Investigating Neuronal Mechanisms of Multisensory Integration in V1 and AL of Mice: Enhancement of Multisensory Neurons and Noise Decorrelation are Associated with Higher Performance on Detection Task

Daphnee Chabal

11200898

Supervision:

Guido Meijer, MSc

Dr Carien Lansink

Table of Contents

Abstract

Multisensory integration is a neuronal process which combines sensory information from different modalities, which improves perception of faint multimodal events. Studies using passive stimuli presentation paradigms have suggested that multisensory integration involves neurons in both primary sensory cortices and areas receiving inputs from different sensory cortices. This study however combined an audiovisual detection task with two-photon calcium imaging and eye-tracking to verify the occurrence of multisensory integration via its expected behavioral effects. The role of arousal and noise structures in stimuli processing and perception were also taken into account. We additionally proposed area AL as a candidate side of multisensory integration. In trained mice, we found that multisensory integration improved perception of faint audiovisual stimuli, through enhancement of multisensory neurons and suppression of visual neurons in V1 but not in AL. Arousal did not impact stimuli perception, nor did neuronal response reliability. However, we found that noise correlation in both V1 and AL improved stimuli perception. Implications of these findings and limitations are discussed.

Word count: 164

Introduction

In the environment, reliable information about meaningful events is necessary to enable timely detection, localization, and appropriate reactions to such events. Sensory information about a single event originating from several modalities enhances the reliability of perceiving such an event, compared to unimodal sensory information. For example, seeing someone talking enhances the reliability of understanding correctly the meaning that person is trying to convey. In an evolutionary context, an enhanced ability to detect potential threatening events enables survival (Stein & Stanford, 2008). This effect of multimodal cues on perception is commonly studied using stimuli detection tasks, with varying cue intensity, that is, varying cue reliability (Ernst & Banks, 2002; Gleiss & Kayser, 2012; Gu et al., 2008; Lippert et al., 2007). These studies have observed that presenting multimodal stimuli rather than unimodal ones lowers the perceptual threshold of an animal and enables detection of faint multimodal stimuli. Indeed, the lower the visual contrast or auditory loudness, the lesser the chances of unimodal stimuli perception.

Perceiving multimodal information as originating from a unified event is an automatic and effortless process, attributed to the neuronal phenomenon called multisensory integration (Stein & Stanford, 2008). Multisensory integration occurs when sensory information from different modalities is perceived almost-simultaneously as originating from one location, and is combined on a neuronal level. The merging of multimodal information during multisensory integration increases chances of detecting events of interest. This process is not necessary when unimodal cues are individually distinct enough to be perceived reliably. However, when unimodal cues are weak, noisy, and unreliable, perception of an event only occurs when these unimodal inputs converge with inputs from other modalities (Fetsch et al., 2013). For example, multisensory integration is more likely to occur and thus to have a greater contribution on one's perception in a noisy and poorly lit room, when understanding an interlocutor is not possible if only hearing them but is enabled by the sight of their facial expressions as they talk. Multisensory integration

thus obeys the principle of inverse effectiveness, which states that multisensory enhancement of stimuli detection is greater when stimulus components are weaker (Stein & Stanford, 2008).

Multisensory integration was initially thought to occur in higher order association areas that receive inputs from two or more primary sensory cortices (Stein & Stanford, 2008). These areas seemed to contain multisensory neurons that have a greater response to multimodal stimuli compared to unimodal ones. For example, Olcese and colleagues (2013) found that a brain region in mice called RL, located between the tactile (i.e. S1) and the visual (i.e. V1) primary sensory cortices, contained multisensory neurons that were more active when tactile and visual stimuli were presented simultaneously rather than individually. These multisensory neurons were more active when multimodal stimuli were individually weak, that is, at below perceptual threshold intensity. Similarly, the superior colliculus in cats receives visual, auditory, and somatosensory inputs that converge on multisensory neurons (Anastasio et al., 2000; Meredith & Stein, 1986; Stein & Stanford, 2008, Wallace & Stein, 1997). In mice, visual and auditory multisensory neurons could therefore be located in an area that receives projections from both the primary auditory and the visual primary sensory cortices. Little is known about area AL in the visual cortex of mice (Anderman et al., 2011), but it shares bidirectional connections with A1 and V1 (Laramée et al., 2011; Wang & Burkhalter, 2007; Wang et al., 2011; Wang et al., 2012). We thus presently propose area AL as a candidate multisensory integration site.

Recent evidence revealed that stimuli from a modality could trigger activity in another modality's primary sensory cortex. Specifically, multimodal modulation of layer II/III neurons at the action potential level was observed in auditory and visual primary sensory cortices (Ibrahim et al., 2016; Iurilli et al., 2012; Kayser et al., 2010; Kayser et al., 2008). Audiovisual multisensory integration in V1 of mice is particularly interesting to investigate presently, because V1 activity has been shown to be strongly modulated by changes in stimuli intensity (Glickfeld et al., 2013) and V1 is a region that is surgically easily accessible in mice. Additionally, V1 shares numerous and direct connections with the primary

auditory cortex (Abrahim et al., 2016; Cappe et al., 2009), thus indicating that multisensory neurons in primary cortices receives crossmodal influences. Similarly to higher order areas, multisensory integration in primary sensory cortices occurs through an increased response of multisensory neurons to the sensory input from that cortex's main modality (e.g. visual input in V1) when it is paired with input from a non-dominant modality (e.g. auditory input in V1) compared to when it is presented alone. Multisensory neurons also show less activity in response to unimodal stimuli compared to other neighboring neurons (Kayer et al., 2010). This mechanism is called multisensory enhancement (Anastasio et al., 2000) and is particularly effective when stimuli intensity is weak in both modalities (Ibrahim et al., 2016; Kayser et al., 2008). However, multimodal information processing in sensory cortices also involves multisensory suppression of neuronal response in a modality's primary sensory cortex upon crossmodal or multimodal stimuli presentation (Ibrahim et al., 2016; Iurilli et al., 2012; Kayser et al., 2010).

The role of multisensory neurons and the exact mechanisms of multisensory integration in terms of enhancement or suppression of primary sensory cortices activity thus remain to be determined. Indeed, it remains unclear which neurons in V1 show multisensory enhancement, which neurons show multisensory suppression and which stimuli intensity activates these mechanisms. Additionally, these studies (Ibrahim et al., 2016; Iurilli et al., 2012; Kayser et al., 2010; Kayser et al., 2008; Olcese et al., 2013) recorded activity in passive mice, which presents a two-fold problem. First, animals had no explicit need to combine information neuronally to convert it into a motor output, thus these studies generate no evidence that multisensory integration occurred, but only that crossmodal stimuli add sensory information processed in unimodal sensory areas. Second, the perceptual threshold of animals were not estimated. It is therefore not certain that unimodal cues were individually weak enough (i.e. below unimodal perceptual threshold) to require the mechanisms of multisensory integration.

Performance on detection tasks and underlying neuronal activity depend on how much an animal pays attention to relevant information about its surrounding (Bennet et al., 2013; Ecker et al., 2014; Reimer et

al., 2014; Vinck et al., 2015). When studying the neural correlates and behavioral effects of multisensory integration, it is thus important to take arousal into consideration. The arousal system consists of several neurotransmitter systems that collectively regulate the cholinergic pathway, which has been found to determine pupil dilatation and constriction (Bradley et al., 2008; Coull et al., 2004). Arousal is therefore positively correlated with pupillometry, which is measurable using non-invasive infrared cameras (Bradshaw, 1967; Zoccolan et al., 2010). In mice, optimal task performance occurs at intermediate levels of arousal (Yerkes & Dodson, 1908; McGinley et al., 2015a); low levels of arousal indicating non-engagement with the task, and high levels of arousal being associated with distractingly threatening situations. The effect of arousal on detection task performance and neuronal multisensory integration mechanisms are still unknown. Moreover, most research that studied the relationship between arousal and behavior categorized arousal binarily, thus overlooking the Yerkes-Dodson-Law of optimal task performance at intermediate levels of arousal.

To understand multisensory integration fully, one must consider neuronal activity that is not accounted for during multimodal and unimodal stimuli processing, that is, activity that is not stimuli-induced, and which may influence whether a stimulus is perceived or not. Neuronal response reliability and noise correlations have been suggested as two noise structures that influence and are affected by arousal states, activeness, and animal performance (McGinley et al., 2015b) as well as stimuli perception (Montijn et al., 2014).

Neuronal response reliability is inversely proportional to the trial-by-trial variability of a neuron's response to the same type of stimuli (e.g. visual stimuli with below perceptual threshold stimuli intensity). On average across neurons, high variability suggests a type of stimulus was encoded through inconsistent responses of neurons in sensory cortices, which could negatively impact an animal's perception and reaction to the stimuli. In 2015, McGinley and colleagues found that, in the auditory cortex of mice, neuronal response reliability, and animal task performance, were optimal at intermediate levels of arousal

(i.e. the optimal arousal state). It is yet to be determined if neuronal response reliability at intermediate levels of arousal improve multimodal or unimodal stimuli perception.

Noise correlations refer to the correlations between pairs of neurons in terms of their responses' trial-to-trial variability. Neuronal noise correlations have been shown to either be beneficial or detrimental to stimuli encoding (Averbeck et al., 2006; Faisal et al., 2008; Franke et al., 2016). Several studies (Bennett et al., 2013; McGinley et al., 2015b; Reimer et al., 2014; Schölvinck et al., 2015; Vinck et al., 2015) showed that there are noise decorrelations during more active and aroused states of mice. However, noise correlations and arousal have not yet been studied in relations to stimuli detection, or lack thereof.

This present study investigated the neuronal correlates of multisensory integration by comparing audiovisual stimuli-induced modulation of neuronal activity of mice. Neuronal activity were recorded using two-photon calcium imaging, which enabled the simultaneous measurement of activity of neuron populations in layer II/III of V1 and AL, over several sessions. To insure neural activity recorded reflects the mechanisms of multisensory integration, mice were trained on an audiovisual detection task with varying stimuli intensity, which they performed during neuronal activity recordings. Training increased confidence that successful performance on the detection task became equivalent to stimuli perception. The behavioral effects of multimodal compare to unimodal stimuli on task performance, as well as the animals' arousal were then analyzed.

We first hypothesized that our detection task would successfully demonstrate the behavioral effect of underlying multisensory integration, as we expected performance of mice on the detection task to be enhanced by audiovisual stimuli, compared to unimodal stimuli, especially when animals' internal state was at intermediate levels of arousal, using a ternary approach to arousal states, even when stimuli were individually faint. We hypothesized that multimodal stimuli, when successfully detected, would activate multisensory neurons in both V1 and AL. Multisensory neurons, given their acute sensitivity to multimodal stimuli, were expected to be identified as being among the least active neurons during visual

stimuli presentation. Multisensory neurons were expected to be more active during audiovisual stimuli compared to visual stimuli, especially when stimuli were faint. Other neurons in these regions were expected to respond preferentially to visual stimuli, thus suppressing their response to audiovisual stimuli. We did not expect such enhancement/suppression mechanisms when the animals did not respond to the stimuli, as it indicated that they did not perceive the stimuli and thus, that multisensory integration did not occur.

We then hypothesized that neuronal response reliability and noise correlations, in relation with arousal, influence stimuli perception, and thus task performance, to investigate whether such noise structures are mechanisms involved in multisensory integration. Less response variability may indicate coherent neuronal computations between trials, such as stimuli-induced activity, which would enable stimuli processing and thus perception. We thus hypothesized that trials where mice perceived and responded to a visual or audiovisual stimulus show higher neuronal response reliability on average, compared to trials where mice did not detect nor respond to stimuli. Moreover, it was hypothesized that neuronal response reliability would lead to higher task performance when the animals had intermediate levels of arousal. Similarly, to gather further insights on noise correlations, we hypothesized that trials where mice responded to and thus perceived visual or audiovisual stimuli showed more stimuli-induced activity and thus less noise correlations compared to trials where mice did not perceive nor respond to stimuli. Noise decorrelation was additionally hypothesized to lead to higher task performance when animals had intermediate, and thus optimal, arousal level.

Method

Animals

The experiment was conducted in accordance with Dutch national guidelines on the conduct of animal experiments, following the Dier Experimenten Commissie (DEC) protocol number 295. Six mice (C57B1/6) were obtained from Harlan, and were housed socially in groups of 2 to 4. Mice were chosen as

they can be subject to genetic modification, they have a smooth cortex, and a visual network spanning over a compact area. The mice lived on a reversed 12-hour light/dark cycle (lights on: 8pm-8am) so that training and experiments in the lab were performed during their active hours. The mice were 21 days old at the beginning of training and 144 days old during the last session.

Surgical Procedures

Headbar implantation. The mice were implanted with a titanium headbar on the skull over the left visual cortex. The headbar enabled head fixation to the experimental set up during training and imaging of neuronal activity of awake mice in V1. The headbar contained an 8 mm circular window that was centered 3 mm posterior and 2.5 mm lateral from Bregma (Paxinos & Franklin, 2004). This window was an opening within which later surgeries could be performed as well as the imaging and recording of cortical neuronal activity. The mice were administered 0.1 mg-kg of buprenorphine analgesia via subcutaneous injection, 20-30 minutes prior to surgery, to exempt the mice from pain for 24 hours after injection. The mice were placed on a heating mat with variable warmth to help maintain the body temperature of mice at 37°C via a temperature controller connected to an anal probe. The mice were under 1.5-2.7% isoflurane anesthesia via a vaporizer during surgery. During surgery, the skin under the desired location for the circular window of the headbar was cut with disinfected surgical scissors. Xylocaine was applied to anesthetized the wound. The headbar was cemented to the skull with C&B Superbond (Sun Medical, Japan). A protective see-through cover of cyanoacrylate glue (Locktite 401, Henkel, CT, USA) was applied over the within the circular window, to prevent infections until the next surgical procedure. Training started after a recovery period of seven days.

Craniotomy. This surgery was two-fold, as 3 mm in diameter of the skull was removed within the headbar's window to allow for subsequent imaging and recording of neuronal activity, and the viral construct AAV1.Syn.GCaMP6m/f.WPRE.SV40 (Penn Vector Core) was injected in two areas, V1 and AL. Areas were identified with the help of an IOS image (Intrinsic Optical Signal) containing blood

vessels visible to the naked eye, obtained prior to the surgery. The calcium indicator GCaMP6f was chosen for its fast rise and decay time and its low calcium accumulation. Using the same analgesic and anesthetic conditions described above, the cyanoacrylate glue and a 3 mm round area of the skull were drilled away with an electric Desk 300 IN dentist drill. Next, 150 nl of viral construct, combined with the color pigment ALEXA FLUOR, was injected in each area at a 500-700 µm depth and at a rate of 23 nl/sec (to prevent back-spill and to help spread the virus), in three pulses of 50 nl, using the NANOJECT III (Drummond Scientific Company) and 8 µm pipettes. There were five minutes waiting time between area injections to allow the virus to spread in the tissue. The craniotomy was closed with a bottom round coverglass of 3 mm diameter to prevent skull regrowth and apply pressure on the brain (Goldey et al., 2014). A top coverglass of 5 mm diameter was glue to the skull using cyanoacrylate glue identified above.

Intrinsic Optical Signal Imaging

IOS operates under the principle that active cortical populations of neurons refracted more light than inactive ones. IOS was performed twice; once to guide the site of the viral injections during the craniotomy surgery and once to insure the microscope was placed above the correct cortical region of interest during calcium imaging recordings. IOS images were obtained, prior to both procedures, using ultra-red (810 nm wavelength) light, projected onto the skull, and a CCD camera (1 Hz sampling rate) to record the amount of reflected light. The area AL, V1, and LM were visualized when mice anesthetized under 1% isoflurane were presented with bidirectionally drifting grating visual stimuli of 8 different orientations, each for 1 second with a 17 second inter-stimulus intertrial. AL and V1 specifically were later identified using VDAQ software (Optical Imaging Ltd) and anatomical knowledge on the location of both areas in the visual cortex of mice (Wang & Bulkhalter, 2007). Images obtained also contained the blood vessels present in the areas, to further guide later surgical and imaging procedures.

Behavioral paradigm

The purpose of the study was to observes effects and mechanisms of multisensory integration on the behavioral and neuronal levels. Mice were therefore trained to an audiovisual detection task, which presented both unimodal and multimodal stimuli to the animals. Performance on the task gave a behavioral indication that multisensory integration occurred neuronally.

The mice were shown randomly an audio, visual, or audiovisual stimulus for 1 second, upon which they were trained to lick to get a reward (within a 1 second window). This started an inter-trial with random variation of duration from 3 to 5 seconds, followed by a period of 1 to 2 seconds during which no licks should be detected to start the next stimulus. The stimuli were at full intensity (100% visual contrast or 70 dB) during training. However, to make the stimuli harder for mice to perceive and respond to, as well as to get an estimation of the mice's relative perceptual thresholds (i.e. perceive stimuli at specific intensity level 50% of the trials), visual contrast and loudness of the audio stimuli were changed and progressively reduced according to performance, during imaging, using an adaptive staircase procedure (Psychtoolbox). The staircase procedure allowed to detect the animals' internal states via a non-linear psychometric function (Buss et al., 2009) where the contrast or loudness of the stimuli were viewed according to their probability of being detected. The whole task was coded in a custom-made MATLAB script, which communicated with the different components of the experimental set-up via field programmable gate arrays.

Behavioral Training

Apparatus. Mice were head-fixed on a custom-made experimental construction. The set up were placed in soundproof cages with no light. A pump pushed the milk reward (18 ml of powdered milk mixed with 60 ml room temperature water) out of a syringe, connected to a spout positioned at licking distance from the mice, at a rate of 1 ml per pulse. Licks were detected using a custom-made laser ray with infrared beam. The visual stimuli were presented on a 15 inch TFT screen positioned 16 cm from the

right eye of mice, 45˚ from the midline of the head. Auditory stimuli were amplified (TA 1630, Sony Amplifier, Japan) and presented with a tweeter (Neo CD 3.0, Audaphon, Germany) positioned at a 22 cm distance from mice under the left side of the screen.

Stimuli. Visual stimuli consisted of square-wave bi-directionally drifting gratings with a temporal speed of 1 Hz and a spatial frequency of 0.05 cpd, presented at 3 orientations (0°, 120° or 240°). The stimuli were confined within a grey cosined-tampered circular window with a diameter of 60 retinal degrees to eliminate edge effects. Visual stimuli were at full contrast during training. Auditory stimuli at maximum intensity consisted of 15 Hz tones at 70 dB during training and 88 dB during imaging to compensate for the 64 dB background noise of the scanner, as measured by a sound meter (Phonic PAA3 Audio Analyzer).

Procedure. Mice were habituated to the experimental setup and to head fixation for 4 days (~3 minutes/session). The mice were then put on a water deprivation schedule during week days to motivate them to earn their daily ration of liquid (min. 1 ml) during training (1 session/week day). The first sessions consisted of 20 visual or audio full-intensity blocks of 40 trials, in random order. Training sessions were terminated when mice did not respond to 5-10 stimuli in a row, indicating satiation or tiredness. Reward was administrated automatically for the first session, to condition mice to the stimuli-reward paradigm. Probe trials were introduced from the second session onwards, during which false-alarm licks were recorded. Stimuli length were progressively reduced from 5 to 1 second within the first three sessions. Performance on the task was measured using a D-prime value (Andermann et al., 2010), where *norminv*() is the inverse of the cumulative function (see Equation 1).

$$\mathrm{d}' = \mathrm{norminv}(\text{Hit rate}) - \mathrm{norminv}(\text{False alarm rate})$$

Equation 1

When mice reached the criteria for proficient performance, set at a D-prime of 1.5 per session for 3 consecutive sessions, the block design was replaced by a random-trial version of the detection task, where audio and visual stimuli were presented in random order, at full intensity. After successful (D-prime < 1.5) learning of this version of the task, the mice were trained on the final staircase procedure, which introduced audiovisual trials and varying stimuli intensity in accordance with performance. Imaging started when mice became proficient in the staircase detection task, and when their visual cortex expressed measurable amount of fluorescence (~3 weeks after craniotomy). The first and last 10 trials of each session were removed from data analyses, as they contain the behavioral confounds of habituation to and disengagement with the task.

The perceptual threshold of mice was determined by the stimulus intensity (visual contrast and auditory loudness) in trials where mice displayed a correct licking response only 50% of the time, indicating that they did not perceive the stimuli 50% of the time. All trials where stimuli intensity was below the animals' perceptual thresholds was termed sub-threshold trials, and all trials where stimuli was above the animals' perceptual thresholds but not at full intensity – contrast or maximum loudness – were termed supra-threshold trials. Audiovisual threshold was calculated both in visual contrast units (%) and loudness units (dB SPL) for trials where the cross-modal stimuli was slightly weaker than the other unimodal stimulus.

Two-Photon Calcium Imaging

Background. Neuronal activity causes fast increases/decreases in intra-cellular calcium concentrations, corresponding to the depolarization of a cell during an action potential. Two-photon calcium imaging is enabled when a viral construct, combined with a calcium indicator protein (i.e. GCaMP6f), is introduced into an organism's genes. The genetically encoded protein is then expressed in the soma of all neurons, and not glia, and becomes fluorescent when intracellular calcium binds to it and changes its configuration, with the fluorescent intensity corresponding to the amount of action potentials

(Chen et al., 2013). Indeed, the protein is a fluorophore that absorbs high power laser light. Wavelength being inversely proportional to energy, the laser needs to have a long wavelength (i.e. infrared laser) but a high concentration of photons to avoid photodamage of the cortical tissue. Therefore, it takes two photons, delivered in the form of brief high-power pulses, arriving simultaneously at the same fluorophore site, to gather the combined energy necessary to activate the fluorophore, thus emitting light that can be measured.

Apparatus. Two-photon calcium imaging was achieved using a Leica SP5 resonant laser scanning microscope, with a Spectra Physics Mai Tai High Performance Mode Locked Ti:Sapphire laser. The wavelength of the laser was set at 910 nm and the resonant mirrors were set to a sampling frequency of 27 Hz. Th focal point of the lasers targeted layer II/III of the visual cortex (V1 or AL) at a depth of 140-200 µm. Refracting fluorescent light was measured with a photo-multiplier tube at a 525 nm wavelength detection sensitivity. The resulting image plane was of 365x365 µm (512x512 pixels), using the 25x Leica objective. Laser and imaging settings were set using the LAS AF software, which coordinated frame capturing, laser pulse, and stimuli synchronization.

Procedure. Brain movements, inevitable despite head-fixation, resulted in micro movements on the x-y directions of the frames, which were corrected using a single step discrete Fourier realignment procedure (Guizar-Sicairos et al., 2008). All neurons expressing fluorescence were selected using a semi-automatic MATLAB interface. Cell bodies were selected manually, but the script performed neuropil subtraction by averaging the fluorescence signal of one soma from an annulus of 2-5 µm around it, and subtracting from that soma's fluorescence signal. The neuropil signal was multiplied by 0.7 prior, to avoid subtraction (Chen et al., 2013). Activity of each neuron was quantified using the $\Delta F/F_0$ metric (see Equation 2), where ΔF is the relative increase in fluorescence, F_0 is mean baseline fluorescence (calculated over the lower half of all values during a 30 second sliding window before time i) and F^i is the activity of a neuron at imaging frame i. The $\Delta F/F_0$ value for each trial was the highest value between

stimulus onset and starting of the licking response of mice (0.5 seconds). A neuron was considered responsive, using the inclusive criteria of significantly more activation during stimuli presentation compared to baseline activity, using a paired sample t-test. Statistical analyses were performed with MATLAB 'Statistics and Machine Learning Toolbox'.

$$\Delta F/_{F_0} = \frac{F^i - F_0}{F_0}$$

Equation 2

Neuronal response enhancement or suppression during audiovisual trials compared to visual trials were quantified by computing a response change index (RCI) of the $\Delta F/F_0$ trace (see Equation 3). In the RCI equation, F_{AV} and F_V stand for the $\Delta F/F_0$ trace of all responsive neurons across mice in response to audiovisual and visual trials respectively. This metric ranged from -1 to 1 with negative values indicating suppression of neuronal response and positive values indicating response enhancement. The response change index is a normalized method for quantifying multisensory cue integration (Meijer et al., under review).

$$RCI = \frac{F_{AV} - F_V}{F_{AV} + F_V}$$

Equation 3

Eye-tracking

The pupils of the mice from the eye ipsilateral to the stimulus presentation were recorded using a near-infrared camera (CV-150 IR, JAI, Denmark), using the light emitted from the laser microscope, at an imaging frequency of 25 Hz. Images were processed with MATLAB.

Analyses. A series of custom-made algorithm (Grzelkowski & Meijer, unpublished), based on Li and colleagues' (2005) starburst algorithm, allowed for the detection and quantification of the pupil diameter from raw images. Manually, image frames were cropped to have the eye boundaries at its center, and the

pupil outline were connected with points, to decrease the processing power required for the algorithm (i.e. less pixels). A Gaussian filter algorithm was used to smooth the new images. The fast radial symmetry transform was applied to find the center of the pupil, where the center is the highest obtained value (FRST; Loy and Zelinsky, 2003). The sobel filter was applied to find the edge of the pupil, where rays were projected from the center of the pupil in all directions with equal radial distance. The maximal value of the sobel filtered image were evaluated, in relation to the ray-points closest to the center. Only the edge points (along the rays-axis) that were at equal distance (± 2 SD) from the center were kept for the ellipse fitting algorithm (Brown, 2015), where an ellipse fitted through the remaining points outlined the pupil. The area of the ellipse (i.e. the area of the pupil), its X-Y displacement, as well as the deviation of the ellipse from an expected round shape (i.e. absolute difference between the vertical and horizontal axes of the ellipse) were manually checked for each frame of each session videos. Pupil information was deleted from frames where the algorithm did not succeed in producing a correctly fitting ellipse (e.g. during eye-blink). Pupil diameters were averaged across the 0.5 s window before stimulus onset, to avoid the effect of visual stimulus induced pupil contraction. Pupil diameters were split in three even groups according to their size, namely, the highest third, the intermediate third, and the smallest third. These groups were used as a measure of three states of arousal (i.e. low, medium, high) of the mice (Aston-Jones & Cohen, 2005; Yerkes & Dodson, 1908).

Results

Audiovisual stimuli improve performance regardless of arousal levels

We trained 6 mice on an audiovisual detection task to determine whether multimodal stimuli improved stimuli detection and thus task performance compared to unimodal stimuli, even when stimuli were below the animals' unimodal perceptual threshold. An animal's perceptual threshold for each type of stimuli was computed as the level of stimuli intensity which led to stimuli perception on 50% of the trials. Task performance was measured by the hit rate - ratio of detected trials to total trial number - calculated

across animals. Hit rates were consistently higher across stimuli types during supra-threshold compared to sub-threshold trials, as revealed by Bonferroni-corrected (p = .007) one-tailed sample t-tests (audio: $t(18) = -4.79$, $p < .007$, visual: $t(18) = -2.98$, $p < .007$, audio-visual: $t(18) = -3.79$, $p < .007$). This finding was expected as louder sounds and brighter visuals are more salient and thus easier to detect. In support of the hypothesis, at both stimuli intensity levels, hit rates were higher for audiovisual compared to visual (supra-threshold: $t(18) = -0.12$, $p < .007$) and audio trials (supra-threshold: $t(18) = -0.16$, $p < .007$; sub-threshold: $t(18) = -0.12$, $p < .007$). At sub-threshold stimuli intensity however, there was no significant difference in hit rate between visual and audiovisual trials ($t(18) = -0.04$, p = .01). Thus, multimodal stimuli improved stimuli detection compared to unimodal stimuli, when unimodal stimuli were easily detectable but also when stimuli were presented at a stimuli intensity below unimodal perceptual threshold (Figure 1). These findings gave an indication that multisensory integration occurred neuronally during the task.

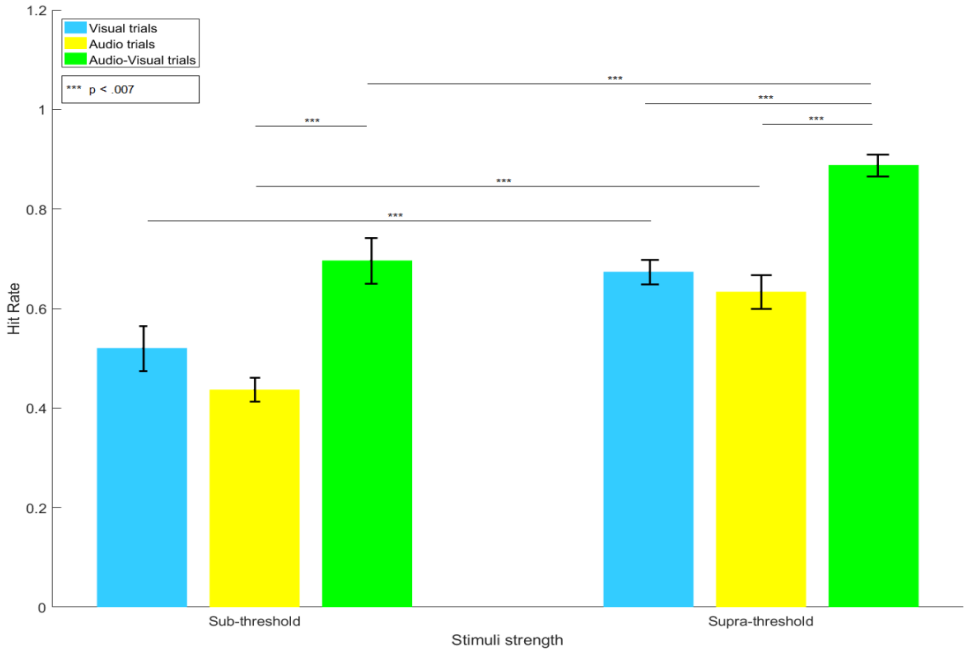

Figure 1.
Multisensory stimuli result in increased perceptual performance. Bar chart of the average hit rate across animals, on visual (blue), audio (yellow), and audiovisual (green) trials with sub-threshold (left) and supra-threshold (right) stimuli intensity.

To investigate the role of arousal on detection of stimuli of different types and intensities, we further divided compared trials based on the corresponding pupil size that animals had directly before the beginning of each trial. We found no effect of arousal on performance during trials with sub-threshold stimuli intensity, $F(2,80) = .81$, $p = .45$ (see Figure 2.A). However for supra-threshold trials, we found a significant effect of pupil sizes of hit rates ($F(2,80) = 4.86$, $p = .01$), with a post-hoc Bonferroni test showing that hit rates were significantly lower when pupil sizes were small compared to when pupil sizes were big, at $p < .01$ (see Figure 2.B). These analyses maintained the significant effect of stimuli type on hit rate for both sub ($F(2, 80) = 22.45$, $p < .001$) and supra-threshold ($F(2, 80) = 24.34$, $p < .001$). Thus, high rather than intermediate levels of arousal facilitated stimuli detection, contrarily to our hypothesis and in contradiction with the Yerkes-and-Dodson law of optimal task performance. Moreover, arousal had an impact when stimuli were easily detectable but not when stimuli intensity was below the unimodal perceptual threshold.

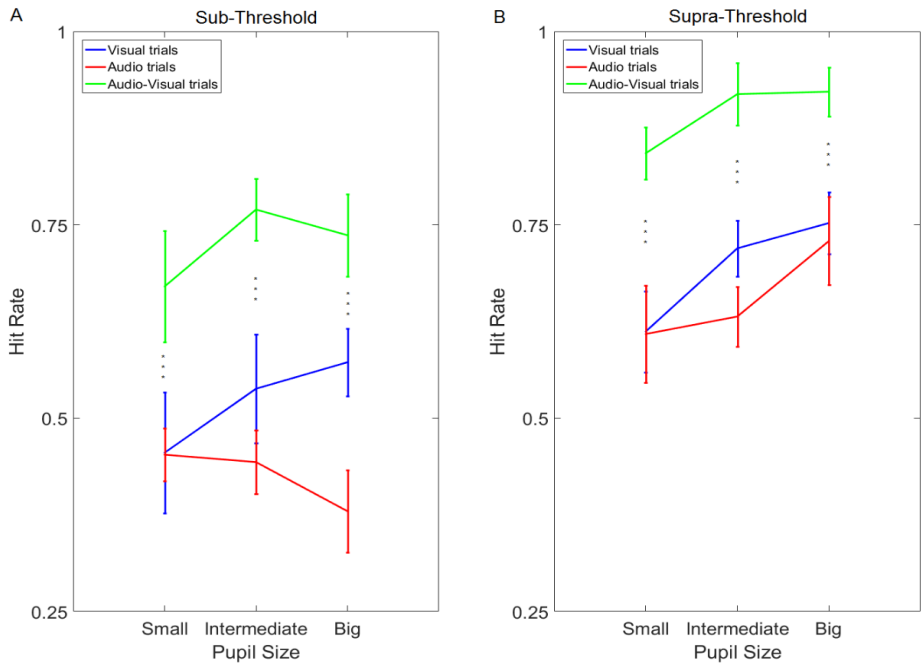

Figure 2.
Performance on detection task is improved by stimuli type (with higher rates with multimodal rather than unimodal trials) but not influenced by arousal level. (A). Plot of average hit rate over all trials with sub-threshold stimuli intensity, when trials were split into 3 quantiles of pupil size and based on their stimulus type. (B) Same as 2.A for trials with supra-threshold stimuli intensity. Bars indicate SEM. *** p < .001

To further investigate the effect of multisensory stimuli on stimuli detection and thus behavior, the perceptual threshold of animals was compared between multisensory and unisensory trials. The visual threshold of animals was significantly lowered during audiovisual trials (M = .475, in % visual contrast units, SEM = .21) compared to visual-only trials (M = 1.88, in % visual contrast units, SEM = .92), as revealed by a one-tailed independent-sample t- test, $t(10) = 3.64$, $p < 0.01$ (see Figure 3). In line with previous observations, presenting a sensory event composed of stimuli from several modalities lowers an animal's perceptual threshold, allowing it to detect and respond to very faint stimuli.

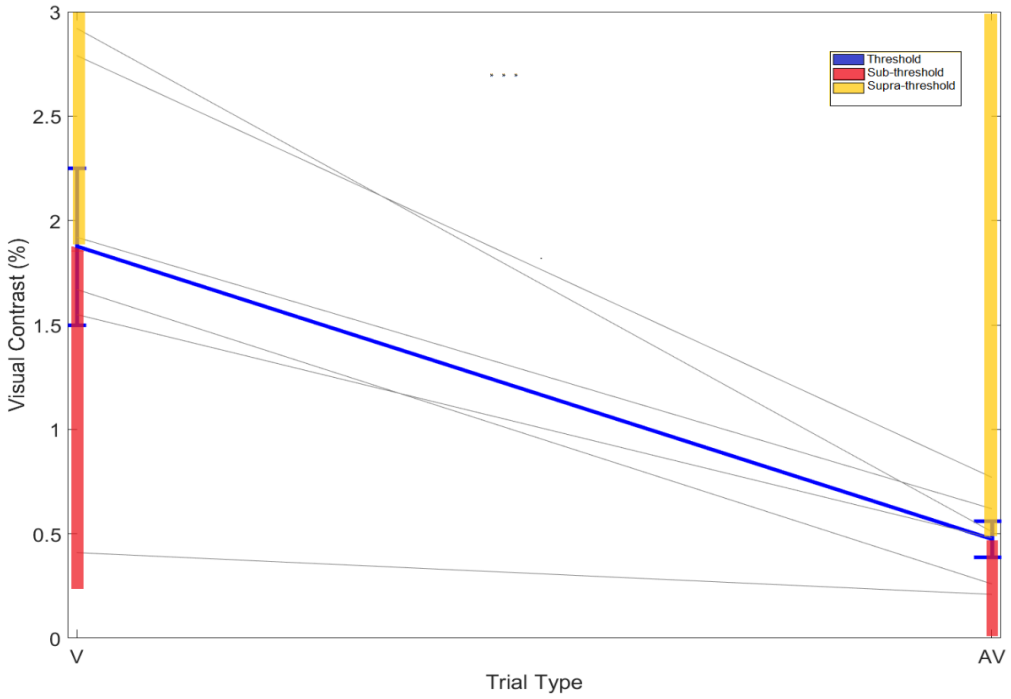

Figure 3.
Multisensory stimuli decrease perceptual threshold of visual stimuli. Plot of average visual contrast threshold (50% detection rate) of all animals (blue) and per animal (grey), with sub-threshold and supra-threshold intensity range, for both visual and audiovisual trials, where audiovisual threshold were calculated only for trials with higher intensity of visual stimulus compared to audio stimulus. Bars indicate SEM. *** p < .01

Audiovisual stimuli affect neuronal response in V1 and AL differently

We performed two-photon imaging of the V1 and AL areas of the mice while they performed on the detection task to study the neuronal correlates and mechanisms of multisensory integration. Given the behavioral effects observed during audiovisual trials compared to unimodal ones, multisensory integration was assumed to occur when animals successfully detected and thus responded to audiovisual stimuli and when these stimuli were at below unimodal perceptual threshold intensity, that is, hardly detectable when presented individually. The activity of responsive neurons was analyzed using the average $\Delta F/F_0$ value across trials of the same type, per animal. The response change index (RCI) was then calculated for responsive neurons in both V1 and AL, to observe the differential effect of multimodal versus unimodal stimuli on neuronal activity in the visual cortex. Trials were split according to stimulus intensity, at either sub or supra-threshold, and according to behavioral response to the trials, HIT trials representing trials where mice have seen and responded to stimuli and MISS trials representing trials where mice did not see nor respond to the stimuli within a 1 second window from stimulus onset.

We found that the mean response change index was significantly bigger than 0 across AL neurons during MISS trials ($t(246) = 4.46$, $p = .00001$; see Figure 4.D) but not during HIT trials ($p = .34$; see Figure 4.C) and was not different from 0 across V1 neurons during HIT ($p = .09$; see Figure 4.A) nor MISS trials ($p = .39$; see Figure 4.B). These results were consistent when splitting responses based on the stimuli intensity of trials and thus indicated that overall, neuronal activity in V1 and AL did not differ for visual and audiovisual stimuli.

Multisensory integration was however hypothesized to occur via the activation of a specific sub-population of neurons called multisensory neurons, which respond preferentially to audiovisual stimuli but respond weakly to visual stimuli. V1 and AL neurons were thus split in three equal groups according to the intensity of their response to visual stimuli compared to baseline (i.e. weak, medium, strong responses to visual-only trials). Multisensory neurons were expected to be ones that responded the

weakest to unimodal (i.e. visual) stimuli while visual neurons were expected to respond the strongest to visual stimuli. Non-parametric tests were used for analysis because RCIs of neurons in each of the three groups were not normally distributed and their variance was not homogenous.

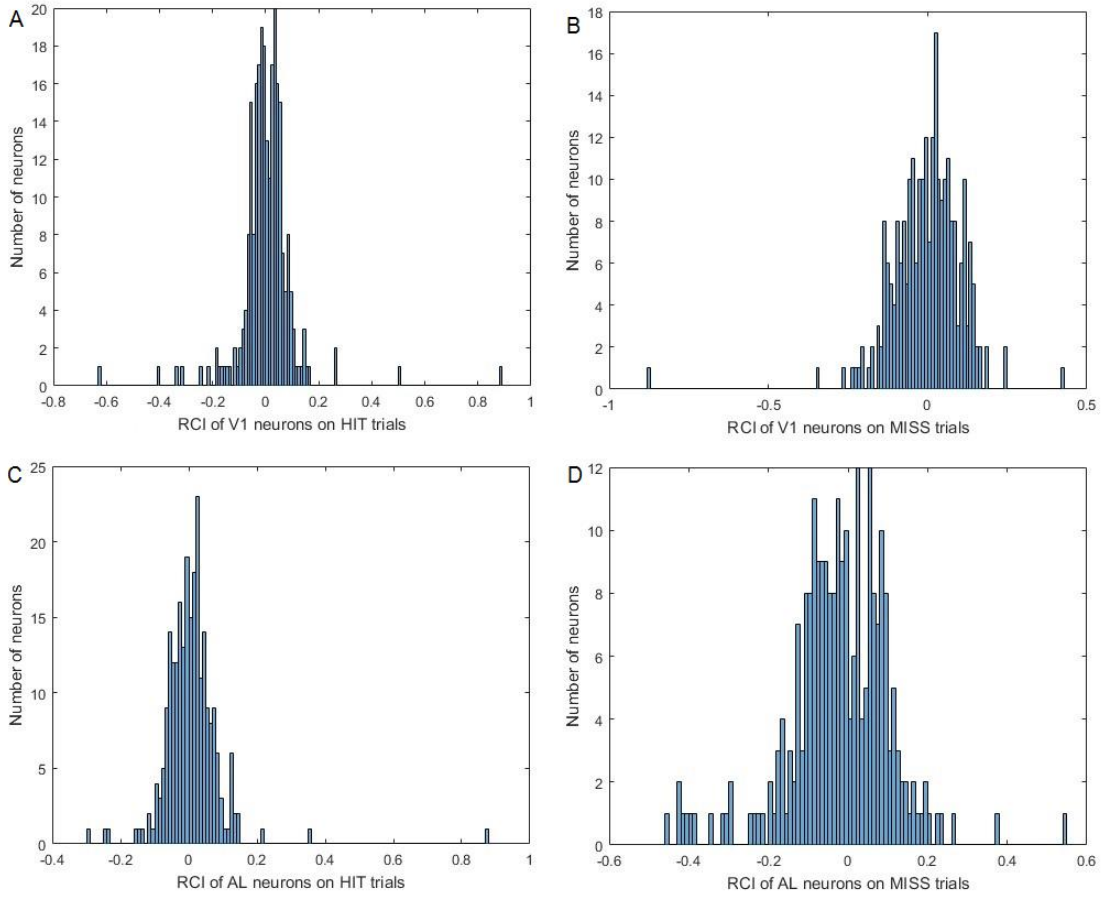

Figure 4.
Neurons in V1 and AL have a Response Change Index (RCI) nearing zero. (A) Histogram of the distribution of V1 neurons based on their RCI value during HIT trials (i.e. stimuli detected and responded to) at around threshold stimulus intensity. (B) Histogram of the distribution of V1 neurons based on their RCI value during MISS trials (i.e. stimulus not detected). (C) Histogram of the distribution of AL neurons based on their RCI value during HIT trials. (D) Histogram of the distribution of AL neurons based on their RCI value during MISS trials.

Consistently with our hypothesis, we found that in V1, the RCI of neurons (n = 260) during sub-threshold HIT trials was significantly different between neuron groups (i.e. sorted depending on their response to visual-only trials), as revealed by a Kruskal Wallis Test, $H(2) = 12.92$, $p = 0.002$. Indeed, multimodal neurons showed an enhanced response during audiovisual trials (RCI: M = .06, SEM = .02) while visual neurons showed a suppressed response during audiovisual trials, (RCI: M = -.008, SEM =

.02, see Figure 5.A). A post-hoc Bonferroni test yielded that RCI of multisensory neurons was significantly higher than the index of visual neurons, at $p < .01$. As expected, this significant enhancement-suppression dynamic was not present for neuronal response to sub-threshold MISS trials in V1, $H(2) = 5.64$, $p = 0.06$, with RCI of multisensory ($M = .01$, $SEM = .02$) and visual neurons ($M = -.04$, $SEM = .02$) being not significantly different at a $p < .01$ level (see Figure 5.B). The RCIs of neurons in V1 were not significantly different between groups of neurons during supra-threshold HIT trials ($H(2) = 5.46$, $p = 0.07$; see Figure 5.C) and MISS trials ($H(2) = 5.72$, $p = 0.06$; see Figure 5.D). In sum, detection of faint stimuli during the task was accompanied by underlying activation of multisensory neurons in V1 and suppression of visual neurons. This enhancement/suppression dynamic was not observed in conditions where multisensory integration is not thought to take effect, that is, when stimuli are not detected and/or when multimodal stimuli are perceptible individually. We therefore concluded that such enhancement/suppression of multisensory neuronal activity is a mechanism of multisensory integration.

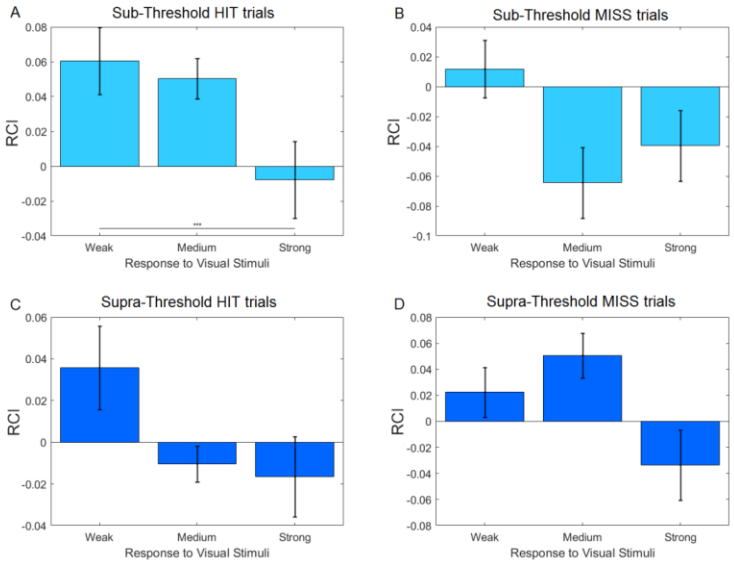

Figure 5.
RCI is significantly bigger for multisensory neurons in V1 compared to visual neurons, during sub-threshold HIT trials. (A) Response Change Index of V1 neurons during HIT trials, based on their response (weak, medium, or strong) to visual trials at sub-threshold stimuli intensity. (B) Response Change Index of V1 neurons during MISS trials, based on their response to visual trials at sub-threshold stimuli intensity. (C) Response Change Index of V1 neurons during HIT trials, based on their response to visual trials at supra-threshold stimuli intensity. (D) Response Change Index of V1 neurons during MISS trials, based on their response to visual trials at supra-threshold stimuli intensity. Bars indicate SEM. *** p < .01

In AL, the RCIs of neurons (N = 247) were not significantly different between groups of neurons during supra-threshold MISS trials ($H(2) = 0.09$, $p = 0.96$; see Figure 6.D) nor sub-threshold MISS trials ($H(2) = 4.30$, $p = 0.12$; see Figure 6.B), as expected. Contrarily to hypothesis however, sub-populations of neurons in AL did not respond differently during sub-threshold HIT trials($H(2) = 2.82$, $p = 0.24$; see Figure 6.A). Instead, sub-populations of neurons responded significantly differently during supra-threshold audiovisual HIT trials ($H(2) = 10.01$, $p = 0.007$). A post-hoc Bonferroni test yielded significantly higher RCI for multisensory neurons (M = .02, SEM = .02) compared to the RCI of visual neurons (M = -.03, SEM = .01) at $p < .01$ (see Figure 6.C). Thus, the expected mechanisms of multisensory integration did occur in AL but not under conditions considered favorable for multisensory integration, that is, when detected stimuli were faint. Instead, enhancement of multisensory neurons' and suppression of visual neurons' activity in AL was only observed during trials where detected stimuli were at a perceptible intensity.

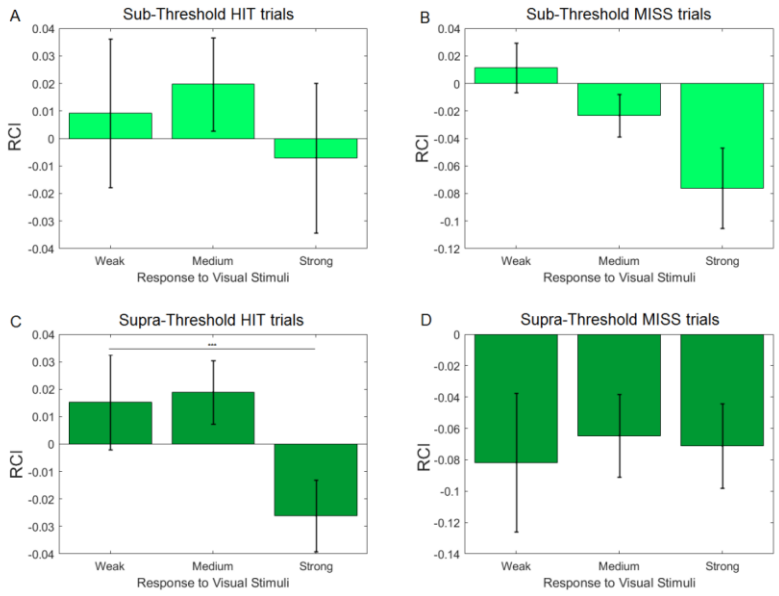

Figure 6.
RCI is significantly bigger for multisensory neurons compared to visual neurons in AL, during sub-threshold HIT trials. (A) Response Change Index of AL neurons during HIT trials, based on their response (weak, medium, or strong) to visual trials at sub-threshold stimuli intensity. (B) Response Change Index of AL neurons during MISS trials, based on their response to visual trials at sub-threshold stimuli intensity. (C) Response Change Index of AL neurons during HIT trials, based on their response to visual trials at supra-threshold stimuli intensity. (D) Response Change Index of AL neurons during MISS trials, based on their response to visual trials at supra-threshold stimuli intensity. Bars indicate SEM. *** p < .01

Variability in neuronal response does not affect performance

Neuronal response reliability over trials with the same stimuli type and intensity was hypothesized to be positively associated with task performance. Indeed, higher reliability is equivalent to lower variability in stimuli-induced neuronal activity, possibly indicating more efficient sensory processing, enabling stimuli perception and thus performance. Intermediary and thus optimal levels of arousals were also hypothesized to enhance the effect of neuronal response variability on task performance. Visual and audiovisual trials were split based on pupil diameter (see above) and the response of mouse (i.e. HIT trials where the animal perceived and responded to stimuli or MISS trials where the animal did not perceive the stimuli).

To avoid splitting data in too many groups, thus decreasing the amount of trials and neurons per group, we sampled all trials with around threshold stimulus intensity (i.e. trials with ±25% stimulus intensity). In each group, the coefficient of variation $\Delta F/F_0$ over trials for all responsive neurons in both V1 (N = 212) and AL (N = 168) across animals was calculated, as a measure of neuronal response reliability, where smaller coefficients of variation indicated more consistent, and thus more reliable neuronal responses to same-type stimuli at around-threshold (i.e. trials with stimuli intensity of ±25% around perceptual threshold) intensity.

Data was analyzed using four 2-way ANOVAs, with pupil size and lick response (i.e. HIT or MISS) predicting coefficients of variation. We found no significant effect of neuronal activity for pupil size and lick response on the coefficient of variation for visual trials in V1 (see Figure 7.A) nor AL (see Figure 7.B). Similarly, no results were found audiovisual trials, in V1 (see Figure 7.C), nor AL (see Figure 7.D). Thus, neuronal response reliability did not influence mice performance during the task, nor was it impacted by the animal's arousal state. It is therefore unlikely that neuronal response reliability influences multisensory integration mechanisms.

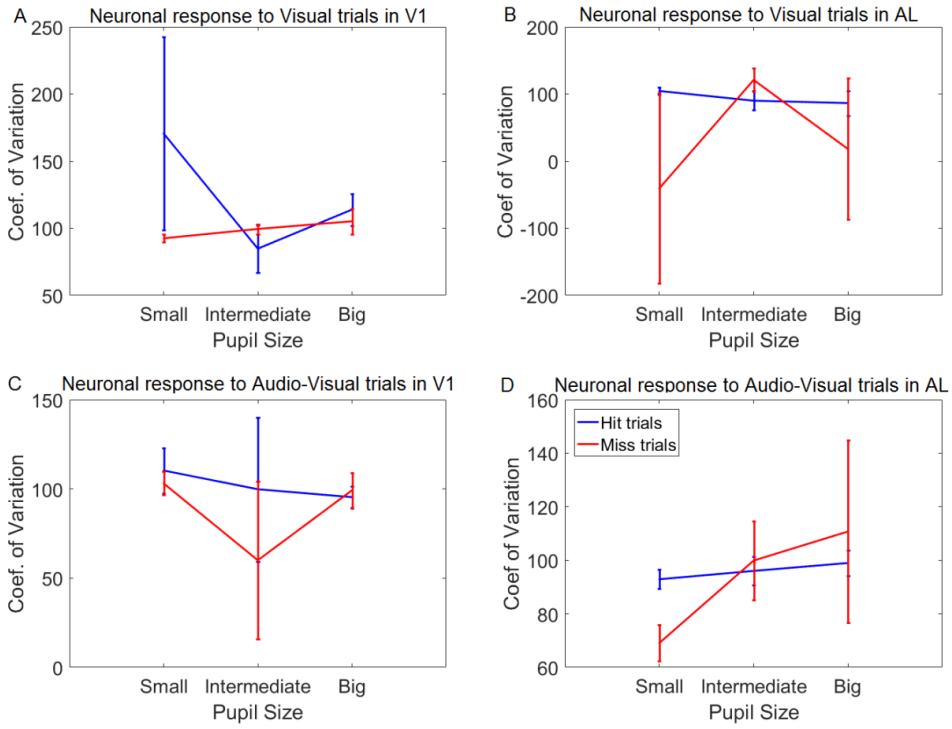

Figure 7.
Variability of neuronal response does not differ significantly between arousal states nor stimuli detection. (A) Plot of the coefficient of variation of V1 neurons during visual trials at around-threshold stimuli intensity (y-axis) for HIT trials (blue) and MISS trials (red) based on the corresponding average pupil size during these trials (x-axis). (B) Same as 7.A for AL neurons. (C) Plot of the coefficient of variation of V1 neurons during audiovisual trials at around-threshold stimuli intensity on the y-axis for HIT trials (blue) and MISS trials (red) based on the corresponding average pupil size during these trials on the x-axis. (D) Same as 7.C for AL neurons. Bars indicate SEM.

Decorrelation of neural responses improves information encoding

To further explore the effect of noise correlations on stimuli perception, as well as the influence of arousal on such noise structure, we calculated noise correlations in $\Delta F/F_0$ over visual and audiovisual trials at threshold intensity for all responses neurons in both V1 (N = 212) and AL (N = 247) across animals. Trials were split according to their respective HIT or MISS animal response, where a HIT response indicates stimuli perception.

We found that during both audiovisual and visual trials, for neurons in V1 and AL, noise correlations were affected by both pupil size and licking response (four 2-way ANOVA: F(2, 1271) = 6.93, p = .001, F(1, 1271) = 124.71, p < .0001 F(1, 1481) = 72.57, p < .0001 F(2, 1271) = 4.97, p = .007). Additional Bonferroni-corrected (p = .003) one-tailed independent-sample t-tests revealed that neurons in V1 were

significantly more decorrelated during HIT trials compared to MISS trials when the pupil was at intermediate ($t(377) = -4.73.$, $p < .003$) and small size ($t(409) = 4.43$, $p < .001$). Moreover, noise decorrelation was higher during HIT trials at intermediate pupil size compared to big pupil size ($p = .0009$; Fig. 8A). During audiovisual trials, for the same neurons in V1, noise correlation was significantly smaller during HIT trials compared to MISS trials, regardless of pupil size (see Figure 8.C). Similarly, for neurons in AL, during visual trials, noise was significantly less correlated during HIT trials compared to MISS trials, regardless of pupil size, (see Figure 8.B). During audiovisual trials, for the same neurons in AL, noise was significantly less correlated for HIT trials compared to MISS trials at small ($t(355) = -5.13.$, $p < .003$) and big pupil sizes ($t(390) = -3.15.$, $p < .003$), with noise correlation being significantly smaller during HIT trials at small pupil size compared to big pupil size ($p < .0001$; see Figure 8.D).

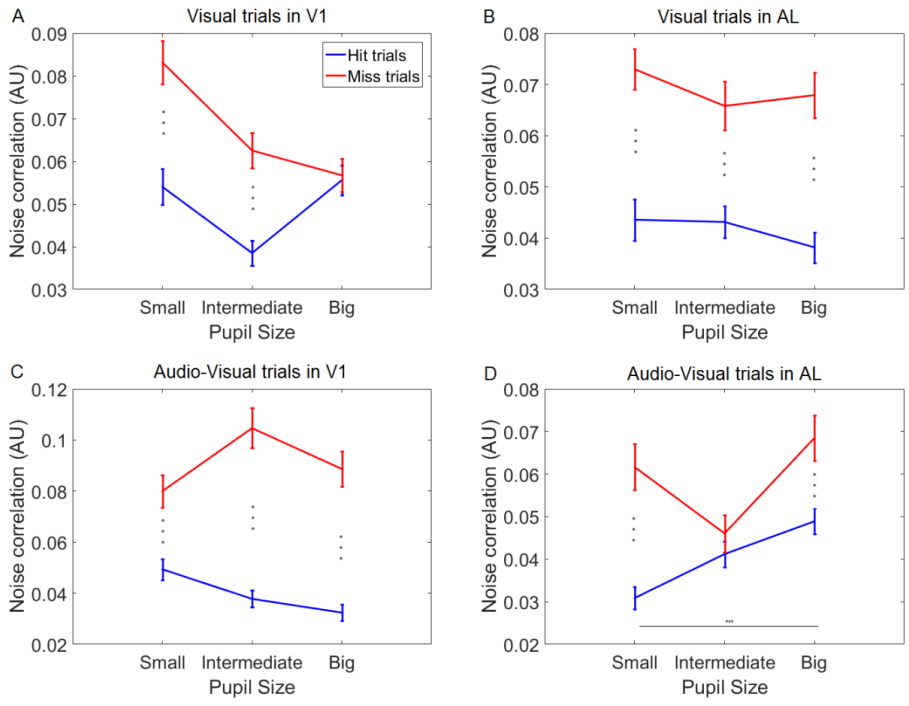

Figure 8.
Neuronal responses of neurons in V1 and AL are significantly less correlated during HIT trials compared to MISS trials. (A) Plot of the average noise correlation index of V1 neurons during visual trials at around-threshold stimuli intensity (y-axis) for HIT trials (blue) and MISS trials (red) based on the corresponding average pupil size during these trials (x-axis). (B) Same as 8.A for neurons in AL. (C) Plot of the average noise correlation index of V1 neurons during audiovisual trials at around-threshold stimuli intensity (y-axis) for HIT trials (blue) and MISS trials (red) based on the corresponding average pupil size during these trials (x-axis). (D) Same as 8.C for neurons in AL. Bars indicate SEM. *** $p < .001$

Thus, when animals successfully perceived stimuli, multi and unimodal alike, noise in the response of V1 and AL neurons underwent significant decorrelations. In other words, neuronal activity cortices which is unaccounted for in primary sensory, contains information relevant to stimuli processing and thus perception; specifically, correlation between neurons' response variability impairs stimuli processing and subsequent perception in both V1 and AL. The level of arousal upon stimuli presentation also impacted amounts of noise correlations, however inconsistently so between trials with different stimuli type and between areas. Unexpectedly, intermediate levels of arousal only induced higher decorrelation of neurons in V1 during visual trials.

Discussion

Multisensory integration is a neuronal process in which sensory information from different modalities are combined to enable perception of a unified event, and improve detection of such event when stimuli are individually faint. We performed calcium imaging of neuronal populations in V1 and AL of mice when they performed an audiovisual stimuli detection task with varying stimuli intensity to observe the neuronal mechanisms of multisensory integration in both regions, taking into consideration the role of arousal. Hence, performance on the detection task was first analyzed to determine whether there was a behavioral difference between multimodal and unimodal trials: Improved performance on the task for multimodal stimuli, as found at both levels of stimuli intensity, indicated the multimodal enhancement effect attributed to underlying multisensory integration (Ernst & Banks, 2002; Fetsch et al., 2013; Gleiss & Kayser, 2012; Gu et al., 2008; Kayser et al., 2010; Lippert et al., 2007; Stein & Stanford, 2008). We were especially interested in this effect for trials at sub-threshold stimuli intensity, as multisensory integration is theorized to be activated and take more effect when unimodal stimuli are too faint to be perceived when presented individually (Stein & Stanford, 2008). For sub-threshold trials, stimuli were detected below or around chance level when presented individually, but were detected more than 70% of the times when presented together. An alternative way to observe the effect of multimodal stimuli on

performance was to compute the perceptual thresholds of animals, that is, the minimum stimuli intensity necessary for animals to detect stimuli more often than not (i.e. >50%). We found that the perceptual threshold of mice was lowered during multimodal trials compared to unimodal ones, an effect once again attributable to underlying multisensory integration (Hollensteiner, et al., 2015).

These results altogether gave us confidence that neuronal activity recorded during the task would reflect the mechanisms of multisensory integration. Additionally, mice had been trained successfully to respond only upon stimuli detection, and their neuronal activity was recorded once their false alarm rate decreased below 20%, which gave us confidence that the hit rate analyzed for all trials reflected stimuli detection. We thenceforth assumed that when mice responded during trials (i.e. HIT trials), it was solely because they successfully detected the stimuli.

We then investigated the role of arousal on detection task performance to determine whether arousal levels interacted with multisensory integration in improving stimuli detection. We however found no effect of arousal on performance, regardless of stimuli type or intensity. These results contradict past findings that performance on detection tasks depend on how much an animal pays attention to relevant information about its surrounding (Bennet et al., 2013; Ecker et al., 2014; Reimer et al., 2014; Vinck et al., 2015) and that intermediate levels of arousal improve task performance in general (Yerkes & Dodson, 1908; McGinley et al., 2015a). We thus concluded that arousal did not, in this present study, influence the effects of multisensory integration, meaning that multisensory integration occurs and enables stimuli perception regardless of arousal levels upon stimuli presentation.

We then proceeded to analyze the neuronal activity in V1 and AL while the animals detected task stimuli to uncover the mechanisms of multisensory integration. We firstly expected to observe the existence of multisensory neurons in both V1 and AL, which enabled multisensory integration. These multisensory neurons were expected to be more active during multisensory trials compared to unisensory trials, and to respond weakly to visual stimuli as their primary function was to process multimodal

information. On the other hand, visual neurons were expected to process visual information, as being located in the visual cortex, but suppress their activity during multisensory trials. According to the theory of multisensory integration, we expected the difference between multisensory neurons and others to be apparent when stimuli were weak but not when stimuli were above the perceptual threshold. Equally, we did not expect these enhancement and suppression effects when the animals did not detect the stimuli, as it was assumed that in these cases multisensory integration did not occur. Our data confirmed our hypothesis in V1.

This finding indicates that only when faint multisensory stimuli were detected, which is a context favorable for the activation of multisensory integration, V1 contains neurons dedicated to process such multisensory information while visual neurons suppress their activity. We presently propose that detection of multimodal stimuli below the unimodal perceptual threshold was enabled by such specialized coding scheme in V1, and that such mechanism enables multisensory integration, corroborating similar findings (Kayser et al., 2010; Kayser et al., 2008; Meijer, Montijn, Pennartz, & Lansink, under review). Our findings also clarify the suppression of neuronal activity in primary sensory cortices in response to multimodal stimuli found by several studies (Ibrahim et al., 2016; Kayser et al., 2010). Indeed, multisensory stimuli do not alter the overall amount of activity in primary sensory cortices, but rather lead to visual-neuron suppression, which are otherwise strongly activated by visual stimuli. Importantly, our study is the first to our knowledge to record neuronal activity when mice are performing on a stimuli detection task. Taking in consideration animal responses to stimuli improved control over the selection of trials in which multisensory integration was expected to happen, and insured the proper manipulation of stimuli intensities with respond to perceptual thresholds. Future studies may explore further differences (e.g. morphological) between multisensory and visual neurons, a distinction that was not accessible using two-photon imaging.

The enhancement/suppression dynamic was present in AL when detected stimuli at a strong intensity. These findings are unprecedented as AL has not, to our knowledge, yet been studied in the context of multisensory integration. It is difficult to interpret such findings, as multisensory integration was unlikely to have occurred during strong stimuli intensity trials when it did not seem to occur during weak stimuli intensity trials. Thus, if AL is a site of multisensory integration, its neurons use a different mechanism to integrate multisensory cues compared to neurons in V1, a suggestion that should be specifically investigated. It is also possible that AL contains neurons that respond preferentially to audio-only stimuli, a sub-population which was not taken in consideration when we categorized AL neurons. According to these results, the question of whether or not AL is a site of multisensory integration remains unanswered.

This study innovatively took into consideration stimuli type and intensity, as well as arousal and behavioral outcome (i.e. stimuli detection or failure to detect) when considering the mechanisms of multisensory integration. This allowed valuable and wholesome insights into the mechanisms of multisensory integration but involved a noteworthy limitation as splitting trials several ways to obtain 12 to 18 trial groups lowered our analytical power to produce statistically robust results and increased changes of Type I and Type II errors. We therefore suggest further research to keep this limitation in mind and attempt to replicate our paradigm using a bigger number of trials.

Finally, we wanted to explore further (McGinley et al., 2015b) the role of noise structures in neuronal activity encoding unisensory and multisensory information in stimuli detection and the effect of arousal on such computations in V1 and AL. Specifically, we investigated whether neuronal response variability and noise correlations was significantly different when stimuli were detected or undetected, and whether arousal was a factor that influenced these two noise structures. We found that neuronal response variability in neither V1 nor AL changed when stimuli were detected compared to when they were not, and were not affected by arousal. We concluded that the neuronal response variability was not a factor influencing the effect of multisensory integration on stimuli detection. However, it is possible that the

mean neuronal response in each group, coupled with the deviation of such responses over trials, contains relevant information about the effect of arousal and behavioral responses on neuronal activity and noise, as observed by McGinley and colleagues (2015a, 2015b). Furthermore, neuronal response during trials was calculated during the pre-lick window, to reflect neuronal activity induced by stimuli perception and to avoid analyzing neuronal activity induced by another factor, such as motor behavior. Thus, our go-no-go detection task may not be suited to observe and interpret significant neuronal response variability.

Stimuli detection occurred during underlying noise decorrelation, in both V1 and AL. These results corroborated similar findings on the effect of active states on noise correlations, where decorrelation of noise improved stimuli encoding in sensory cortices (Bennett et al., 2013; McGinley et al., 2015b; Reimer et al., 2014; Schölvinck et al., 2015; Vinck et al., 2015). Thus, one may suggest that failure to detect stimuli was partially due to the state of the neuronal network upon stimuli presentation, where sensory information was not efficiently distributed between neurons (i.e. neurons process the same information; Schölvinck et al., 2015; Vinck et al., 2015), and that higher noise correlations are thus detrimental to sensory information neuronal encoding. It is unlikely that noise decorrelation is a mechanism specific to multisensory integration as it occurs during visual trials as well as audiovisual trials. These results however represent an opportunity for future research to clarify the exact role that noise correlations have on encoding sensory information and stimuli perception. The effect of arousal in noise decorrelations is difficult to interpret as our results were inconsistent across trials with different stimuli modalities. It is possible, as mentioned earlier, that splitting trials into so many subtypes reduced statistical power in our analysis. Thus, replication of these analyses with more trials and neurons per area are required, along with further developments in theorizing the role of noise correlations in sensory information processing.

In conclusion, this study simultaneously analyzed the effects of multimodal stimuli on both performance and underlying neuronal activity. This study also investigated whether arousal and noise structures were factors influencing stimuli detection during multimodal and unimodal stimuli

presentation, to provide original insights into the context (i.e. animal engagement with task) and mechanisms of sensory information processing and detection performance. Thus, our study both corroborated previous findings on the mechanisms of multisensory integration and raised important questions on the validity and replicability of some previous findings, namely, on the role of arousal and neuronal response variability on stimuli detection.

References

Anastasio, T. J., Patton, P. E., & Belkacem-Boussaid, K. (2000). Using Bayes' rule to model multisensory enhancement in the superior colliculus. *Neural Computation*, *12*(5), 1165-1187.

Andermann, M.L., Kerlin, A.M., Reid, R.C., 2010. Chronic Cellular Imaging of Mouse Visual Cortex During Operant Behavior and Passive Viewing. Frontier in Cellular Neuroscience. 4.

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.

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annual Review Neuroscience*, *28*, 403-450.

Averbeck, B. B., Latham, P. E., & Pouget, A. (2006). Neural correlations, population coding and computation. *Nature reviews Neuroscience*, *7*(5), 358-366.

Bennett, C., Arroyo, S., & Hestrin, S. (2013). Subthreshold mechanisms underlying state-dependent modulation of visual responses. *Neuron*, 80(2), 350-357.

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, *45*(4), 602-607.

Bradshaw, J. (1967). Pupil size as a measure of arousal during information processing. *Nature*, 216, 515–516

Buss, E., Hall III, J. W., & Grose, J. H. (2009). Psychometric functions for pure tone intensity discrimination: Slope differences in school-aged children and adults. *The Journal of the Acoustical Society of America*, *125*(2), 1050-1058.

Cappe, C., Rouiller, E.M., and Barone, P. (2009). Multisensory anatomical pathways. *Hearing Research, 258,* 28–36.

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.

Coull, J. T., Jones, M. E., Egan, T. D., Frith, C. D., & Maze, M. (2004). Attentional effects of noradrenaline vary with arousal level: selective activation of thalamic pulvinar in humans. *Neuroimage, 22*(1), 315-322.

Ecker, A. S., Berens, P., Cotton, R. J., Subramaniyan, M., Denfield, G. H., Cadwell, C. R., ... & Tolias, A. S. (2014). State dependence of noise correlations in macaque primary visual cortex. *Neuron, 82*(1), 235-248.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*(6870), 429-433.

Faisal, A. A., Selen, L. P., & Wolpert, D. M. (2008). Noise in the nervous system. *Nature reviews neuroscience, 9*(4), 292-303.

Fetsch, C. R., DeAngelis, G. C., & Angelaki, D. E. (2013). Bridging the gap between theories of sensory cue integration and the physiology of multisensory neurons. *Nature Reviews Neuroscience, 14(6),* 429-442.

Franke, F., Fiscella, M., Sevelev, M., Roska, B., Hierlemann, A., & da Silveira, R. A. (2016). Structures of neural correlation and how they favor coding. *Neuron, 89*(2), 409-422.

Guizar-Sicairos, M., Thurman, S. T., & Fienup, J. R. (2008). Efficient subpixel image registration algorithms. *Optics letters, 33*(2), 156-158.

Gleiss, S., and Kayser, C. (2012). Audiovisual Detection Benefits in the Rat. *PLoS ONE 7*, e45677.

Glickfeld, L. L., Histed, M. H., & Maunsell, J. H. (2013). Mouse primary visual cortex is used to detect both orientation and contrast changes. *Journal of Neuroscience, 33(50),* 19416-19422.

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. *Nature Protocol. 9*, 2515–2538.

Grzelkowski, S., Meijer, G.T., (unpublished). Influence of arousal state on encoding of visual stimuli by neuronal populations in primary visual cortex.

Gu, Y., Angelaki, D. E., & DeAngelis, G. C. (2008). Neural correlates of multisensory cue integration in macaque MSTd. *Nature neuroscience*, *11*(10), 1201-1210.

Hollensteiner, K. J., Pieper, F., Engler, G., König, P., & Engel, A. K. (2015). Crossmodal integration improves sensory detection thresholds in the ferret. *PloS one, 10(5),* e0124952.

Ibrahim, L.A., Mesik, L., Ji, X., Fang, Q., Li, H., Li, Y., 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.

Iurilli, G., Ghezzi, D., Olcese, U., Lassi, G., Nazzaro, C., Tonini, R., Tucci, V., Benfenati, F., and Medini, P. (2012). Sound-Driven Synaptic Inhibition in Primary Visual Cortex. *Neuron 73*, 814–828.

Kayser, C., Petkov, C.I., and Logothetis, N.K. (2008). Visual Modulation of Neurons in Auditory

Cortex. *Cerebral Cortex 18*, 1560–1574.

Kayser, C., Logothetis, N.K., and Panzeri, S. (2010). Visual Enhancement of the Information Representation in Auditory Cortex. *Current Biology, 20*, 19–24.

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. European. *Journal of Neuroscience, 34*, 65–78.

Lippert, M., Logothetis, N.K., and Kayser, C. (2007). Improvement of visual contrast detection by a simultaneous sound. *Brain Research, 1173*, 102–109.

Meijer, G.T., Montijn, J.S., Pennartz, C.M.A., Lansink, C.S., (under review). Dynamic encoding of cross-modal stimuli in the primary visual cortex of the mouse.

McGinley, M.J., David, S.V., and McCormick, D.A. (2015). Cortical Membrane Potential Signature of Optimal States for Sensory Signal Detection. *Neuron*, *87*, 179–192.

McGinley, M. J., Vinck, M., Reimer, J., Batista-Brito, R., Zagha, E., Cadwell, C. R., ... & McCormick, D. A. (2015). Waking state: rapid variations modulate neural and behavioral responses. *Neuron*, *87*(6), 1143-1161.

Meredith, M.A., and Stein, B.E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of neurophysiology*, *56*(3), 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. *Frontier in Computational Neuroscience, 8,* 58.

Olcese, U., Iurilli, G., & Medini, P. (2013). Cellular and synaptic architecture of multisensory integration in the mouse neocortex. *Neuron, 79*(3), 579-593.

Paxinos, G., and Franklin, K.B.J. (2004). The Mouse Brain in Stereotaxic Coordinates (Elsevier/Academic Press).

Reimer, J., Froudarakis, E., Cadwell, C. R., Yatsenko, D., Denfield, G. H., & Tolias, A. S. (2014). Pupil fluctuations track fast switching of cortical states during quiet wakefulness. *Neuron, 84*(2), 355-362.

Schölvinck, M. L., Saleem, A. B., Benucci, A., Harris, K. D., & Carandini, M. (2015). Cortical state determines global variability and correlations in visual cortex. *Journal of Neuroscience, 35*(1), 170-178.

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews Neuroscience, 9(4),* 255-266.

Vinck, M., Batista-Brito, R., Knoblich, U., & Cardin, J. A. (2015). Arousal and locomotion make distinct contributions to cortical activity patterns and visual encoding. *Neuron, 86*(3), 740-754.

Wallace, M. T., & Stein, B. E. (1997). Development of multisensory neurons and multisensory integration in cat superior colliculus. *Journal of Neuroscience, 17(7),* 2429-2444. Chicago

Wang, Q., & Burkhalter, A. (2007). Area map of mouse visual cortex. *Journal of Comparative Neurology, 502*(3), 339-357.

Wang, Q., Gao, E., & Burkhalter, A. (2011). Gateways of ventral and dorsal streams in mouse visual cortex. *Journal of Neuroscience, 31(5),* 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. *Journal of Neuroscience*, *32*, 4386–4399.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of intensity of stimulus to rapidity of habit-formation. *Journal of comparative neurology*, *18*(5), 459-482.

Zoccolan, D., Graham, B.J., and Cox, D.D. (2010). A self-calibrating, camera-based eye tracker for the recording of rodent eye movements. *Decision in Neuroscience, 4*, 193.