure 4A), not in orthogonal ($p = 0.95$; Figure 4B) and opposite neurons ($p = 0.90$; Figure 4C). Next, we tested whether the relation between orientation selectivity and sound modulation in matching neurons was expressed differentially depending on activity levels (Figure 4D). We found that sound mostly enhanced the response of matching neurons that were inactive and moderately active in the visual-only context but were highly orientation selective for the test stimulus orientation (OSI > 0.75; two-sided sign tests: $p < 0.001$ $p < 0.05$, respectively). Inactive matching neurons that were highly orientation selective were significantly more potentiated than moderately active and highly active neurons with similar orientation tuning (Figure 4D (left panel), Kruskal-Wallis: $\chi^2 = 35.4$, $p = 2.1x10^{-8}$; Dunn-Sidak post-hoc comparison: $p_{inactive\text{-}active} = 0.02$, $p_{inactive\text{-}high} = 1.3x10^{-8}$, $p_{active\text{-}high} = 0.02$). On the other hand, sound significantly suppressed matching neurons that were less orientation selective (OSI < 0.75) regardless of their responsiveness (Figure 4D, middle and right panel).

Next, we examined whether sound modulation was expressed differentially depending on the neurons' direction selectivity (Figure 5). No linear correlation between sound modulation and DSI was found for matching and orthogonal neurons (Figure 5A, B; $p = 0.73$ and $p = 0.58$ respectively). However, a negative correlation was found for opposite neurons (Linear fit: $p = 0.015$; Figure 5C). We tested whether the relationship between direction selectivity and sound modulation in opposite neurons was expressed differentially depending on the amplitude of the response measured in the

unimodal context (Figure 5D). We did not find significant differences between neurons moderately and highly active in the unimodal context that were either poorly direction selective neurons (DSI < 0.25; Kruskal-Wallis: χ^2 = 15.2, p = 0.0005; Dunn-Sidak post-hoc comparison: $p_{inactive\text{-}active}$ =0.01, $p_{inactive\text{-}high}$ = 0.02, $p_{active\text{-}high}$ = 0.67; Figure 5D left panel) or moderately direction selective neurons (0.25 < DSI < 0.75; Kruskal-Wallis: χ^2 = 27.1, p = $1.3x10^{-6}$; Dunn-Sidak post-hoc comparison: $p_{inactive\text{-}active}$ < 0.0001, $p_{inactive\text{-}high}$ = 0.01, $p_{active\text{-}high}$ = 0.99).

Locomotion and arousal are similar in visual only and audiovisual contexts

As locomotion and arousal were both previously found to modulate the response of V1 neurons to visual stimuli (Niell and Stryker, 2010; Polack et al., 2013; McGinley et al., 2015; Vinck et al., 2015), we tested whether the simultaneous presentation of a sound with the test stimulus triggered an increase in locomotion or arousal that could explain our results. For this, we performed another set of experiments involving four-channel EEG recordings under the same stimulus conditions (Figure 6A) while monitoring the animal's pupil size (used to monitor arousal) and locomotor activity (n = 6 mice, 17 recording sessions). The probability of locomotor activity did not differ during unimodal and audiovisual blocks (Kruskal Wallis test: χ^2 = 0.4, p = 0.81; Figure 6B). Similar findings were found for the imaging experiments presented above (Kruskal Wallis test: χ^2 = 0.8, p = 0.66). Moreover, we found that although pupil size was larger during locomotion vs. no locomotion trials (Kruskal Wallis test: χ^2 = 33.0, p = $3.7x10^{-6}$), there was no difference in pupil diameter between the unimodal and bimodal contexts in either stationary or locomotion trials (Figure 6C; Dunn-Sidak post-hoc test, p > 0.99). Consistent with these results, gamma power did not differ between the two contexts in V1 (Friedman test: χ^2 = 1.4, p = 0.50; Figure 6D) and in the vicinity of the primary auditory cortex (Friedman test: χ^2 = 1.1, p = 0.57; Figure 6E).

Discussion

In this study, we investigated how pure tones modulate the representation of orientation and direction by L2/3 neurons in V1. We found that: [1] Sound modulation had a different outcome on neurons whose preferred orientation matched the orientation of the test stimulus (matching neurons) compared to neurons matching orthogonal and opposite orientations/direction, leading to a relative suppression of the activity of orthogonal and opposite neurons; [2] Sound modulation particularly enhanced the response of matching neurons that were highly orientation selective but presented weak responses in the unimodal context; [3] Sound modulation particularly suppressed the responses of neurons matching for the orientation of the test stimulus but preferring the opposite direction; [4] V1 L2/3 neurons recruited by the test stimulus in the audiovisual context were better tuned to the orientation and direction of the stimulus, leading to a more concentrated distribution of the preferred orientations of the responding neurons around the orientation of the test stimulus; [5] Overall, pure tones improved the representation of orientation and direction of the stimulus in the V1 L2/3 neuronal population; [6] The modulation of visually evoked responses in the audiovisual context was not due to an increase in locomotion or arousal, two behavioral factors known to improve the gain of V1 neurons (Niell and Stryker, 2010; Polack et al., 2013; McGinley et al., 2015; Vinck et al., 2015).

Our findings provide several advances on the previous studies that so far documented sound modulation in V1 at the individual neuron level (Iurilli et al., 2012; Ibrahim et al., 2016; Meijer et al., 2017). First, our finding of a significant difference between the sound modulation of matching neurons and orthogonal neurons agrees with the report of Ibrahim and colleagues (2006), which showed a sharpening of the tuning curve of V1 neurons by sound. This finding was debated as Meijer and colleagues were unable to reproduce it (Meijer et al., 2017). We also showed that sound modulation particularly enhanced the visual response of matching neurons highly selective for the orientation of the test stimulus but responding weakly, following the principle of inverse effectiveness

which states that multisensory enhancement is most prominent when individual unimodal inputs are weak (Meredith and Stein, 1983, 1986; Serino et al., 2007; Gleiss and Kayser, 2012). Simultaneously, matching neurons that were less selective and/or more active showed no modulation or were suppressed by sound. Altogether, our findings suggest that the mechanisms responsible for the enhanced response of matching neurons saturates for higher firing rates which could explain why the sharpening of the neurons' tuning curve, in the audiovisual context, was shown by previous reports to be more prominent at low contrasts (Ibrahim et al., 2016). On the other hand, the suppression of the response of orthogonal and opposite neurons was more ubiquitous across response amplitudes and tuning characteristics, which might explain why pioneer studies using whole cell recordings reported that sound systematically hyperpolarized V1 L2/3 neurons (Iurilli et al., 2012). Second, contrarily to previous studies that used stimuli recruiting large neuronal populations in the primary auditory cortex (A1) such as broadband noise, white noise, or frequency-modulated tones (Iurilli et al., 2012; Ibrahim et al., 2016; Meijer et al., 2017), we used pure tones (sine waves) likely to activate distinct subpopulations of A1 neurons (Guo et al., 2012), and therefore possibly distinct bundles of A1 to V1 cortico-cortical projections. We chose two frequencies that activate A1 neuronal populations of different sizes (as 10 kHz is more represented in A1 than 5 kHz (Guo et al., 2012)). Our intent was to test the robustness of tone modulation as well as to detect possible changes in the strength of the effect that would correlate with the size of A1 population activated by the auditory stimulus. We found that simple tones are sufficient to induce sound modulation in V1, yet both tones had a similar impact on the representation of orientation in V1. Further experiments will be necessary to determine if the signal sent by A1 to V1 carries information on the frequency of the auditory stimulus. Finally, our study shows for the first time that sound modulation improves the representation of the stimulus direction in the audiovisual context by suppressing the response of neurons matching to the orientation of the test stimulus but preferring the other direction. This suppression of the null direction likely results from modulation of the intracortical inhibition that was shown to control direction selectivity of V1 L2/3 neurons by suppressing responses to the null direction of motion (Wilson et al., 2018). Sound suppression of the null direction could also explain

why no sound modulation was found for neurons matching to the orientation of the test stimulus when the direction of the test stimulus was not taken into account (Meijer et al., 2017).

Several cellular and network mechanisms have already been proposed to underpin sound modulation in V1. The absence of sound modulation in the layer 4 of V1 (Ibrahim et al., 2016) suggests that sound modulation originates in V1 L2/3. Moreover, direct excitatory projections from the deep layers of the primary auditory cortex, shown to predominantly innervate layer 1 (L1) inhibitory neurons in V1 and to a lesser extent L2/3 excitatory and inhibitory (VIP and somatostatin) neurons (Ibrahim et al., 2016), were proposed to be the anatomical substrate conveying sound modulation to V1. L1 interneurons, which are orientation tuned and project to both V1 L2/3 excitatory and inhibitory neurons (Xu and Callaway, 2009; Ibrahim et al., 2016), were shown to be potentiated by sound (Ibrahim et al., 2016). However, another study proposed that direct A1 to V1 projections activate a subpopulation of infragranular V1 neurons (layer 5), which in turn could mediate the inactivation of supragranular neurons (layer 2/3), likely through local interneurons (Iurilli et al., 2012). Further experiments are required to determine if the sound-induced enhanced response of matching neurons is due to a disinhibition or a direct excitation from A1 neurons.

One of the main goals of this study was to determine if the representation of the visual stimulus by the V1 L2/3 neuronal population was significantly modified in the audiovisual context. The neuronal representation of a sensory stimulus is the information provided by neurons activated during the presentation of the stimulus, that can be potentially extracted by computations performed at later stages of cerebral processing (deCharms and Zador, 2000). We limited the information carried by V1 neurons to their preferred orientation and did not take into account the amplitude of the response for the following reasons. First, V1 neuron's preferred orientation is stable across contrasts (Alitto and Usrey, 2004), and temporal frequency (Moore et al., 2005) and only slightly affected by spatial frequency (Ayzenshtat et al., 2016). On the other hand, the amplitude of a neuron's response depends not only on its orientation tuning but also on its contrast, spatial frequency and temporal frequency tuning, and brain state. Therefore, a given firing rate can be obtained in a same neuron

by different combinations of orientation, contrast and spatial frequency limiting decoding strategies using the orientation tuning curve of the neuron as an estimation of the uncertainty about the stimulus orientation (Ma et al., 2006).

The functional role of cross-modal modulation at such an early stage of cortical sensory processing remains poorly understood. It is possible that interactions between primary sensory cortices are crucial to improve the detection (Odgaard et al., 2004; Lippert et al., 2007; Gleiss and Kayser, 2014) and discrimination thresholds (Vroomen and de Gelder, 2000), as well as to decrease the reaction times (Hershenson, 1962; Posner et al., 1976; Gielen et al., 1983) during object perception tasks in multi-modal contexts. Moreover, we found that the representation of orientation provided by the neuronal population of V1 in the audiovisual context is more concentrated around the orientation of the test stimulus. This more compact representation could potentially reduce the uncertainty on the orientation of the stimulus and therefore improve angular discrimination (i.e. a better capacity for discriminating between stimuli having smaller angular difference). Altogether our results suggest that sound modulation improves the signal to noise of the representation of orientation and direction in the primary visual cortex. These results are compatible with the hypothesis that early stage cross-modal interactions enhance the salience and weight of the sensory signals provided by sensory cortices to higher-order multisensory cortices in order to improve the performance of multisensory integration (Bizley et al., 2016).

References

Alitto HJ, Usrey WM (2004) Influence of contrast on orientation and temporal frequency tuning in ferret primary visual cortex. J Neurophysiol 91:2797-2808.

Atilgan H, Town SM, Wood KC, Jones GP, Maddox RK, Lee AKC, Bizley JK (2018) Integration of Visual Information in Auditory Cortex Promotes Auditory Scene Analysis through Multisensory Binding. Neuron 97:640-655 e644.

Ayzenshtat I, Jackson J, Yuste R (2016) Orientation Tuning Depends on Spatial Frequency in Mouse Visual Cortex. eNeuro 3.

Berens P (2009) CircStat: A MATLAB Toolbox for Circular Statistics. Journal of Statistical Software 31:1-21.

Bizley JK, Jones GP, Town SM (2016) Where are multisensory signals combined for perceptual decision-making? Curr Opin Neurobiol 40:31-37.

Brainard DH (1997) The Psychophysics Toolbox. Spat Vis 10:433-436.

Brosch M, Selezneva E, Scheich H (2005) Nonauditory events of a behavioral procedure activate auditory cortex of highly trained monkeys. J Neurosci 25:6797-6806.

Cappe C, Barone P (2005) Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. Eur J Neurosci 22:2886-2902.

Chen TW, Wardill TJ, Sun Y, Pulver SR, Renninger SL, Baohan A, Schreiter ER, Kerr RA, Orger MB, Jayaraman V, Looger LL, Svoboda K, Kim DS (2013) Ultrasensitive fluorescent proteins for imaging neuronal activity. Nature 499:295-300.

Cronin B, Stevenson IH, Sur M, Kording KP (2010) Hierarchical Bayesian modeling and Markov chain Monte Carlo sampling for tuning-curve analysis. J Neurophysiol 103:591-602.

deCharms RC, Zador A (2000) Neural representation and the cortical code. Annu Rev Neurosci 23:613-647.

Dombeck DA, Khabbaz AN, Collman F, Adelman TL, Tank DW (2007) Imaging large-scale neural activity with cellular resolution in awake, mobile mice. Neuron 56:43-57.

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

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

Ghazanfar AA, Schroeder CE (2006) Is neocortex essentially multisensory? Trends Cogn Sci 10:278-285.

Giard MH, Peronnet F (1999) Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. J Cogn Neurosci 11:473-490.

Gielen SC, Schmidt RA, Van den Heuvel PJ (1983) On the nature of intersensory facilitation of reaction time. Percept Psychophys 34:161-168.

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

Gleiss S, Kayser C (2014) Acoustic noise improves visual perception and modulates occipital oscillatory states. J Cogn Neurosci 26:699-711.

Guo W, Chambers AR, Darrow KN, Hancock KE, Shinn-Cunningham BG, Polley DB (2012) Robustness of cortical topography across fields, laminae, anesthetic states, and neurophysiological signal types. J Neurosci 32:9159-9172.

Hershenson M (1962) Reaction time as a measure of intersensory facilitation. J Exp Psychol 63:289-293.

Ibrahim LA, Mesik L, Ji XY, Fang Q, Li HF, Li YT, Zingg B, Zhang LI, Tao HW (2016) Cross-Modality Sharpening of Visual Cortical Processing through Layer-1-Mediated Inhibition and Disinhibition. Neuron 89:1031-1045.

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

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

Ma WJ, Beck JM, Latham PE, Pouget A (2006) Bayesian inference with probabilistic population codes. Nat Neurosci 9:1432-1438.

Markram H, Toledo-Rodriguez M, Wang Y, Gupta A, Silberberg G, Wu C (2004) Interneurons of the neocortical inhibitory system. Nat Rev Neurosci 5:793-807.

McGinley MJ, Vinck M, Reimer J, Batista-Brito R, Zagha E, Cadwell CR, Tolias AS, Cardin JA, McCormick DA (2015) Waking State: Rapid Variations Modulate Neural and Behavioral Responses. Neuron 87:1143-1161.

Meijer GT, Montijn JS, Pennartz CMA, Lansink CS (2017) Audiovisual Modulation in Mouse Primary Visual Cortex Depends on Cross-Modal Stimulus Configuration and Congruency. J Neurosci 37:8783-8796.

Meredith MA, Stein BE (1983) Interactions among converging sensory inputs in the superior colliculus. Science 221:389-391.

Meredith MA, Stein BE (1986) Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. J Neurophysiol 56:640-662.

Molholm S, Sehatpour P, Mehta AD, Shpaner M, Gomez-Ramirez M, Ortigue S, Dyke JP, Schwartz TH, Foxe JJ (2006) Audio-visual multisensory integration in superior parietal lobule revealed by human intracranial recordings. J Neurophysiol 96:721-729.

Moore BDt, Alitto HJ, Usrey WM (2005) Orientation tuning, but not direction selectivity, is invariant to temporal frequency in primary visual cortex. J Neurophysiol 94:1336-1345.

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

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

Odgaard EC, Arieh Y, Marks LE (2004) Brighter noise: sensory enhancement of perceived loudness by concurrent visual stimulation. Cogn Affect Behav Neurosci 4:127-132.

Petro LS, Paton AT, Muckli L (2017) Contextual modulation of primary visual cortex by auditory signals. Philos Trans R Soc Lond B Biol Sci 372.

Polack PO, Friedman J, Golshani P (2013) Cellular mechanisms of brain state-dependent gain modulation in visual cortex. Nat Neurosci 16:1331-1339.

Posner MI, Nissen MJ, Klein RM (1976) Visual dominance: an information-processing account of its origins and significance. Psychol Rev 83:157-171.

Rockland KS, Ojima H (2003) Multisensory convergence in calcarine visual areas in macaque monkey. Int J Psychophysiol 50:19-26.

Serino A, Farne A, Rinaldesi ML, Haggard P, Ladavas E (2007) Can vision of the body ameliorate impaired somatosensory function? Neuropsychologia 45:1101-1107.

Song YH, Kim JH, Jeong HW, Choi I, Jeong D, Kim K, Lee SH (2017) A Neural Circuit for Auditory Dominance over Visual Perception. Neuron 93:940-954 e946.

Vinck M, Batista-Brito R, Knoblich U, Cardin JA (2015) Arousal and locomotion make distinct contributions to cortical activity patterns and visual encoding. Neuron 86:740-754.

Vinje WE, Gallant JL (2000) Sparse coding and decorrelation in primary visual cortex during natural vision. Science 287:1273-1276.

Vroomen J, de Gelder B (2000) Sound enhances visual perception: cross-modal effects of auditory organization on vision. J Exp Psychol Hum Percept Perform 26:1583-1590.

Wilson DE, Scholl B, Fitzpatrick D (2018) Differential tuning of excitation and inhibition shapes direction selectivity in ferret visual cortex. Nature 560:97-101.

Wohrer A, Humphries MD, Machens CK (2013) Population-wide distributions of neural activity during perceptual decision-making. Prog Neurobiol 103:156-193.

Xu X, Callaway EM (2009) Laminar specificity of functional input to distinct types of inhibitory cortical neurons. J Neurosci 29:70-85.

Legends

Figure 1: Responses of the V1 L2/3 neuronal population to drifting gratings in visual only and audiovisual contexts. (**A**) Experimental setup. (**B**) Stimuli of the unimodal block and audiovisual blocks. (**C**) Variation of the fractional fluorescence (dF/F) of the GCaMP6f signal of the same neuron during the presentation of the test stimuli 45° (top panel) and 135° (bottom panel) in the unimodal (blue traces) and multimodal contexts (with 10 kHz tone: green traces; with 5 kHz tone: red traces). Overlap of 15 consecutive presentations of the same stimulus (colored traces). Median response (black traces). Colored arrows: mean across the duration of the stimulus (grey boxes, duration: 3 seconds) of the median response. (**D**) Tuning curves of the neuron presented in

(C) obtained with a resampling-based Bayesian method. Overlapped traces around the polar plot are the single trial responses of the neuron to the presentation of each of the twelve orientations. Preferred orientation: 123°; Orientation Selectivity Index: 0.92; Direction Selectivity Index: 0.62. (**E**) Plot of the response amplitude to the presentation of the 45° test stimulus during 'Visual only' blocks as a function of the angular distance between the orientation of the test stimulus and the neuron's preferred orientation (δ orientation) for all excitatory neurons in the database (n = 2,376 neurons). Inset: distribution of the response amplitudes. Gaussian fit: R^2 = 0.99. One-tailed 0.05 alpha threshold: 0.403 (activity threshold). (**F**) Plot of the response amplitude of active neurons to the presentation of a 135° drifting grating as a function of the δ orientation in the unimodal context. (**G**) (Left panel) Same representation as in (F) when a 10 kHz tone is presented simultaneously with the 135° drifting grating. (Right panel) Same representation as in (F) when a 5 kHz tone is presented simultaneously with the 135° drifting grating.

Figure 2: Orientation of the population of V1 L2/3 neurons active in the visual only and audiovisual contexts. (**A**) Polar histograms representing the distribution (as a probability density function) of the preferred orientations of the active neurons in the visual only (blue, left) and audiovisual contexts (green: tone 10 kHz, center; red: tone 5 kHz, right). Inset: circular means of the three distributions (scale: 2-fold the polar plot's scale). (**B**) Plot of the mean resultant length and mean direction of the 1,000 circular mean vectors obtained by bootstrapping the preferred orientations of the active neurons during the visual only (blue dots) and audiovisual (green dots: 10 kHz tone, red dots: 5 kHz tone) contexts. Ovals indicate the probability density functions at the level 0.05 computed from a two-dimension gaussian fit. (**C**) Mean and confidence interval of the difference in direction (top), mean resultant length (center), and concentration (bottom) between the unimodal and audiovisual contexts. * Indicates a significant difference.

Figure 3: Sound modulation as a function of V1 L2/3 neurons' orientation tuning. (**A**) Box plots (median and 25 – 75 percentile box) and individual data points of the modulation of active neurons matching the orientation of the test stimulus (left) or matching orthogonal orientations and

opposite directions to that of the test stimulus (middle and right, respectively). Colored boxes indicate medians that significantly differ from 0 (two-sided sign test with $p<0.05$). *: Kruskal–Wallis test ($p = 2.0x10^{-5}$) followed by Dunn-Sidak multiple comparisons ($p_{1-2} = 0.001$, $p_{1-3} = 0.0001$, $p_{2-3} = 0.50$, with 1: matching, 2: orthogonal, and 3: opposite). (**B**) Similar representation as in (A) when the activity threshold was brought to 2.0 z-score. *: Kruskal–Wallis test ($p = 1.0x10^{-4}$) followed by Dunn-Sidak multiple comparisons ($p_{1-2} = 0.001$, $p_{1-3} = 0.009$, $p_{2-3} = 0.81$). (**C**) Sound modulation of matching, orthogonal, and opposite neurons inactive (response < 0.4 z, yellow box), moderately active (0.4 z < response < 2.0 z, orange box), and highly active (response > 2.0 z, brown box) during the presentation of the test stimulus in the unimodal context. Whiskers extend to the most extreme data not considered outliers * Indicates a significant difference between groups (Kruskal-Wallis test followed by a Dunn-Sidak multiple comparison, $p<0.05$). • Indicates a median significantly different from zero (two-sided sign test, $p <0.05$).

Figure 4: Sound modulation as a function of V1 L2/3 neurons' orientation selectivity. (**A**) Plot of the sound modulation index of the matching neurons as a function of their orientation selectivity indexes (OSI). Blue line: linear fit. Shaded blue area: confidence bounds of the linear fit. Linear fit: r^2: 0.01; F-statistic vs. constant model: 4.8; p-value = 0.03 (**B**) Same representation as in (A) for neurons matching orientations orthogonal to the orientation of the test stimulus. Linear fit: r^2: $1x10^{-5}$; F-statistic vs. constant model: 0.004; p-value = 0.95. (**C**) Same representation as in (A) for neurons matching the direction opposite to that of the test stimulus. Linear fit: r^2: $1x10^{-4}$; F-statistic vs. constant model: 0.02; p-value = 0.90. (**D**) Sound modulation of matching neurons with low (OSI < 0.25), moderate (0.25< OSI < 0.75), and high orientation selectivity (OSI > 0.75). * Indicates a significant difference between groups (Kruskal-Wallis test followed by a Dunn-Sidak multiple comparison, $p<0.05$). • Indicates a median significantly different from zero (two-sided sign test, $p <0.05$).

Figure 5: Sound modulation as a function of V1 L2/3 neurons' direction selectivity. (**A**) Plot of the sound modulation index as a function of the neuron direction selectivity indexes (DSI) for

neurons matching the test stimulus orientation. Linear fit: R^2: $3x10^{-4}$; F-statistic vs. constant model: 0.1; p-value = 0.73. (**B**) Same representation as in (A) for neurons matching orientations orthogonal to that of the test stimulus. Linear fit: R^2: $1x10^{-4}$; F-statistic vs. constant model: 0.3; p-value = 0.58. (**C**) Same representation as in (A) for neurons matching the direction opposite to that of the test stimulus. Linear fit: R^2: 0.04; F-statistic vs. constant model: 6.0; p-value = 0.015. (**D**) Sound modulation of opposite neurons with low (ODI < 0.25), and moderate direction selectivity (0.25< ODI < 0.75). High direction selective neurons (ODI > 0.75) were not included because of a too low count. * Indicates a significant difference between groups (Kruskal-Wallis test followed by a Dunn-Sidak multiple comparison, p<0.05). ● Indicates a median significantly different from zero (two-sided sign test, p <0.05).

Figure 6: Brain states during visual presentation in the visual and audiovisual contexts. (**A**) Example of a four-channel EEG recording (top traces) of the motor cortex (Motor Cx), primary auditory cortex (A1), primary visual cortex ipsilateral (V1 ipsi.) and primary visual cortex contralateral (V1 contra.) made in a head-fixed mouse placed on a spherical treadmill. Recording were performed while presenting blocks of visual and audiovisual cues (bottom traces) and monitoring pupil size and locomotion. (**B**) Box plots of the probability of locomotion during visual only, auditory only and audiovisual cues. Kruskal–Wallis test: p = 0.81. (n = 17 recording sessions in 6 animals). (**C**) Plots of the pupil area (mean ± s.e.m (black) and individual data points (orange); n = 17 recording sessions in 6 animals) as a function of the stimulus modality (visual only, auditory only or audiovisual) and the presence or absence of locomotion. Kruskal-Wallis test: χ^2 = 33.0, p = 3.7×10^{-6}, Dunn-Sidak post-hoc test: p = 0.99 between stimulus conditions in locomotion and no locomotion trials. (**D**) Plots of the power of the gamma band measured in the contralateral V1 to the visual stimulus as a function of the stimulus modality and the presence or absence of locomotion. Friedman test: Chi-squared locomotion χ^2 = 20.6, p = $5.7x10^{-6}$; Chi-squared stimulus modality χ^2 = 1.4, p = 0.50. (**E**) Same representation as in (D) for the primary auditory cortex. Friedman test: Chi-squared locomotion χ^2 = 17.8, p = $2.5x10^{-5}$; Chi-squared stimulus modality χ^2 = 1.14, p = 0.57.

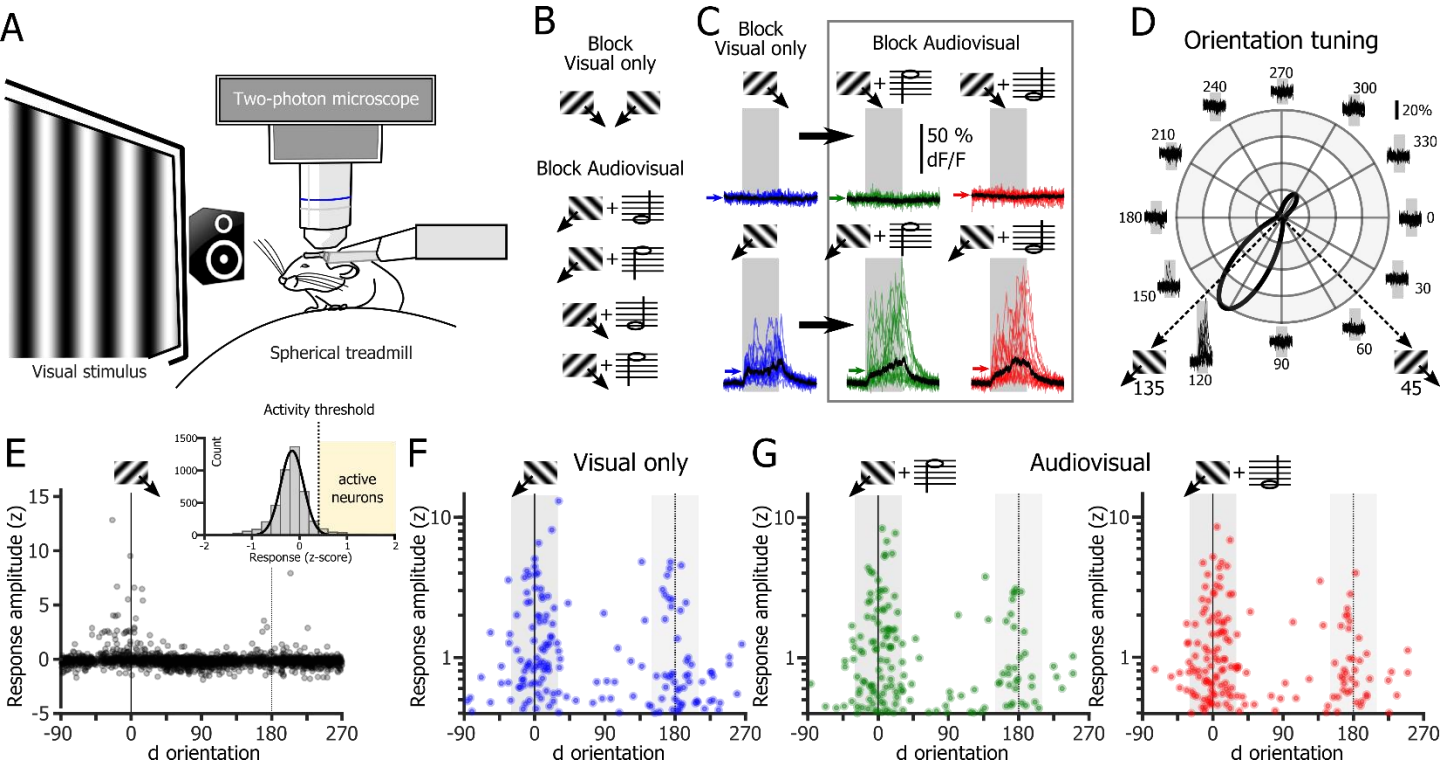
A
Two-photon microscope
Visual stimulus
Spherical treadmill
B
Block
Visual only
Block Audiovisual
C
Block
Visual only
Block Audiovisual
50 %
dF/F
D
Orientation tuning
270
240
300
20%
210
330
180
0
150
30
90
60
135
120
45
E
Activity threshold
Count
active
neurons
Response (z-score)
Response amplitude (z)
d orientation
F
Visual only
Response amplitude (z)
d orientation
G
Audiovisual
Response amplitude (z)
d orientation

Figure 1

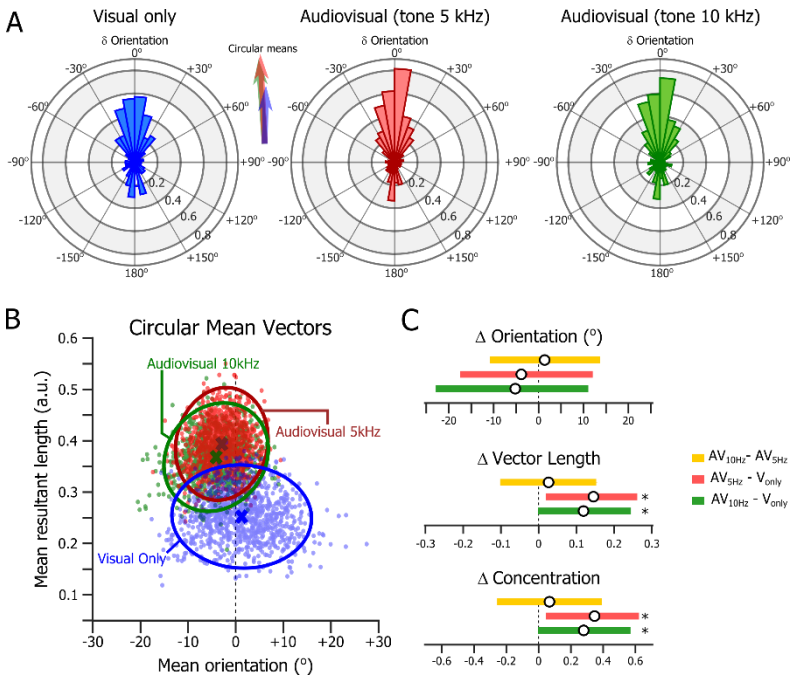
A
Visual only
Audiovisual (tone 5 kHz)
Audiovisual (tone 10 kHz)
δ Orientation
Circular means
B
Circular Mean Vectors
Audiovisual 10kHz
Audiovisual 5kHz
Visual Only
Mean resultant length (a.u.)
Mean orientation (°)
C
Δ Orientation (°)
Δ Vector Length
Δ Concentration
AV10Hz - AV5Hz
AV5Hz - Vonly
AV10Hz - Vonly

Figure 2

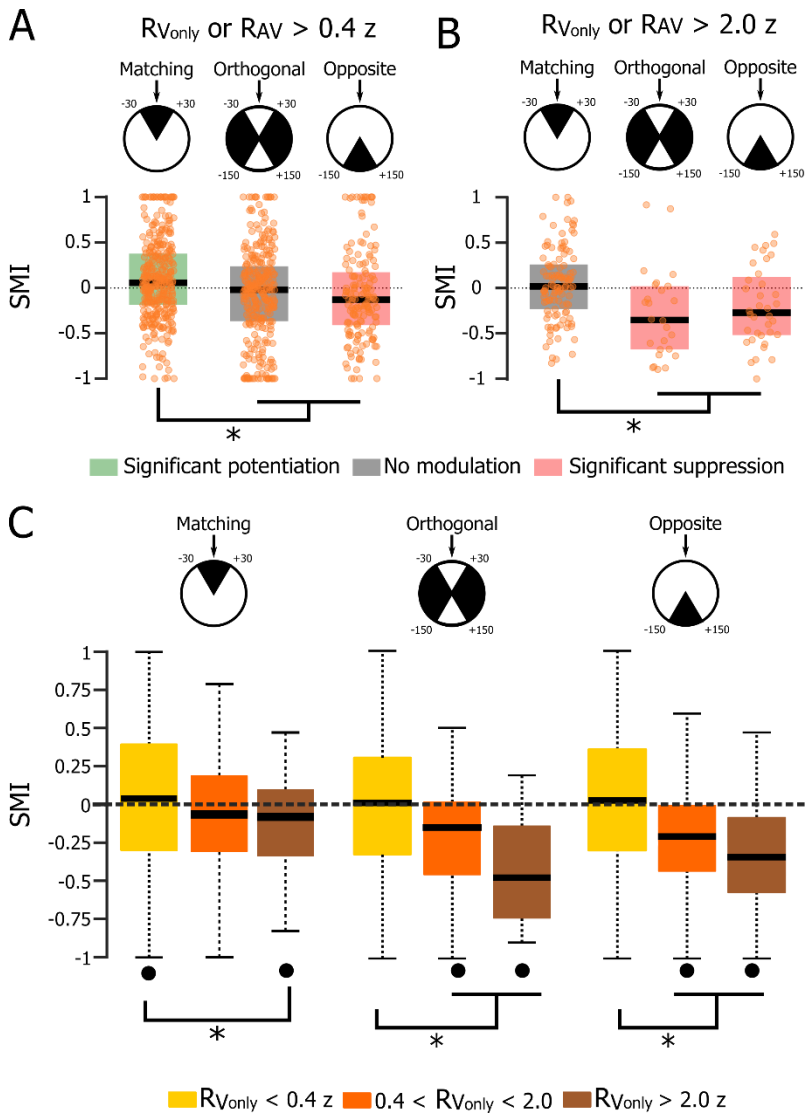
A
RVonly or RAV > 0.4 z
Matching
Orthogonal
Opposite
-30
+30
-150
+150
SMI
1
0.5
0
-0.5
-1
*
B
RVonly or RAV > 2.0 z
Matching
Orthogonal
Opposite
-30
+30
-150
+150
SMI
1
0.5
0
-0.5
-1
*
Significant potentiation
No modulation
Significant suppression
C
Matching
Orthogonal
Opposite
-30
+30
-150
+150
SMI
1
0.75
0.5
0.25
0
-0.25
-0.5
-0.75
-1
*
*
*
RVonly < 0.4 z
0.4 < RVonly < 2.0
RVonly > 2.0 z

Figure 3

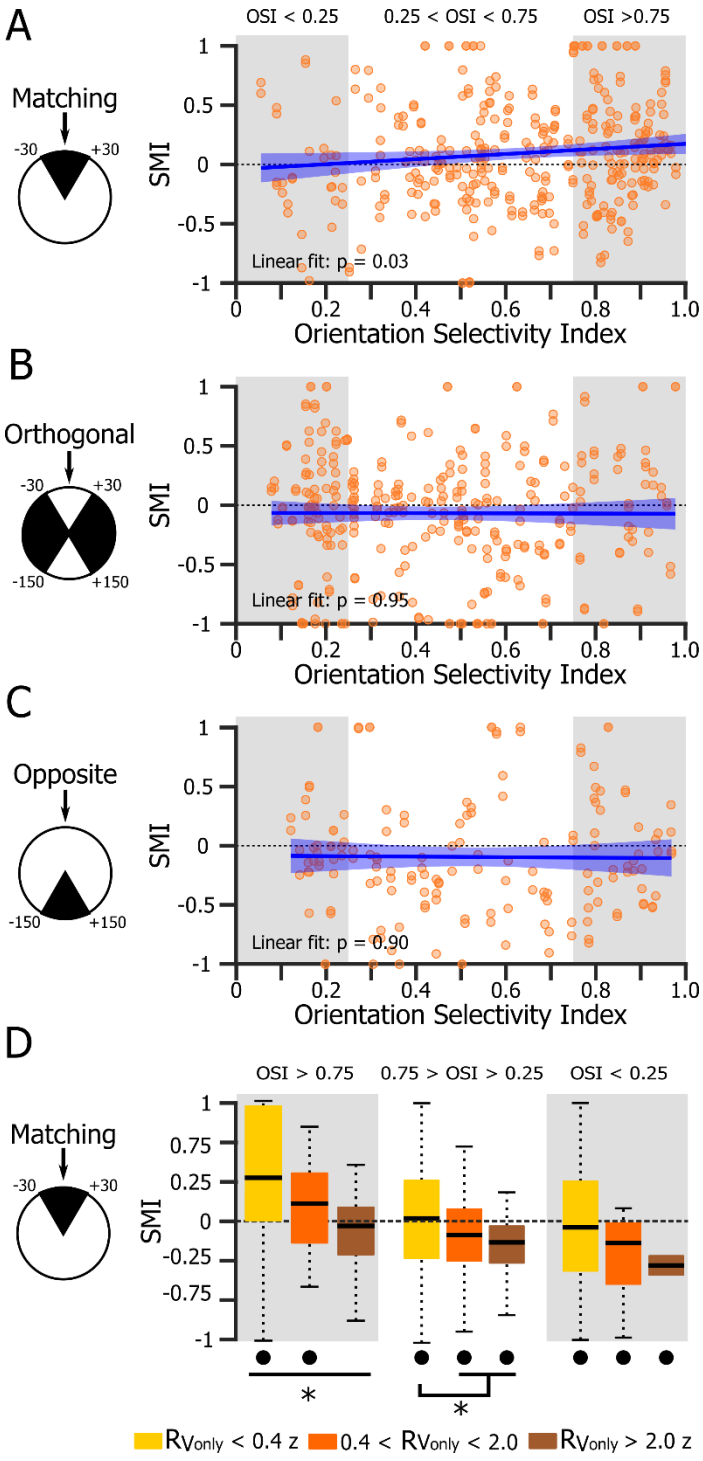
A
Matching
-30
+30
OSI < 0.25
0.25 < OSI < 0.75
OSI >0.75
SMI
Linear fit: p = 0.03
Orientation Selectivity Index
B
Orthogonal
-30
+30
-150
+150
Linear fit: p = 0.95
Orientation Selectivity Index
C
Opposite
-150
+150
Linear fit: p = 0.90
Orientation Selectivity Index
D
Matching
-30
+30
OSI > 0.75
0.75 > OSI > 0.25
OSI < 0.25
SMI
RVonly < 0.4 z
0.4 < RVonly < 2.0
RVonly > 2.0 z

Figure 4

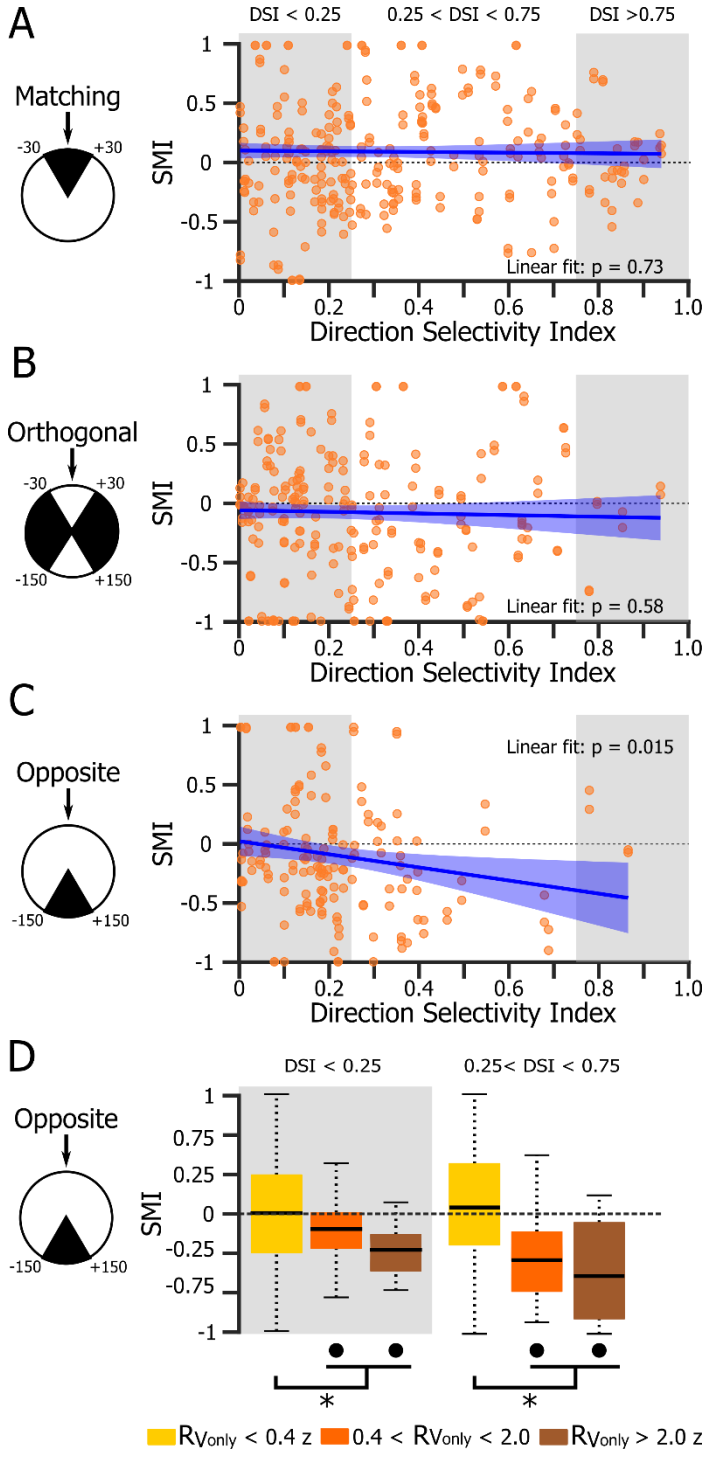
A
Matching
-30
+30
DSI < 0.25
0.25 < DSI < 0.75
DSI >0.75
SMI
Linear fit: p = 0.73
Direction Selectivity Index
B
Orthogonal
-30
+30
-150
+150
SMI
Linear fit: p = 0.58
Direction Selectivity Index
C
Opposite
-150
+150
SMI
Linear fit: p = 0.015
Direction Selectivity Index
D
Opposite
-150
+150
DSI < 0.25
0.25< DSI < 0.75
SMI
*
*
RVonly < 0.4 z
0.4 < RVonly < 2.0
RVonly > 2.0 z

Figure 5

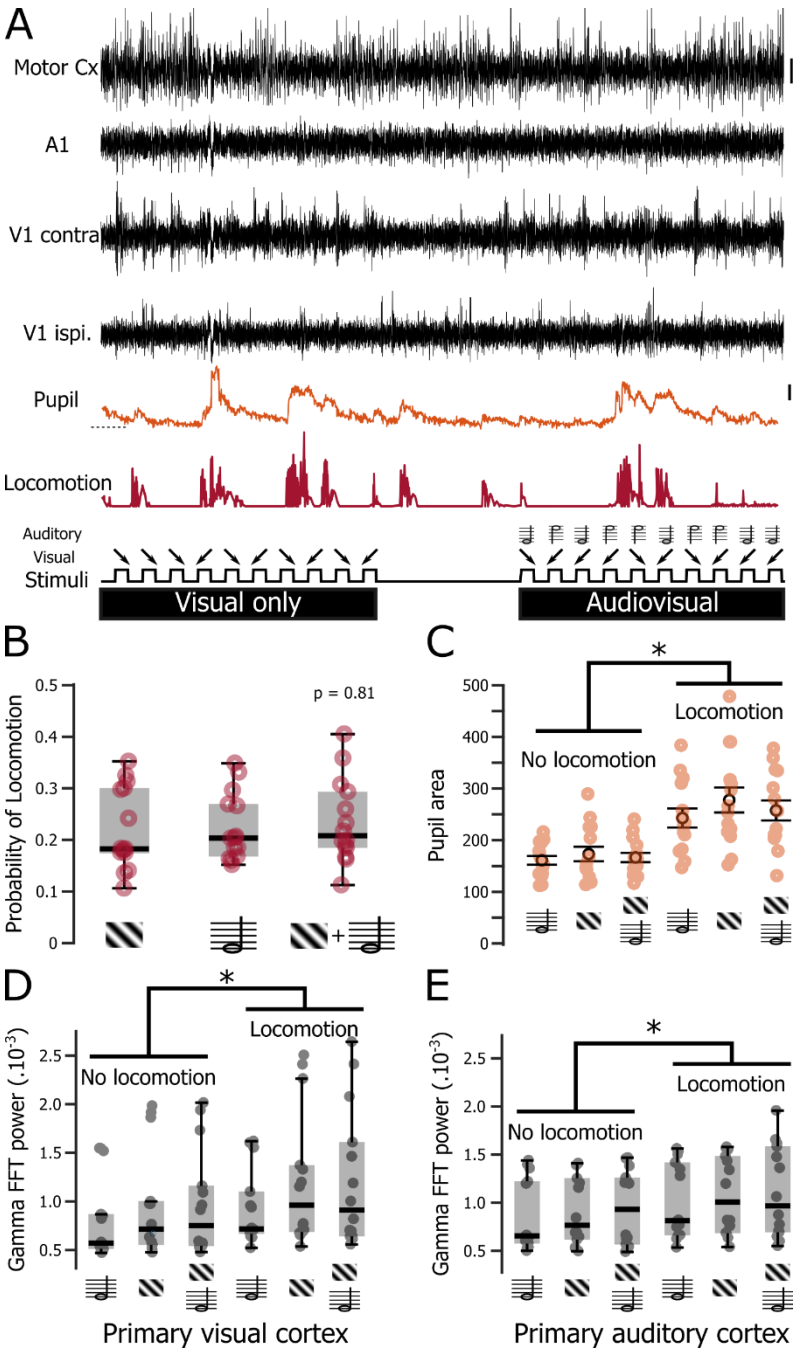
A
Motor Cx
A1
V1 contra
V1 ispi.
Pupil
Locomotion
Auditory
Visual
Stimuli
Visual only
Audiovisual
B
Probability of Locomotion
p = 0.81
C
Pupil area
No locomotion
Locomotion
D
Gamma FFT power (.10-3)
No locomotion
Locomotion
Primary visual cortex
E
Gamma FFT power (.10-3)
No locomotion
Locomotion
Primary auditory cortex

Figure 6