Sound-induced contextual modulation of the representation of oriented visual stimuli in the primary visual cortex of awake behaving mice

by

John P. McClure Jr.

A dissertation submitted to the
Graduate School-Newark, Rutgers, The State University of New Jersey
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
Graduate Program in Behavioral and Neural Sciences
written under the direction of
Dr. Pierre-Olivier Polack
and approved by:

_____

Bart Krekelberg, PhD

_____

Pierre-Olivier Polack, DVM., PhD

_____

Tibor Koós, PhD

_____

Laszlo Zaborszky, MD, PhD

_____

David Margolis, PhD

Newark, New Jersey
January 2020

.

Abstract of the dissertation

Sound-induced contextual modulation of the representation of oriented visual stimuli in the primary visual cortex of awake behaving mice

By JOHN PATRICK MCCLURE JR.

Dissertation Advisor: Dr. Pierre-Olivier Polack

The cerebral cortex is continuously receiving sensory information provided by the five senses. The role of sensory processing is to identify and integrate the sensory information necessary for the selection and execution of behaviors most adapted to the environmental conditions. Most of the time, combining or comparing the signals provided by different sensory modalities leads to optimal processing of the information. The mechanisms responsible for the interactions between at least two different sensory modalities are critical for the production of a relevant behavioral response. However, the cellular and network activities underpinning multimodal integration are poorly understood. For decades, scientist thought that multisensory interactions occur only in 'higher-order' cortical regions. Only recently, converging evidence has shown that part of cross-modal cortical integration occurs as early as the primary sensory cortices, such as the existence of direct cortico-cortical projections between the primary auditory cortex (A1) and the primary visual cortex (V1). Despite these findings, the effects of sounds on visually-evoked responses along with the influence of the audiovisual context on visual processing and visual perception remain unclear. Therefore, two-photon calcium imaging was performed in V1 of awake mice

passively or actively processing oriented visual stimuli in order to determine the effect of sounds on visual processing. The results demonstrated an association between auditory stimulation and improvements on V1 representation of the orientation and the direction of the visual stimulus. This improvement of the representation of the visual stimulus was associated with a decrease in the limits of angular perception of mice performing an angular perception task. Finally, the results indicated that during active audiovisual processing, the modulation of the representation of the visual stimuli in V1 dynamically adapted to the needs of the behavioral task. Altogether, the results showed that other sensory modalities influence visual processing in V1 and visual perception, and are adaptable to the behavioral goal.

Acknowledgements

I want to thank my research advisor, Dr. Pierre-Olivier Polack for his mentorship and support throughout this process. Dr. Polack gave me the necessary tools to confidently conduct all of my experiments, critically evaluate my work, and the work of others. Through his guidance, I can present my thesis for consideration of a Ph.D. I also want to thank all of my lab family, both past and present members as follows: Norwin Ahmed (Nishe), Dr. Julien Corbo, Hussein Khdour, and Han Xu. Thank you for your contribution to my growth as a young scientist. A special thanks go to Dr. Corbo for helping with data analysis and the writing of this thesis document. I want to thank the members of my thesis committee: Dr. Bart Krekelberg, Dr. Tibor Koós, Dr. Laszlo Zaborszky, and Dr. David Margolis for their timeless advice and feedback on my research. I am also thankful for all of the philosophical discussions and fun adventures I shared with my BNS friends. Specifically, I want to thank Dr. Sally Wang for all of our excursions and for the friendly rivalry we shared on the squash courts.

I acknowledge and give thanks to my father John McClure Sr., my mother Kimberlyn Leader, my brother Jaquan McClure, and my sister Shanice McClure for always believing in me and for making sure I finish what I started. Finally, I am truly indebted to the love of my life Taytiana Welch-McClure. Thank you for sticking with me through this process and for your encouragement and boost along the way. This thesis is dedicated to you.

Preface

A version of the work presented in Chapter III has been published (McClure and Polack, 2019, *Journal of Neurophysiology*). Currently, Chapters IV and V are in preparation.

Table of Contents

CHAPTER I: GENERAL INTRODUCTION ..... **- 1 -**

1.1 SENSORY PROCESSING OF VISUAL AND AUDITORY STIMULI BY THE CEREBRAL CORTEX ..... **- 1 -**

1.1.1 Flow of visual information from the eye to the cerebral cortex ..... - 2 -

1.1.2 The primary visual cortex (V1): the first step of cortical processing ..... - 4 -

1.1.3 V1 to extrastriate visual cortices: the 'Where' and 'What' visual pathways ..... - 8 -

1.1.4 The auditory system: from the ear to the cerebral cortex ..... - 9 -

1.1.5 A1: Connectivity and function ..... - 10 -

1.2 PERCEPTUAL EFFECTS OF AUDIOVISUAL INTERACTIONS ..... **- 11 -**

1.2.1 Audiovisual integration of congruent stimuli ..... - 12 -

1.2.2 Audiovisual integration of incongruent stimuli ..... - 14 -

1.3 MULTISENSORY INTEGRATION: CONVERGENCE OF UNIMODAL INPUTS ..... **- 16 -**

1.3.1 Late integration model of multisensory interactions ..... - 16 -

1.3.2 Audiovisual integration in association cortices ..... - 18 -

1.3.3 The principles of multisensory interactions ..... - 21 -

1.4 MULTISENSORY INTERACTIONS IN PRIMARY SENSORY AREAS ..... **- 22 -**

1.4.1 Early integration model of multisensory interactions ..... - 22 -

1.4.2 Anatomical evidence for multimodal interactions in primary sensory cortices ..... - 23 -

1.4.3. Functional evidence for multimodal interactions in primary sensory areas ..... - 24 -

1.5 SOUND-INDUCED MODULATION OF V1 ACTIVITY IN MICE ..... **- 27 -**

1.5.1 Sound-Driven Synaptic Inhibition in Primary Visual Cortex ..... - 27 -

1.5.2 Cross-Modality Sharpening of Visual Cortical Processing through Layer-1-Mediated Inhibition and Disinhibition ..... - 30 -

1.5.3 Audiovisual Modulation in Mouse Primary Visual Cortex Depends on Cross-Modal Stimulus Configuration and Congruency ..... - 32 -

1.5.4 A comparison of the key findings from Iurilli, Ibrahim, and Meijer ..... - 34 -

1.6 GOALS OF THESIS ..... - 35 -

CHAPTER II: GENERAL METHODS ..... - 37 -

2.1 ANIMALS ..... - 37 -

2.2 STEREOTAXIC SURGERY ..... - 37 -

2.2.1 Head bar implant ..... - 37 -

2.2.2 Adeno-associated virus injections ..... - 39 -

2.2.3 Coverslip implant surgery ..... - 40 -

2.3 *IN VIVO* CALCIUM IMAGING ..... - 41 -

CHAPTER III ..... - 44 -

3.1 PURE TONES MODULATE THE REPRESENTATION OF ORIENTATION AND DIRECTION IN THE PRIMARY VISUAL CORTEX ..... - 44 -

3.2 BACKGROUND ..... - 44 -

3.3 HYPOTHESIS ..... - 45 -

3.4 MATERIALS AND METHODS ..... - 46 -

3.4.1 Animals ..... - 46 -

3.4.2 Stereotaxic surgery ..... - 46 -

3.4.3 In vivo calcium imaging ..... - 46 -

3.4.4 Audiovisual stimuli ..... - 46 -

3.4.5 EEG electrode implants ..... - 48 -

3.4.6 EEG recordings ..... - 49 -

3.4.7 Data analysis ..... - 49 -

3.4.8 Statistics ..... - 52 -

3.5 RESULTS ..... - 53 -

3.5.1 Representation of orientation and direction in the unimodal and audiovisual contexts ..... - 57 -

3.5.2 Sound modulation as a function of the neurons' orientation preferences ..... - 59 -

3.5.3 Sound modulation as a function of the orientation and direction selectivity of V1 neurons.. ..... - 62 -

3.5.4 Locomotion and Arousal are Similar in Visual-Only and Audiovisual Contexts ..... - 67 -

3.6 DISCUSSION ..... - 69 -

CHAPTER IV ..... - 75 -

4.1 PURE TONES IMPROVE THE REPRESENTATION OF ORIENTED STIMULI IN V1 AND FACILITATE ANGULAR DISCRIMINATION ..... - 75 -

4.2 BACKGROUND ..... - 75 -

4.3 MATERIALS AND METHODS ..... - 76 -

4.3.1 Animals ..... - 77 -

4.3.2 Stereotaxic surgery ..... - 77 -

4.3.3 In vivo calcium imaging ..... - 77 -

4.3.4. Audiovisual stimuli ..... - 77 -

4.3.5 Data analysis ..... - 80 -

4.3.6 Statistics ..... - 84 -

4.4 RESULTS ..... - 85 -

4.4.1 Behavioral performance as a function of the angular distance between the Go and NoGo visual cues ..... - 86 -

4.4.2 Sound-induced modulation of the behavioral performances as a function of the angular distance between the Go and NoGo cues ..... - 87 -

4.4.3 Representation of the orientation in V1 of a Go and NoGo stimuli distant by 45° ..... - 90 -

4.4.4 Representation of the orientation in V1 of a Go and NoGo stimuli distant by 30° ..... - 94 -

4.4.5 Representation of the orientation in V1 of a Go and NoGo stimuli distant by 25° ..... - 97 -

4.4.6 Representation of the orientation in V1 of a Go and NoGo stimuli distant by 20° ..... - 101 -

4.4.7 Representation of the orientation in V1 of a Go and NoGo stimuli distant by 15° ..... - 104 -

4.5 DISCUSSION ..... - 106 -

4.5.1 Effect of learning in V1: the presence of attractors in the representations ..... - 107 -

4.5.2 Linking evoked activity in V1 and perceptual orientation discrimination ..... - 108 -

CHAPTER V ..... - 110 -

5.1 ADAPTATION TO THE BEHAVIORAL GOAL OF THE REPRESENTATION OF THE ORIENTATION AND DIRECTION OF THE VISUAL STIMULI IN A CROSS-MODAL DISCRIMINATION TASK ..... - 110 -

5.2 BACKGROUND ..... - 110 -

5.3 HYPOTHESIS ..... - 111 -

5.4 MATERIALS AND METHODS ..... - 111 -

5.4.1 Animals ..... - 111 -

5.4.2 Stereotaxic surgery ..... - 111 -

5.4.3 In vivo calcium imaging ..... - 111 -

5.4.4. Audiovisual stimuli ..... - 112 -

5.4.5 Data analysis ..... - 114 -

5.4.6 Statistics ..... - 119 -

5.5 RESULTS ..... - 119 -

5.5.1 Behavioral performance in a cross-modal discrimination task ..... - 119 -

5.5.2 Sound shapes licking onset and offset during the cross-modal behavioral task ..... - 121 -

5.5.3 Locomotion and arousal are similar in visual-only and audiovisual contexts ..... - 122 -

5.5.4 Modulation by sound of the representation of orientation in V1 in naïve mice. ..... - 123 -

5.5.5 Modulation by sound of the representation of orientation in V1 in mice performing a cross-modal discrimination task. ..... - 126 -

5.5.6 Modulation by sound of the representation of orientation in V1 in trained mice passively processing the audiovisual stimuli. ..... - 128 -

5.6 DISCUSSION ..... - 131 -

5.6.1 Cross-modal integration and behavioral performance ..... - 132 -

5.6.2 Assessing the representation of the orientation of the visual stimulus in V1 with shallow neuronal networks (SNNs) ..... - 133 -

CHAPTER VI: GENERAL DISCUSSION ..... - 135 -

6.1 AUDITORY TONES IMPROVE THE REPRESENTATION OF ORIENTATION AND DIRECTION OF VISUAL STIMULI IN V1 ..... - 136 -

6.2 AUDITORY STIMULI SHARPEN THE REPRESENTATION OF ORIENTATION IN V1 DURING AN ANGULAR PERCEPTION TASK AND IMPROVES PERCEPTUAL ORIENTATION DISCRIMINATION - 138 -

6.3 SOUND DYNAMICALLY ADAPTS V1 TO THE BEHAVIORAL GOALS ..... - 140 -

6.4 INNOVATION OF THE THESIS EXPERIMENTS ..... - 141 -

6.5 DISORDERS RELATED TO ALTERATIONS IN CROSS-MODAL PERCEPTION ..... - 142 -

LIST OF REFERENCES ..... - 144 -

List of Figures

Chapter I

FIGURE 1-1. CELLULAR COMPONENTS AND SYNAPTIC INTERACTIONS IN THE PRIMATE RETINA. ..... - 4 -

FIGURE 1-2. ILLUSTRATION OF THE SIMPLIFIED LAMINAR ORGANIZATION OF THE MOUSE V1. ..... - 5 -

FIGURE 1-3. SPIKE TRAINS OF A V1 NEURON DURING VISUAL STIMULATION OF VARIOUS ORIENTED BARS. ..... - 6 -

FIGURE 1-4. ACTION POTENTIAL DISCHARGE OF A NEURON IN THE MACAQUE MONKEY V1 DURING THE PRESENTATION OF MOVING BARS. ..... - 7 -

FIGURE 1-5. SCHEMATIC REPRESENTATION OF THE VISUAL SYSTEM ORGANIZATION IN THE MACAQUE MONKEY. ..... - 9 -

FIGURE 1-6. TONOTOPIC ORGANIZATION OF A1 IN A RAT. ..... - 10 -

FIGURE 1-7. AUDIOVISUAL ENHANCEMENT IN LOUDNESS DETECTION. ..... - 12 -

FIGURE 1-8. THE EFFECTS OF AUDIOVISUAL INTEGRATION ON STIMULUS DETECTION IN MICE. ..... - 13 -

FIGURE 1-9. THE VENTRILOQUISM EFFECT IN MACAQUE MONKEYS. ..... - 16 -

FIGURE 1-10. THE LATE INTEGRATION MODEL OF MULTISENSORY INTEGRATION. ..... - 17 -

FIGURE 1-11. TEMPORAL POLE ACTIVATION DURING AUDIOVISUAL STIMULATION…… ..... - 19 -

FIGURE 1-12. EFFECTS OF AUDIOVISUAL STIMULI ON STS NEURONAL RESPONSES IN RHESUS MONKEYS. ..... - 21 -

FIGURE 1-13. THE EARLY INTEGRATION MODEL OF MULTISENSORY INTEGRATION…… ..... - 23 -

FIGURE 1-14. MULTIUNIT ACTIVITY IN A1 DURING ACOUSTIC AND NON-ACOUSTIC EVENTS OF AN AUDITORY CATEGORIZATION TASK. ..... - 26 -

FIGURE 1-15. THE EFFECTS OF A1 ACTIVATION ON V1 ACTIVITY. ..... - 29 -

FIGURE 1-16. SOUND-INDUCED SHARPENING OF ORIENTATION SELECTIVITY IN V1 L2/3.....- 31 -

FIGURE 1-17. THE IMPACT OF AUDIOVISUAL STIMULUS FEATURES ON MULTISENSORY INTEGRATION IN V1. .....- 33 -

Chapter III

FIGURE 3-1. RESPONSES OF THE PRIMARY VISUAL CORTEX LAYER 2/3 NEURONAL POPULATION TO DRIFTING GRATINGS IN VISUAL-ONLY AND AUDIOVISUAL CONTEXTS.....- 55 -

FIGURE 3-2. ORIENTATION OF THE POPULATION OF PRIMARY VISUAL CORTEX LAYER 2/3 NEURONS ACTIVE IN THE VISUAL-ONLY AND AUDIOVISUAL CONTEXTS.....- 58 -

FIGURE 3-3. SOUND MODULATION AS A FUNCTION OF PRIMARY VISUAL CORTEX LAYER 2/3 NEURONS' ORIENTATION TUNING. .....- 61 -

FIGURE 3-4. SOUND MODULATION AS A FUNCTION OF PRIMARY VISUAL CORTEX LAYER 2/3 NEURONS' ORIENTATION SELECTIVITY. .....- 64 -

FIGURE 3-5. SOUND MODULATION AS A FUNCTION OF PRIMARY VISUAL CORTEX LAYER 2/3 NEURONS' DIRECTION SELECTIVITY.....- 66 -

FIG. 3-6. BRAIN STATES DURING VISUAL PRESENTATION IN THE VISUAL AND AUDIOVISUAL CONTEXTS. .....- 68 -

Chapter IV

FIGURE 4-1. GO/NOGO VISUAL DISCRIMINATION TASK.....- 80 -

FIGURE 4-2. SCHEMATIC OF THE RESAMPLING METHOD FOR THE GENERATION OF POPULATION PSTHS. .....- 83 -

FIGURE 4-3. LICK RESPONSES OF A SINGLE MOUSE DURING THE GO/NOGO ORIENTATION DISCRIMINATION TASK. .....- 87 -

FIGURE 4-4. SOUND-INDUCED EFFECTS ON POPULATION PERFORMANCE IN AN ORIENTATION DISCRIMINATION TASK. ..... - 88 -

FIGURE 4-5. SOUND-INDUCED EFFECTS ON INDIVIDUAL MICE PERFORMANCE IN THE ORIENTATION DISCRIMINATION TASK. ..... - 89 -

FIGURE 4-6. V1 REPRESENTATION OF ORIENTATION AT 45^{O} AD WITH AND WITHOUT AUDITORY STIMULATION. ..... - 93 -

FIGURE 4-7. V1 REPRESENTATION OF ORIENTATION AT 30^{O} AD WITH AND WITHOUT AUDITORY STIMULATION. ..... - 96 -

FIGURE 4-8. V1 REPRESENTATION OF ORIENTATION AT 25^{O} AD WITH AND WITHOUT AUDITORY STIMULATION. ..... - 100 -

FIGURE 4-9. V1 REPRESENTATION OF ORIENTATION AT 20^{O} AD WITH AND WITHOUT AUDITORY STIMULATION. ..... - 103 -

FIGURE 4-10. V1 REPRESENTATION OF ORIENTATION AT 15^{O} AD WITH AND WITHOUT AUDITORY STIMULATION. ..... - 105 -

FIGURE 4-11. THE EFFECTS OF THE GO/NOGO STIMULI ON V1 REPRESENTATION OF ORIENTATION IN TUNING CURVE TRIALS. ..... - 108 -

Chapter V

FIGURE 5-1. GO/NOGO CROSS-MODAL DISCRIMINATION TASK. ..... - 114 -

FIGURE 5-2. ANALYSIS OF THE REPRESENTATION OF THE ORIENTATION OF DRIFTING GRATINGS IN V1 USING A SHALLOW NEURAL NETWORK (SNN). ..... - 118 -

FIGURE 5-3. SOUND MODULATION OF BEHAVIORAL PERFORMANCE ON A CROSS-MODAL TASK. ..... - 121 -

FIGURE 5-4. THE EFFECTS OF AUDIOVISUAL CUES ON LICK ONSET AND OFFSET DURING THE CROSS-MODAL TASK. ..... - 122 -

FIGURE 5-5. CHANGES IN AROUSAL AND LOCOMOTION DURING THE CROSS-MODAL DISCRIMINATION TASK. ..... - 123 -

FIGURE 5-6. MODULATION BY SOUND OF THE POPULATION RESPONSE OF V1 IN NAÏVE MICE. ..... - 126 -

FIGURE 5-7. MODULATION BY SOUND OF THE POPULATION RESPONSE OF V1 IN MICE ACTIVELY PERFORMING THE TASK. .....- 127 -

FIGURE 5-8. MODULATION BY SOUND OF THE POPULATION RESPONSE OF V1 IN MICE TRAINED TO THE TASK BUT PASSIVELY EXPOSED TO THE AUDITORY AND VISUAL STIMULI. .....- 130 -

Chapter VI

FIGURE 6-1: SOUND MODULATION OF V1 L2/3 GABAERGIC NEURONS. .....- 138 -

List of abbreviations

V1: Primary visual cortex

A1: Primary auditory cortex

LGN: Lateral geniculate nucleus

V2: Secondary visual cortex

V3: Third extrastriate visual area

V4: Fourth extrastriate visual area

L1-L6: Cortical layers 1-6

MT (V5): Middle temporal visual area

TEO: Occipitotemporal transition area

TE: Inferior temporal lobe

VIP: Ventral intraparietal area

MST: Medial superior temporal area

LIP: Lateral intraparietal sulcus

MGB: Medial geniculate body

MEG: Magnetoencephalography

TP: Temporal pole

TC: Temporally congruent

TI: Temporally incongruent

STS: Superior temporal sulcus

SC: Superior colliculus

FsB: Fast blue dye

ERP: Event-related potentials

SH: Sound hyperpolarization

AP: Anterior posterior

ML: Medial lateral

dF/F: fractional fluorescence

OSI: Orientation selectivity index

DSI: Direction selectivity index

TC: Tuning curves

CI: Confidence Interval

AV: Audiovisual

AD: Angular distance

EEG: Electroencephalography

SNN: Shallow neuronal network

DY: Diamidino yellow

Chapter I: General Introduction

1.1 Sensory processing of visual and auditory stimuli by the cerebral cortex

Animals, plants, gases, liquids, and objects are all defined by physical and chemical properties such as shape, color, odor, taste, texture, density, and many others. Throughout evolution, animal species have developed sensors to detect these specific physical or chemical properties to facilitate their interaction with their environment. In most cases, these sensors are localized in specialized sensory organs such as the eyes or the ears, in which sensory stimuli are transduced into electrical signals that are later integrated by specialized areas of the central nervous system, such as the cerebral cortex (Lodish et al., 2000).

The mechanisms by which sensory information, originating from sensory organs, is integrated by the cerebral cortex remain the object of an intense research effort. Most of the current studies use a unimodal approach whereby they record the responses of the cortical neurons for a specific unimodal stimulus, within the cortical regions dedicated to that modality. Yet, most objects are characterized by physical properties detected by more than one sensory organ. For example, a car is defined by its visual appearance, but also by its noise and its smell. Therefore, it has been necessary to generate a complete representation of the objects and to develop the capability to integrate multisensory information (Stein and Stanford, 2008). To this date, the mechanisms underpinning multisensory processing remain poorly understood. The integration of audiovisual stimuli is an excellent model to study

multimodal integration because visual and auditory stimuli can be easily, reliably, and accurately generated.

In the remainder of this introduction, I will briefly describe how unimodal visual and auditory information is processed, from the activation of sensory organs to the processing in higher-order unimodal sensory areas. After demonstrating that auditory and visual information interacts with each other during audiovisual processing, I will also highlight audiovisual integration within specialized brain regions. Then, I will show that while unimodal cortices were not initially thought to be involved in multisensory processing, audiovisual interactions do occur in V1.

1.1.1 Flow of visual information from the eye to the cerebral cortex

The role of the eye is to receive light from the environment and convert the photons into electrical signals before sending the information to the cerebral cortex for analysis; a process called phototransduction (Shastry, 1998). The photoreceptors, either rods or cones, are responsible for generating black-and-white (rods) or high-resolution color (cones) information in dim light and daylight conditions, respectively (Wolf, 2004). Rods convey mostly information on luminance intensity to the retinal ganglion cells, the output layer of the retina, due to their low spatial resolution but high light sensitivity (e.g., rods can produce a response to a single photon), while cones provide high-spatial resolution information. The retina also contains other neuronal types: the bipolar cells, which connect the photoreceptors to the ganglion cells, the horizontal, and the amacrine cells. The primary function of horizontal cells is to provide inhibitory feedback to rods and cones as a mechanism of 'local gain control'

of illumination levels. On the other hand, amacrine cells are widely distributed throughout the retina and represent the most diverse inhibitory cell type in the retina. Generally, they receive synaptic input from bipolar cells and other amacrine cells, and in turn, provide information to amacrine and ganglion cells as well as feedback to bipolar cells. While the function of different amacrine cell types remains poorly understood, studies have linked a subtype of these cells to the direction selectivity of some ganglion cells (Hanson et al., 2019; Ray et al., 2018). Finally, bipolar cells project to retinal ganglion cells where visual information is encoded for the first time as an action potential firing rate (Fig 1-1).

Retinal ganglion cells send information in the form of action potential discharges to the lateral geniculate nucleus (LGN) of the thalamus (See Masland, 2001; Masland, 2012; and Van Essen, 1992 for review). Retinal ganglion projections to the LGN represent only ~10% of the projections to the LGN. Indeed, most of the synapses on LGN neurons originate from the cortex and the brainstem (Van Horn et al., 2000), and play a role in adapting LGN activity during different behavioral states such as alertness and attention (Sherman and Guillery, 1998; Van Horn et al., 2000). The receptive fields of LGN neurons present a center/surround organization that is nearly identical to that of their retinal inputs (Alonso et al., 2006). The activity of the LGN is transmitted to V1where the next stage of visual processing takes place.

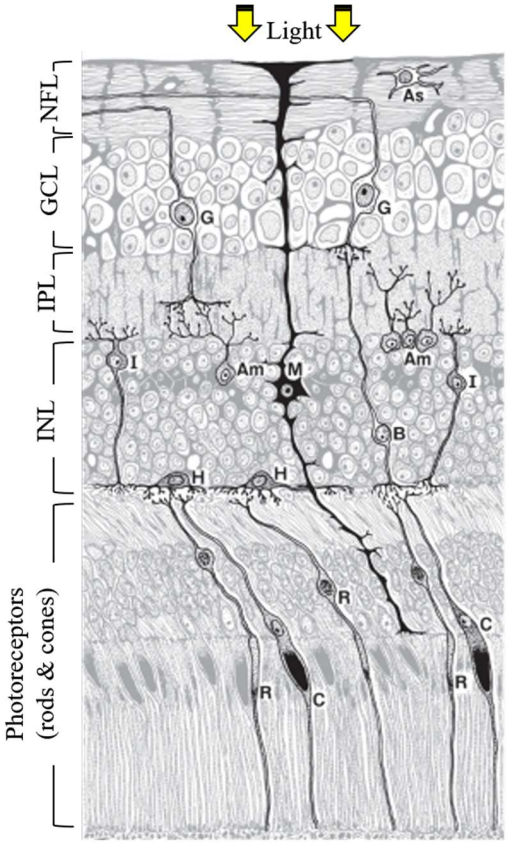

Figure 1-1. Cellular components and synaptic interactions in the primate retina. *Photoreceptors (rods (R) and cones (C)) transmit feed-forward visual information to bipolar cells (B) in the inner nuclear layer (INL) of the retina. In turn, visual information from bipolar cells is propagated through the inner plexiform layer (IPL) onto ganglion cells (G), located in the ganglion cell layer (GCL). Visual information is modulated by the activity of amacrine cells (Am) and horizontal cells (H) located in the INL. Also found in the INL are Müller (M) and Interplexiform cells (I). The axons of many retinal ganglion cells converge and form the nerve fiber, within the nerve fiber layer (NFL), before becoming the optic nerve, and later the optic tract, eventually synapsing onto neurons in the LGN of the thalamus. (From Zezo, 2016).*

1.1.2 The primary visual cortex (V1): the first step of cortical processing

V1 (also known as striate cortex or Brodmann area 17 in primates and felines), receives feed-forward projections from the LGN (Michael-Titus et al., 2010). In humans and non-human primates, V1 is located in the calcarine sulcus within the medial occipital lobe. V1 integrates visual information provided by the LGN and sends its output to other visual cortical areas, such as V2 (Sincich and Horton, 2002; Khibnik et al., 2014), V3 (Baldwin et al., 2012), and V4 (Baldwin et al., 2012).

Neurons in V1 are organized into six horizontal layers based on differences in histological appearance (Balaram, 2014). Each layer receives specific inputs and gives rise to distinct output projections (Self et al., 2013). Layer 4 (L4) is usually described

as the 'input' layer of V1 because it receives most of the feed-forward projections into V1 from the LGN (Self et al., 2013). Layer 2 and layer 3 (often referred to as layer 2/3 or L2/3, due to similarities in histological appearance and connections) receive most of their inputs from L4. L2/3 neurons project directly onto higher-order cortical areas (Remington, 2012). The 'output' layer of V1, Layer 5 (L5), contains relatively few cell bodies compared to the surrounding layers. L5 neurons project mainly to cortical and subcortical brain regions (Kim et al., 2015). Layer 6 (L6), the deeper layer of V1, is dense with cells, receive connections from both the LGN and other layers, and sends feedback projections to the LGN and the perigeniculate sector (PGN) of the thalamic reticular nucleus (Remington, 2012). Layer I (L1), the most superficial layer of V1, consists mainly of axon terminals from subcortical and interareal brain regions (Felleman and Van Essen, 1991; Ji et al., 2015; Rubio-Garrido et al., 2009; Yang et al., 2013). The exact functional role of L1 remains unclear. Figure 1-2 summarizes the efferent and afferent projections within layers 1-6 of V1.

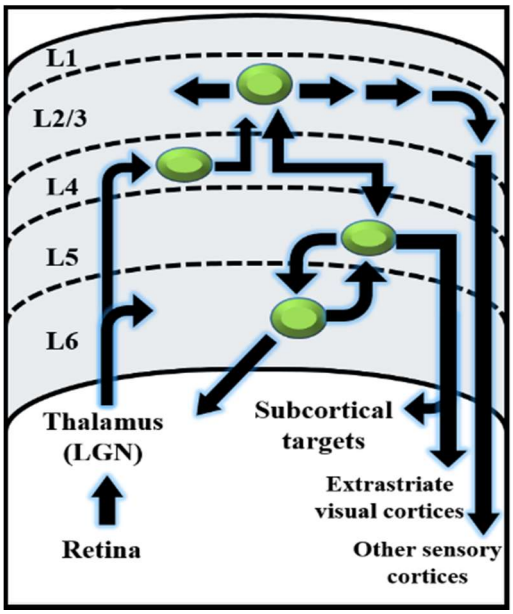

Figure 1-2. Illustration of the simplified laminar organization of the mouse V1. *V1 is composed of six layers, corresponding to distinct densities of neuronal cell bodies, dendrites, and axons. Neurons in the infragranular layers (L5 and L6) send their axons to subcortical structures such as the thalamus as well as extrastriate visual areas, the tectum, hindbrain, and spinal cord. In contrast, neurons located in the supragranular layers (L2/3) target intra- and intercortical regions. L4 (the granular layer) serves as the input layer of V1 and also targets neurons in intercortical areas (From Jabaudon, 2017).*

Pioneering studies by Hubel and Wiesel in the early and late '60s uncovered two main characteristics of V1 responses to a visual stimulus, namely, their high selectivity for stimulus orientation and direction. In a study conducted in anesthetized cats, they performed extracellular recordings while presenting oriented slits of light on a screen facing the cat in the neuron's receptive field (Fig 1-3). They observed that some orientations of the bar would elicit robust action potential discharge from the recorded neurons. However, if they adjusted the orientation of the slit of light away (e.g., > 40^{o}) from this 'preferred orientation' then spiking would reduce drastically (Hubel and Wiesel, 1962).

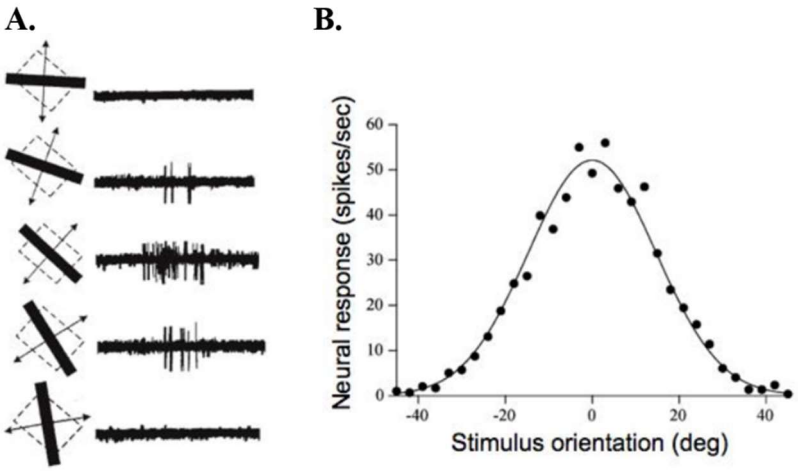

Figure 1-3. Spike trains of a V1 neuron during visual stimulation of various oriented bars. ***A.*** *(Left) A rectangular slit of light (width = 1/8 degrees), oriented at different angles to the receptive field axis, was presented to the contralateral eye of an anesthetized cat. The light stimulus was steadily moved bi-directionally in each neuron's receptive field at 1^{o}/sec while simultaneously performing extracellular recordings in V1. (Right) The spike train of a single neuron during stimulus presentation.* ***B.*** *An example response plot of the neuron on the left to its preferred orientation (i.e., set to the 0^{o} orientation) along with increasing/decreasing stimulus angles from the preferred orientation. The neuron fires the maximum spikes/sec for an oriented bar that matches its' preferred orientation. (Fig 3A from Hubel and Wiesel, 1962; Fig 3B from Khoei, 2014).*

In another study published in 1968, Hubel and Wiesel performed extracellular recordings of V1 neurons in lightly anesthetized macaque and spider monkeys while presenting various orientations of a black bar moving in the two orthogonal directions (Fig 1-4). They found that in some V1 neurons, the response to the preferred orientation was larger when the bar was moving in one direction compared to the other (Fig 1-4D; Hubel and Wiesel, 1968).

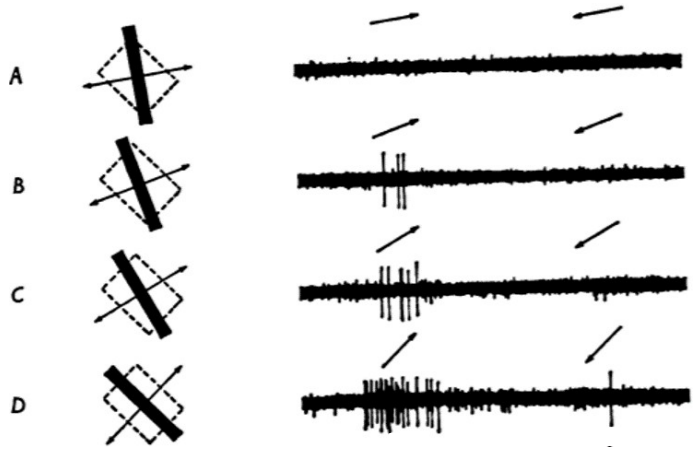

Figure 1-4. Action potential discharge of a neuron in the macaque monkey V1 during the presentation of moving bars. ***A-D.*** *(Left) bars of various orientations moving in both directions (arrows) were presented to the contralateral eye in a lightly anesthetized macaque monkey. (Right) Action potential discharge profile of an L4 neuron in V1. Action potential discharge was scarce or non-existent at most orientations (A-C).* ***D.*** *When the orientation of the stimulus matched the neuron's preferred orientation, the L4 neuron fired preferentially when the bar was moving in one direction (left to right) compared to the other direction (from Hubel and Wiesel, 1968).*

While most V1 neurons show some degree of orientation selectivity, only about 25-35% of V1 neurons are strongly directionally selective (De Valois et al., 1982; Schiller et al., 1976). In addition to orientation and direction selectivity, V1 neurons are also selective for changes in contrast (Dai and Wang, 2012), the spatial

frequency (Foster et al., 1985), and the temporal frequency (Movshon et al., 1978) of a visual stimulus. Taken together, the functional role of V1 is analogous to a 'simple filter' whereby edges and contours within a visual scene are enhanced and transmitted to extrastriate visual areas.

1.1.3 V1 to extrastriate visual cortices: the 'Where' and 'What' visual pathways

Once visual information is processed in V1, it is sent to extrastriate visual areas. Excitatory projections from the supragranular layers of V1 (i.e., L2/3) target extrastriate cortical regions (e.g., V2, V3, V4, MT), through two visual pathways called the dorsal stream and the ventral stream. In 1982, Mishkin and Ungerleider were the first to propose the existence of the two visual streams. They postulated that the visual information propagated along the dorsal stream would process information about an object's location in space (the 'Where' pathway; Fig 1-5). On the other hand, visual information processed through the ventral stream would give rise to object recognition (the 'What' pathway; Fig 1-5). Later studies by Goodale and Milner (Goodale et al., 1994; Goodale and Milner, 1992; Milner and Goodale, 2008) characterized the two streams based on their functional significance. They hypothesized that the ventral stream is involved in the perception of information about objects, while the dorsal stream processes information to guide actions.

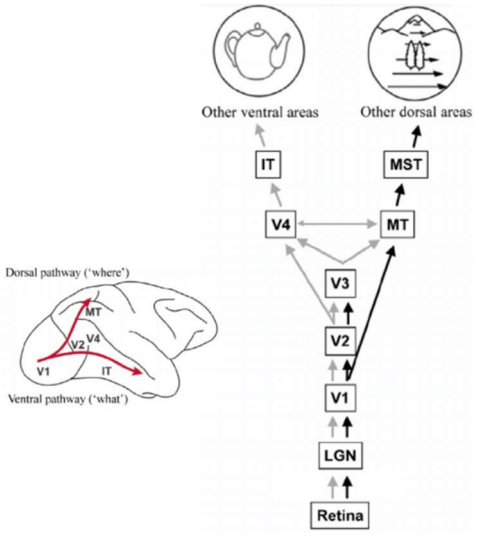

Figure 1-5. Schematic representation of the visual system organization in the macaque monkey. *The ventral pathway consists of the propagation of visual information from V1, V2, V3, and V4 before synapsing in the inferior temporal lobe (IT). Information propagated within the dorsal stream make stops at V1, the medial temporal area (MT; also known as V5) before synapsing in the medial superior temporal area (MST) of the parietal cortex. Information processing in the dorsal and ventral streams give rise to the 'where' and 'what' of objection perception, respectively (From Medathati et al., 2016).*

1.1.4 The auditory system: from the ear to the cerebral cortex

When a sound wave meets the ear, it moves through the auditory canal to the tympanic membrane (also known as the eardrum), a thin section of tissue that separates the outer ear from the middle ear (Roederer, 2008). The sound waves make the tympanic membrane vibrate, and this vibration is relayed by three bones (malleus, incus, and stapes, collectively known as the auditory ossicles) that amplify the sound vibrations before transmitting the signal to the cochlea, a fluid-filled spiral-shaped organ in the inner ear. Sound vibrations cause ripples in this fluid, which in turn, activates the mechanosensing organelles of hair cells within the cochlea, known as stereocilia (O'Malley and Clarke, 1961). Hair cells are responsible for transducing the sound information into electrical impulses, which travel through the auditory and cochlear nerves to the brain. These signals are sent to the cochlear nuclei before ascending to the inferior colliculus in the midbrain and the medial geniculate body (MGB) of the thalamus (Bartlett, 2013). Finally, thalamocortical projections terminate

in A1 (Brodmann's areas 41 and 42), located in the superior temporal gyrus of the temporal lobe (Cappe et al., 2009; Mitani et al., 1984; Smith et al., 2012).

1.1.5 A1: Connectivity and function

The auditory cortex of primates consist of three central regions: the core, the belt, and the parabelt. The core auditory region, which includes A1, receives innervations from the MGB of the thalamus and projects to the lateral belt areas (Kaas et al., 1999). In humans, A1 is located within Heschl's gyrus, which is analogous to core regions described in non-human primates.

The role of A1 is to decode the basic properties or features in sounds. A1 is organized tonotopically (i.e., neurons with similar frequency encoding are co-localized; higher frequencies are localized medially and caudally, while more lateral and rostral areas represent the lower frequencies; Fig 1-6).

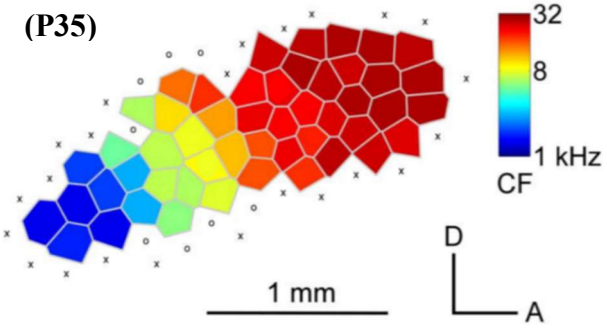

Figure 1-6. Tonotopic organization of A1 in a rat. *Illustration of the frequency inputs into A1 of a rat at postnatal day 35. The colored polygons indicate the characteristic frequency (CF) for neurons recorded in that region (From Zhou and Merzenich, 2008).*

While humans (Langers and Dijk, 2012), monkeys (Bendor and Wang, 2008), and rodents (Eggermont, 2013; Tsukano et al., 2017) all have a tonotopic organization in V1, each species exhibit different audible frequency sensitivities. For example, humans can detect sounds ranging from 20 Hz to ~ 20 kHz (Reynolds et al., 2010).

Among different non-human primates, the audible spectrum ranges from as low as 16 Hz in the Japanese macaque, to as much as 58 kHz in the Ring-tailed lemur (Heffner, 2004). However, in rodents, the audible frequency ranges from 1 kHz up to 100 kHz (Reynolds et al., 2010).

A1 sends auditory signals to the belt regions (secondary areas) where more complex feature extraction processes occur, such as spectral and temporal information extraction. Interestingly, it has been postulated that similarly to the visual cortex, auditory processing may also be organized into a 'what' and 'where' pathways, (Kaas and Hackett, 1999).

1.2 Perceptual effects of audiovisual interactions

As mentioned earlier, most objects in our environment are characterized by an array of physical properties detected by more than one sensory system. Integration of the information across sensory systems can, therefore, confer advantages towards generating an accurate, complete, and unified percept. Multisensory integration is particularly true for the visual and auditory systems that share overlapping properties. Indeed, both sensory systems are involved with detecting and localizing stimuli away from the body and provide high spatial and temporal information (Bulkin and Groh, 2006). Moreover, vision and audition usually convey complementary information about an object. Therefore, combining information provided by the two systems is likely to enhance the accuracy of the resulting percept. We will see below that the accuracy of the percept, resulting from audiovisual integration, is contingent on the congruency of the visual and auditory inputs.

1.2.1 Audiovisual integration of congruent stimuli

Several studies have revealed that binding complementary (i.e., congruent) or neutral visual and auditory inputs, facilitate and enrich numerous behavioral and perceptual processes. In particular, the detection (Gleiss and Kayser, 2014; Lippert et al., 2007) and discrimination thresholds (Vroomen and de Gelder, 2000) of visual stimuli are improved in the audiovisual context, while the reaction times decrease for visual object perception (Gielen et al., 1983; Hershenson, 1962; Posner et al., 1976). For example, human participants report near-threshold light flashes as brighter when presented simultaneously with white noise burst (Stein et al., 1996). Similarly, the audiovisual context was also found to enhance sound perception. When pairing concurrent light flashes with audible white noise, participants reliably described the noise as being louder in the audiovisual condition compared to the auditory-only condition (Fig 1-7; Odgaard et al., 2004). This cross-modal enhancement in loudness detection was replicated across numerous decibel levels and probabilities of concurrent light presentations.

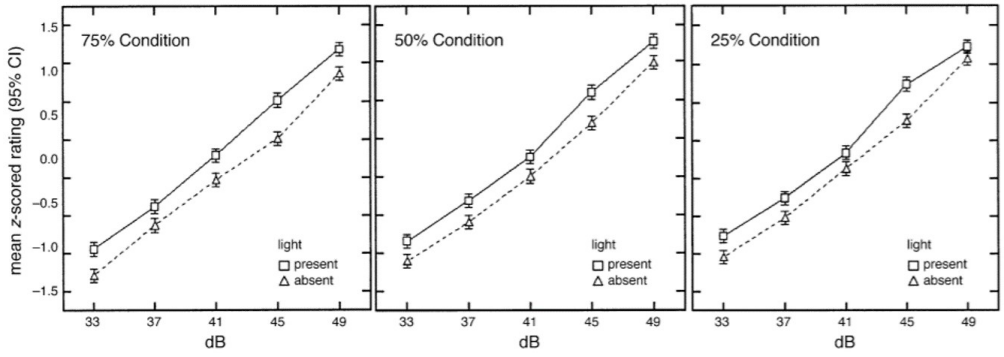

Figure 1-7. Audiovisual enhancement in loudness detection. *Average z-scores of intensity (loudness) ratings of white noise burst at five various dB levels, with (squares) and without (triangles) concurrent light, at three different probabilities of presentation of the light (75% vs. 50% vs. 25%). All participants showed cross-modal enhancement in loudness detection (From Odgaard et al., 2004).*

Studies in monkeys also found audiovisual enhancement of behavioral responses. For example, Lanz and colleagues demonstrated in 2013 that macaque monkeys engaged in a detection sensory-motor task, where visual and auditory stimuli were either presented alone or simultaneously, exhibited a shorter reaction time in the multimodal context, and the number of correct responses increased in response to audiovisual stimuli. The authors proposed that the improvement of the behavioral outcome was due to a reduction in signal ambiguity and an increase in the signal-to-noise ratio in the audiovisual context.

Studies conducted in rats (Tees, 1999; Komura et al., 2005; Hirokawa et al., 2008, 2011) also demonstrated strong links between multisensory integration and improved behavioral outcomes such as faster reaction times and better discrimination thresholds. Mice trained to respond to the presentation of unimodal visual or a unimodal auditory stimulus, by initiating nose pokes for a food reward (Fig 1-8A), made fewer mistakes when the visual and auditory cues were combined (Fig 1-8B; Siemann et al., 2014).

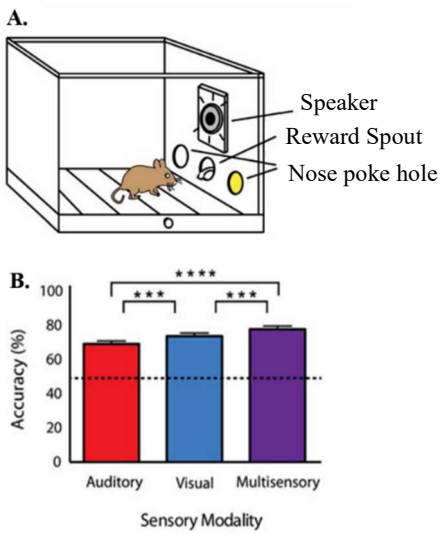

Figure 1-8. The effects of audiovisual integration on stimulus detection in mice. ***A.*** *Depiction of the operant chamber and cue presentation paradigm. Mice responded with nose pokes to the left or right depending on the location of the light stimulus in the visual-only condition or the presentation of white noise (left nose poke) or an 8 kHz pure tone (right nose poke), in the auditory-only condition. During multisensory trials, only congruent pairs of audiovisual stimuli were presented.* ***B.*** *Accuracy of the mice' performance was measured as the number of correct nose pokes/total number of trials presented in each stimulus context (From Siemann et al., 2014).*

1.2.2 Audiovisual integration of incongruent stimuli

Psychophysical research in humans and monkeys have demonstrated a persistent interest in the perceptual experiences produced when incongruent visual and auditory inputs interact in the brain. These studies have established that, in certain conditions, the interactions between sensory inputs can forge illusory percepts such as when visual and auditory inputs are mismatched (i.e., correspond to different known objects). The most well-known example of this phenomenon in humans is the McGurk effect. The McGurk effect is an audiovisual illusion in which the pairing of a spoken audible sound (e.g. 'ba') along with a visual perception of another sound (e.g. 'ga'), gives rise to a unified percept of a third sound (e.g. 'da'; McGurk and MacDonald, 1976).

The McGurk effect suggests that the goal of multimodal integration is not only to associate properties, detected by different senses, to the same object (e.g., a speaker and its voice), but also to reconcile the conflicts existing between various unimodal sources when the information is incomplete or altered. The current hypothesis is that this auditory illusion (discrepant visual inputs changes the auditory percept of a sound) is the result of the production of a 'compromise' between each unimodal component. Interestingly, the McGurk effect still exists even when subjects are aware of the illusion and can also lead to a 'recalibration' of speech perception (Lüttke et al., 2018), suggesting that multimodal interactions are robust and resilient to top-down influences such as expectation, learning, and memory. Another illusion observed in humans, the sound-induced flash illusion (Shams et al. 2000), similarly highlights how the brain integrates incongruent audiovisual sensory inputs. In this visual illusion, when

multiple auditory beeps accompany a single flash of a white circle against a black background, the single flash is incorrectly perceived as multiple flashes. Once again, audiovisual inputs are integrated to produce an illusory percept that reflects a compromise between the information provided by both modalities. While the McGurk effect illustrates that vision can alter auditory perception, the sound-induced flash illusion demonstrates that sound can also qualitatively alter visual perception.

In addition to the psychophysical studies in humans, studies in monkeys have also shown perceptual alterations or illusory-like experiences during audiovisual integration of incongruent sensory stimuli. For example, one study by Woods and Recanzone (2004) demonstrated that monkeys are susceptible to a well-known human illusion, known as the ventriloquism effect. The ventriloquism effect occurs when human subjects experience spatially disparate visual and auditory stimuli but perceive the sound stimulus as originating from the location of the visual stimulus. To determine if monkeys also experience the ventriloquism effect, two macaque monkeys were pre-trained to saccade to either the left or right side of a midline depending on which side the sound originated. During the task (post-training sessions), the monkeys received spatially identical or spatially disparate auditory and visual stimuli. Indeed, the authors found that the monkeys experienced a shift in their auditory spatial perception towards the direction of the visual stimulus when the audiovisual cues were spatially disparate (Fig 1-9). This study, along with the McGurk effect, illustrates that both primates and humans are subject to audiovisual illusions when the visual and auditory stimuli are spatially incongruent.

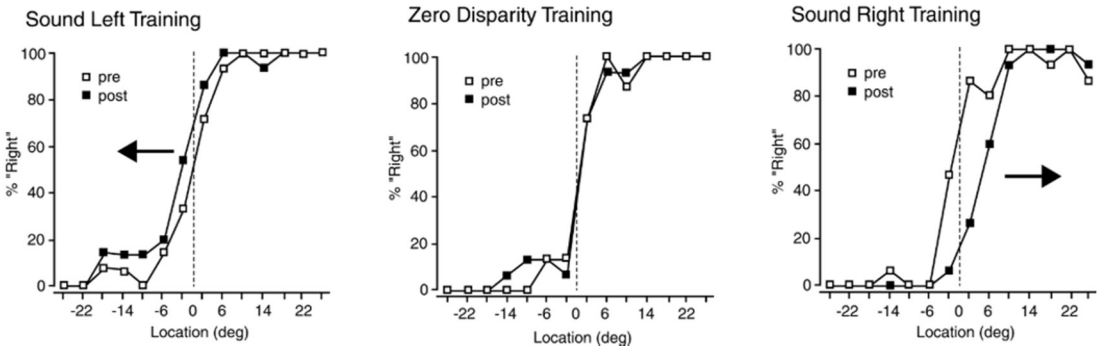

Figure 1-9. The Ventriloquism effect in macaque monkeys. ***(Left panel)*** *Psychometric curves from the sound left training experiment during pre-training (open square) and post training (filled square). This monkey had a clear shift to the left of the pre-trained psychometric function (pre-training results). This shift was due to a bias to respond "right" near the midline (vertical dotted line) when the sound was presented to the left of the visual stimulus.* ***(Middle panel)*** *However, during the experiment where there was zero spatial disparity between the audiovisual cues, there was no systematic shift in the psychometric function.* ***(Right panel)*** *Lastly, when a sound right training was implemented, the monkey showed a clear bias to respond "left" near the midline, resulting in a right shift in the psychometric function. For both the sound left and right training, the bias was in the direction of the visual stimulus during the training paradigm (From Woods and Recanzone, 2004).*

1.3 Multisensory integration: convergence of unimodal inputs

As demonstrated above, audiovisual integration enables the brain to identify images and sounds that correspond to one another. While the mechanisms by which audiovisual information gets integrated remain poorly understood, several high-order multisensory areas have been identified and were shown to be locations where multisensory integration occurs after unimodal processing.

1.3.1 Late integration model of multisensory interactions

The first models of multisensory integration were based on a modular conception of perceptual processing. In particular, multisensory integration was thought to occur only in higher-order multisensory areas; once unisensory processing

had already taken place (Yau et al., 2015; Ghazanfar and Schroeder, 2006). This working hypothesis has since been called the late integration model by Bizley and colleagues (2016). The late integration model was initially supported by studies that failed to find early interactions between various unimodal inputs. For example, early anatomical studies performed in the macaque monkey and cats did not find direct connections between primary sensory cortices (Felleman and Van Essen, 1991; Scannell et al., 1995). In the late integration model, unimodal inputs in the brain (e.g., auditory and visual stimuli) are integrated independently through parallel pathways defined by their sensory systems. The unisensory pathways converge in dedicated multisensory regions of the cortex, known as association cortices (Fig 1-10). In those multisensory areas such as the temporal pole (TP), the superior temporal sulcus (STS), and the posterior parietal cortex (PPC), neurons respond to stimuli from multiple sensory modalities, including vision and audition.

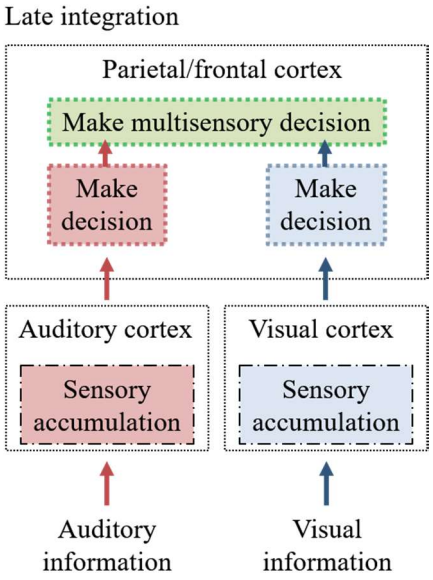

Figure 1-10. The late integration model of multisensory integration. *In this model, unisensory cortices, such as A1 and V1 process sensory information through segregated, non-interacting pathways. Next, unisensory information converges in 'higher-order' multisensory regions in the parietal and frontal cortices, giving rise to multisensory interactions. Multisensory integration in the associational cortices, in turn, gives rise to a multisensory decision, is communicated to motor areas in the brain (e.g., motor cortex), and mediates a behavioral response (From Bizley et al., 2016).*

1.3.2 Audiovisual integration in association cortices

The temporal pole (TP), a multisensory brain area located in the anterior temporal lobe, was identified as a hub of multisensory integration activity. In a study conducted by Ohki and colleagues (2016), the authors investigated the neural mechanisms underlying the binding of audiovisual cues in TP using magnetoencephalography (MEG) in healthy human volunteers, while performing an audiovisual speech matching task (Fig 1-11). In this task, temporally congruent and temporally incongruent visual cues (i.e., the appearance of a speaker mouthing words) were matched with auditory cues (i.e., speech). Participants had to determine if a pair of matching (congruent) or mismatching words (incongruent) appeared in the video they had watched (Fig 1-11A). The neural data showed the presence of delta oscillations in the TP during the task (Fig 1-11B), as well as delta band coherence between TP (seed region) and the occipital lobe, the posterior parietal cortex (PPC), and the prefrontal cortex (Fig 1-11C). They also demonstrated that TP activation correlated with an enhanced performance during an audiovisual speech detection task (Fig 1-11D).

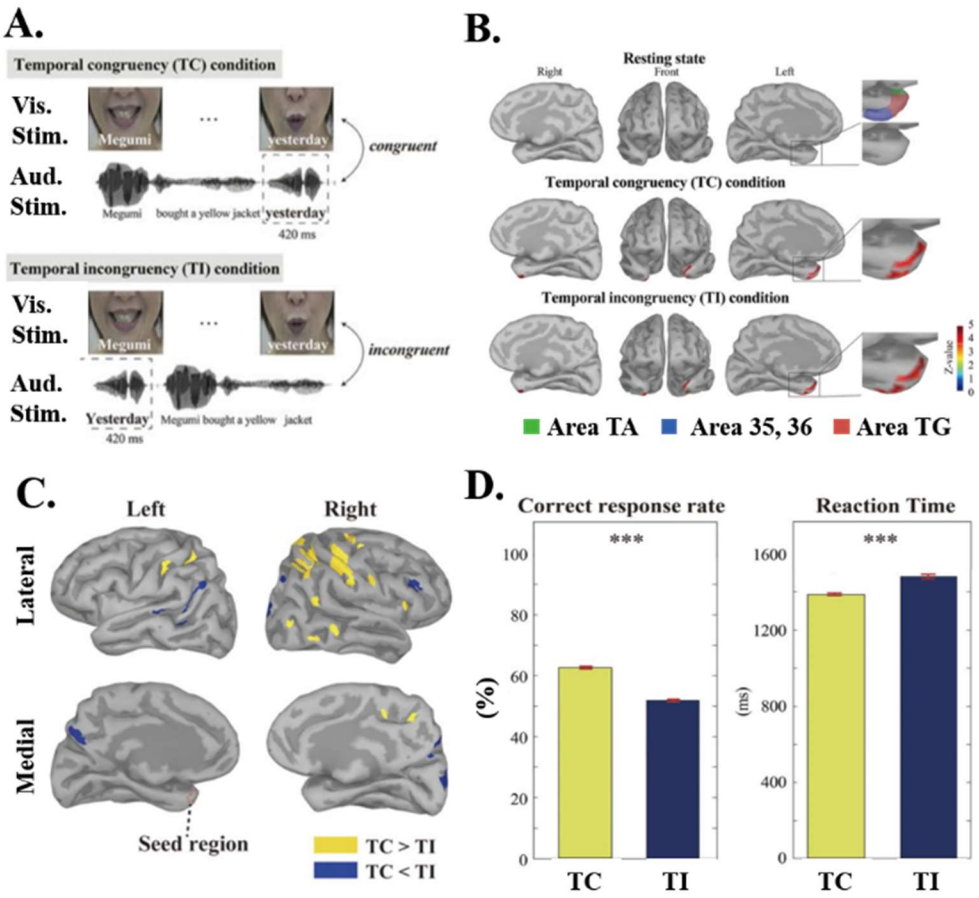

Figure 1-11. Temporal pole activation during audiovisual stimulation. ***A.*** *Illustration of the temporally congruent (TC) and temporally incongruent (TI) audiovisual conditions.* ***B.*** *Area TP was consistently activated by the multisensory signals regardless of stimuli congruency (TC vs. TI). Red-colored regions of the TP represent the highest power spectrum density during audiovisual stimulation. No activation was found in TP during the resting state condition. Bilateral TP consist of the following subregions: the somatosensory network (Area TA), the visual network (Area 35 and 36), and the semantic network (Area TG).* ***C.*** *The yellow clusters indicate coherence values that were significantly higher in the TC condition compared to the TI condition. The blue color depicts clusters considerably higher in the TI compared to the TC conditions. Significant TC clusters tended to overlap with the PPC, while significant TI clusters overlapped with the occipital lobe, STS, and the right ventral lateral prefrontal cortex* ***D.*** *In the left panel, mean correct response rate (%) on a speech detection task and the 95% confidence intervals (the red bar) are shown. In the right figure, mean reaction time (ms) and standard error (the red bar) are also shown for the detection task (From Ohki et al., 2016).*

In mice, the posterior parietal cortex (PTLp) was also recently identified as a critical area for audiovisual integration as well as auditory dominance over vision during audiovisual conflicts. In one study, scientist demonstrated the role of PTLp in mice trained to discriminate visual and auditory stimuli under a Go/NoGo protocol. During the task, the mice experienced visual-only or auditory-only trials with audiovisual trials interleaved (conflict trials). All trained mice showed auditory dominance during conflicting trials (i.e., mice based their behavior on the value of the auditory cue over the visual cue). This auditory dominance over vision required the PTLp. Indeed, PTLp inactivation with muscimol abolished this dominance (Song et al., 2017).

Furthermore, the superior temporal sulcus (STS) was identified to be a hub for audiovisual processing. Barraclough and colleagues (2005) recorded STS neurons *in vivo* in Rhesus monkeys during the presentation of unimodal visual stimuli, unimodal auditory stimuli, and audiovisual stimuli. These recordings revealed that STS neurons are modulated either positively or negatively in the audiovisual context compared to the unimodal stimulus presentations (Fig 1-12). The amplitude of the modulation of the visual response by auditory signals was substantial: visual responses were enhanced by 86% in neurons for which auditory signals had a positive effect, and attenuated by 46% for neurons suppressed in the audiovisual condition. Therefore, similarly to what is found in the superior colliculus (see below), STS neurons exhibit both response enhancement and response suppression during multisensory integration.

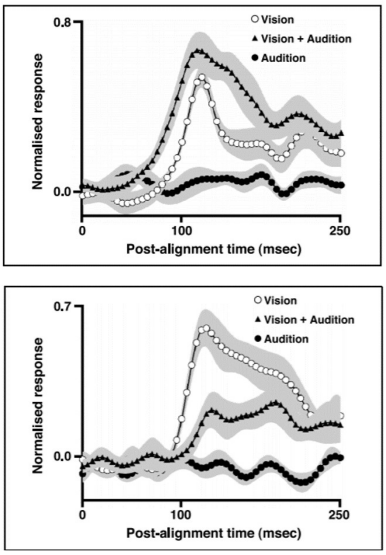

Figure 1-12. Effects of audiovisual stimuli on STS neuronal responses in Rhesus monkeys. ***Top panel)*** *Responses averaged across 14 STS neurons whose visual responses were enhanced in the audiovisual context. Solid triangles indicate responses to the audiovisual stimuli, whereas open and filled circles depict the neural responses to the unimodal visual and auditory stimuli, respectively. Each trace represents the spike density functions averaged for the recorded 14 neurons (gray cloud = SEM).* ***(Bottom panel)*** *Responses averaged across 14 STS neurons whose visual responses decreased in the audiovisual condition compared to the unimodal conditions. The markers indicating the stimulus shown during the task are identical to the top panel (From Barraclough et al., 2005).*

1.3.3 The principles of multisensory interactions

Interestingly, the region of the central nervous system best known for multisensory integration does not belong to the cortex. The superior colliculus is a midbrain structure that integrates visual, auditory, and other sensory inputs. Studies that investigated the effects of audiovisual interactions in the superior colliculus have demonstrated that sensory interactions can lead to a large panel of effects, from a simple summation, to an enhancement (the response to a cross-modal stimulus is considerably larger than the sum of the response to either of the components of the stimulus) or a degradation (the response to a cross-modal stimulus is significantly smaller than the sum of the response to either of the components of the stimulus; Stein et al., 2014). The type of response produced from audiovisual interactions depends on the spatial alignment (the spatial principle) and the temporal proximity (the temporal principle) of unimodal stimuli, as well as the effectiveness of unimodal components at evoking a neuronal response (the inverse effectiveness rule). Indeed, response

amplitudes of multisensory neurons to the auditory and visual stimuli are enhanced when the stimuli are presented simultaneously and at the same place, while auditory and visual stimuli presented at different locations or times attenuate or do not affect the responses of multisensory neurons. Furthermore, the extent to which a response is enhanced or suppressed is inversely related to the effectiveness of the individual component stimuli (Stein and Sanford, 2008; Stein et al., 2014).

1.4 Multisensory interactions in primary sensory areas

Although it is well established that higher-order brain areas such as the temporal lobe, the posterior parietal cortex, and the superior temporal sulcus all serve as multisensory areas where audiovisual inputs converge, recent evidence points to multisensory interactions also occurring during the early stages of sensory processing. In the following sections, highlighted studies, in non-human primates and mice, provide evidence for multisensory interactions in primary sensory areas.

1.4.1 Early integration model of multisensory interactions

Recently, several studies have provided both anatomical and functional evidence for audiovisual interactions in primary sensory areas (Cappe and Barone, 2005; Zingg et al., 2014; Giard and Peronnet, 1999; Brosch et al., 2005). Those new findings resulted in the proposition of an early integration model, defined by Bizley and colleagues (2016), positing that before the convergence of multisensory information in the higher-order brain areas, sensory interactions already occur as early as the primary sensory cortices (Fig 1-13). In this model, multisensory interactions at

the level of primary sensory areas contribute to the accumulation of sensory information before it is sent to the parietal and frontal cortices, where ultimately a decision is made (Bizley et al., 2016).

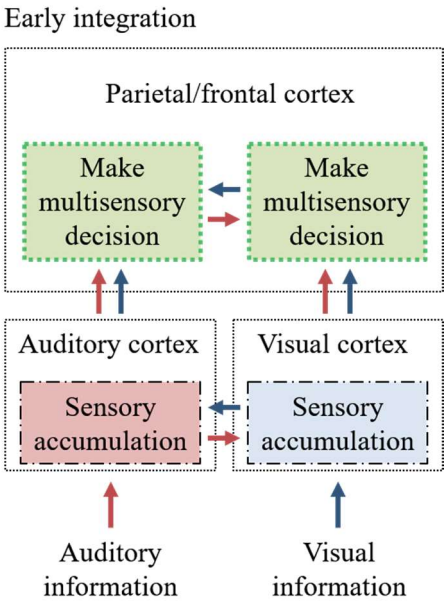

Figure 1-13. The early integration model of multisensory integration. *During early integration of sensory information across multiple modalities, such as vision and audition, multisensory interactions occur as early as the primary sensory cortices. Once multisensory interactions occur in unimodal brain areas, leading to evidence accumulated (i.e., the saliency and temporal congruency of the interacting sensory information), the multisensory signal is sent to higher-order brain regions where a decision is made, and behavioral response is initiated (From Bizley et al., 2016).*

1.4.2 Anatomical evidence for multimodal interactions in primary sensory cortices

Early anatomical studies in monkeys have uncovered direct projections between primary sensory areas. In one study in which retrograde tracer Dyamidino-Yellow was injected into the visual cortex of marmoset monkeys, the authors demonstrated the existence of projections in the auditory cortex, originating from primary and secondary visual areas (Cappe and Barone, 2005). In another study conducted in cynomolgus monkeys, the authors showed that the core (primary) and the surrounding parabelt regions (secondary areas) of the auditory cortex send direct projections to the primary and secondary visual areas. The authors established these

findings using retrograde fluorescent tracers Fast Blue and Dyamidino-Yellow, injected in the primary and secondary visual areas (Falchier et al., 2002).

Recently, anatomical studies in mice also revealed neuronal connections between unimodal sensory areas, further challenging the classical view that sensory processing within each modality is isolated and unaffected by the others. Direct connections between primary sensory areas can be found in the mouse cortical connectivity atlas created by Zingg and colleagues (2014). This exhaustive database, derived from a large-scale mouse connectome project, provides hundreds of labeled neural pathways from tracer injections applied across the entire mouse neocortex. According to both anterograde and retrograde tracer studies, many primary sensory cortices have reciprocal connections within the mouse cortex. In particular, the primary visual cortex both receives and sends cortico-cortical projections to the auditory cortex, thus providing anatomical support for audiovisual interactions between V1 and A1 (see Zingg et al., 2014 for review).

1.4.3. Functional evidence for multimodal interactions in primary sensory areas

In addition to the anatomical evidence for sensory interactions in primary sensory cortices, other studies have provided compelling functional evidence for direct multimodal interactions in primary sensory cortices. Such evidence includes the modulation of visually evoked event-related potentials (ERPs) in V1, by sound (Giard and Peronnet, 1999). In this study, ERPs were recorded from 30 scalp electrodes while human participants performed a forced-choice reaction-time categorization task. For each trial, the subjects identified one of two objects they were previously presented

with by pressing one of two keys. The two objects were defined by auditory features alone, visual features alone, or the combination of auditory and visual features. As expected, participants were more accurate and rapid at identifying multimodal than unimodal objects. Furthermore, analysis of ERPs and scalp current densities revealed audiovisual interactions in the visual, auditory, and frontal-temporal areas less than 200 msec post-stimulus. Activation of sensory-specific and the higher-order regions occurred in a temporally distinct manner with the frontal-temporal activation happening last in the sensory processing chain. These findings led the authors to conclude that audiovisual integration involved temporally and spatially distinct components that can occur early in primary sensory cortices and operate in both sensory-specific and nonspecific cortical structures (Giard and Peronnet, 1999).

Other evidence for functional multisensory interactions in primary sensory areas includes auditory cortical neurons that are modulated by visual and somatosensory stimuli (Brosch et al., 2005). In this study, multi-unit recordings were performed in the primary auditory cortex and posterior belt areas in monkeys while they performed an auditory categorization task. In that task, after a cue light came on, the monkeys could initiate a tone sequence by touching a bar and then earn a reward by releasing the bar only for occurrences of a falling frequency sweep. The electrophysiology data revealed acoustically responsive neurons in A1 whose firing was synchronized to both the cue light and to the touch or release of the bar (Fig 1-14). These findings suggested that A1 can be activated by visual and somatosensory stimulation and by movements. Furthermore, the authors speculated that multimodal activation of A1 arises from the training of the subjects on the behavioral task and that

multisensory integration facilitated the performance in the audio-motor tasks in proficient subjects (Brosch et al., 2005).

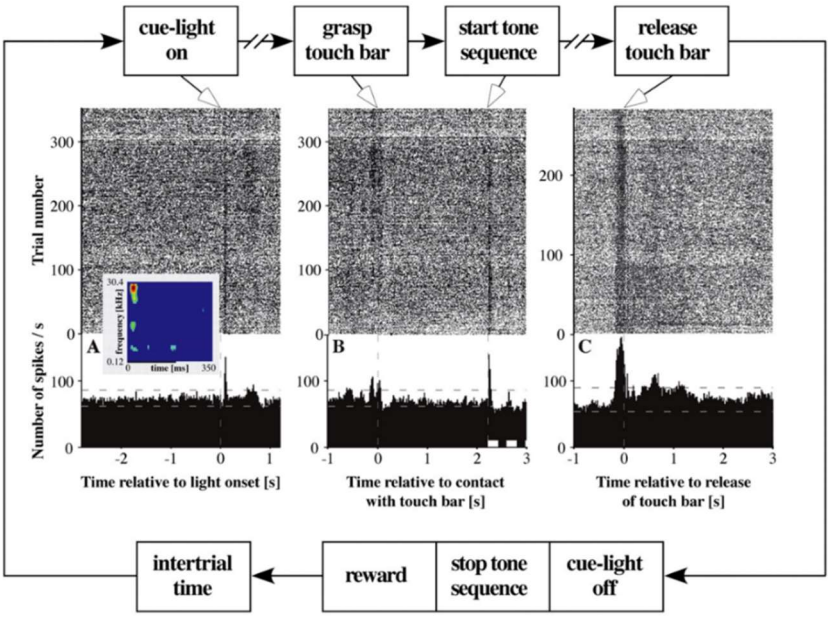

Figure 1-14. Multiunit activity in A1 during acoustic and non-acoustic events of an auditory categorization task. *The outer flow diagram depicts the clockwise sequence of the behavioral task, beginning with the cue-light. Inner panels depict the raster plots and the peristimulus time histograms computed from the neuronal spikes, which were temporally aligned to the onset of the cue light* ***(A)****, the bar touch and the start of the tone sequence* ***(B)****, and the release of the touch bar* ***(C)****. In panel* ***A****, the inset shows a response heat map for a single A1 neuron, computed from the spikes that, after the behavioral task, were recorded during the presentation of pure tones within the frequency range of tones indicated on the ordinate. The black bar denotes tone duration. The dark blue color indicates the average spike rate during the intertrial intervals. Spike rates that were significantly above this rate are plotted with warmer colors. In panel* ***B****, the white bars mark the first and the second tone in the sequence. Panel* ***C*** *was computed from trials with correct responses only. The dashed horizontal lines mark 3 SDs above and below the average baseline firing, calculated from the period of 1.8 sec before light onset (From Brosch et al., 2005).*

Taken together, both the human and the monkey study, detailed above, suggest that multisensory interactions during early-stage sensory processing are involved in multisensory processing.

1.5 Sound-induced modulation of V1 activity in mice

Investigating the effects of multisensory integration on neural activity in the primary sensory cortices, in mice, offer many advantages that include the possibility of using electrophysiological, neuroimaging, and genetic tools not available in human and monkey studies. Furthermore, since unimodal processing in V1 and A1 has been widely studied in different species, investigating audiovisual integration in V1 offers direct comparisons between human, monkey, and rodent experiments. Hence, several laboratories have investigated how multi-modal information is integrated within sensory cortices in mice, by recording the neuronal activity in V1 while presenting audiovisual stimuli (Iurilli et al., 2012; Ibrahim et al., 2016; Meijer et al., 2017). However, these studies yielded contradictory results. The following sections examine the main results from each study, along with their contradictory findings.

1.5.1 Sound-Driven Synaptic Inhibition in Primary Visual Cortex

In 2012, Iurilli and colleagues published a study in which they measured the modulatory effects of A1 activation (noise burst or optogenetic activation) on V1 activity in awake (when using LFP recordings) and lightly anesthetized mice (when using whole-cell recordings). They found that: [1] sounds (e.g., noise burst) hyperpolarized V1 L2/3 pyramidal neurons (Fig 1-15A), [2] sound-induced hyperpolarization (SH) of V1 L2/3 neurons required the activation of A1 (Fig 1-15B), [3] the activation of V1 neurons in the infragranular layers, specifically layer 5 neurons, mediated the sound-induced hyperpolarization observed in V1 L2/3 neurons (Fig 1-15C), and [4] Noise burst caused layer-specific responses (e.g. depolarization

or hyperpolarization) in V1 (Fig 1-15D). Overall, they showed that direct cortico-cortical A1 to V1 connections conveyed the sound-evoked activity in V1. Indeed, the transection of the white matter between A1 and V1 abolished sound-driven hyperpolarizations in V1 L2/3. They also demonstrated that higher sound intensities (48 dB SPL to 72 dB SPL) evoked larger hyperpolarizations in V1 until a saturating plateau was reached (~ 64 dB SPL). Finally, they found that sound-induced hyperpolarization in V1 was associated with behavioral suppression (e.g., reduced motor response) in mice conditioned to a visually-driven motor response. Given that the authors also showed that sound hyperpolarizes the neurons of the primary somatosensory cortex, they proposed that primary sensory cortices 'compete' for the activation of higher cortical areas by inhibiting each other.

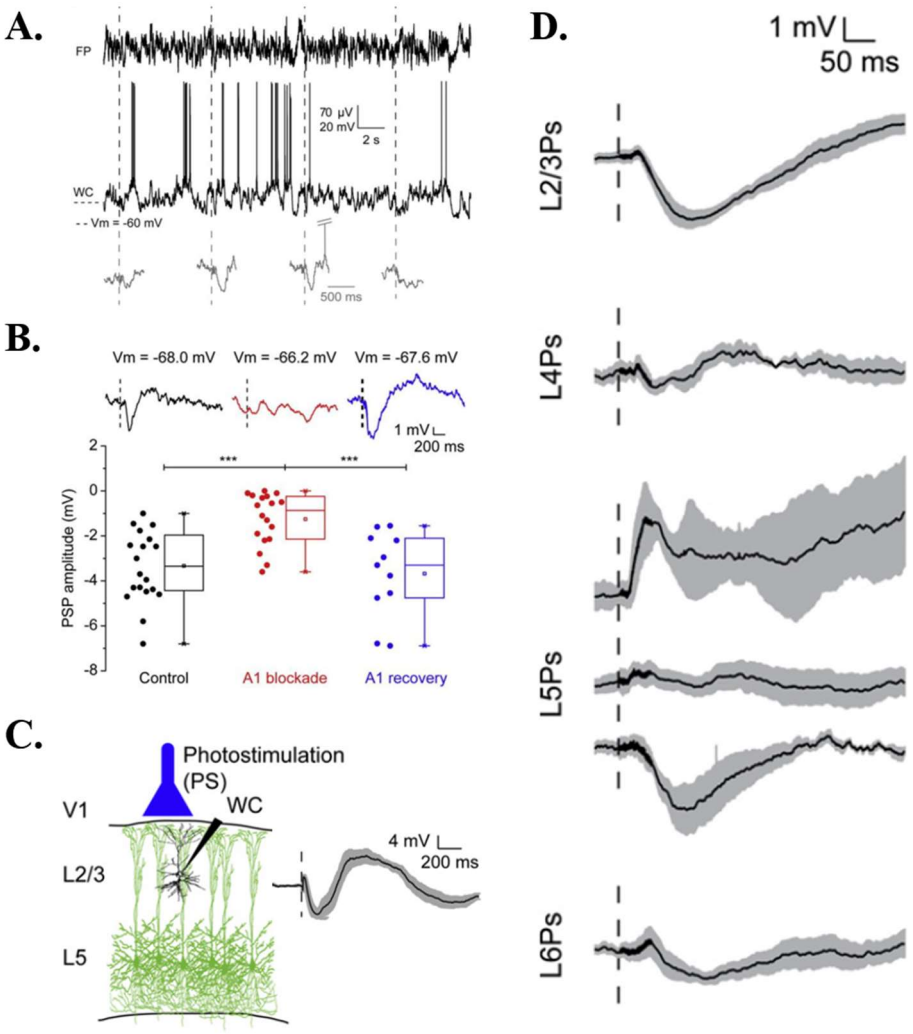

Figure 1-15. The effects of A1 activation on V1 activity. ***A.*** *Simultaneous field potential (FP; Top) and whole-cell (WC; Middle) recordings of the membrane potential (Vm) of L2/3 neurons in V1. (Bottom) Magnified FPs correspond to the sound-induced hyperpolarizations (SHs) recorded in V1.* ***B.*** *(Top) Example WC recordings of sound-driven Vm responses in V1 in controls (black), after A1 inactivation with muscimol (red), and after the functional recovery (i.e., 5 hr after muscimol application) of A1 (blue). (Bottom) Representative box plots (***$p < 0.001$ for post hoc test).* ***C.*** *(Left) Simultaneous optogenetic activation of L5 neurons and WC recording of L2/3 neurons in V1. (Right) Avg. response of V1 L2/3 neurons (black line) ± SEM (Gray cloud).* ***D.*** *Avg. Sub- and suprathreshold responses of layer-specific neurons to sound in V1 (From Iurilli et al., 2012).*

1.5.2 Cross-Modality Sharpening of Visual Cortical Processing through Layer-1-Mediated Inhibition and Disinhibition

A subsequent study refuted the hypothesis that sound modulation was solely suppressive in V1. Using loose-patch recordings and Ca^{2+} imaging in V1 (L2/3) of anesthetized and awake behaving mice, Ibrahim and colleagues (2016) measured the sound modulation of the visually evoked response of V1 neurons. Using CTb-Alexa, a retrograde tracer, in V1, they first established the existence of direct A1 to V1 connections. They found dense labeling throughout A1, specifically in A1 L5, targeting L1 neurons in V1 (Fig 1-16A). Their loose-patch and Ca^{2+} imaging recordings showed that the orientation selectivity of V1 L2/3 neurons, but not V1 L4 neurons, was sharpened in the presence of sound or optogenetic activation of A1 neurons (Fig 1-16B). Sound-induced sharpening of orientation selectivity of L2/3 neurons in V1 was due to an enhanced response of the neuron for its preferred orientation, together with a response suppression at non-preferred orientations (Fig 1-16B). This sharpening of orientation selectivity was more pronounced when the visual stimulus (grifting gratings of different orientations) was presented at low contrast (25%) than at high contrast (95%; Fig 1-16B), suggesting that sound modulation in V1 follows the principle of inverse effectiveness. The activation of L1 neurons in V1 reproduced the sound-induced sharpening of orientation selectivity in V1 (Fig 1-16C), suggesting that A1 connections to V1 L1 could mediate the orientation selectivity sharpening of L2/3 neurons, by sound. Taken together, Ibrahim's findings indicated that the sound-induced modulation of V1 neurons could improve the representation of the orientation of the visual stimulus, at the population level.

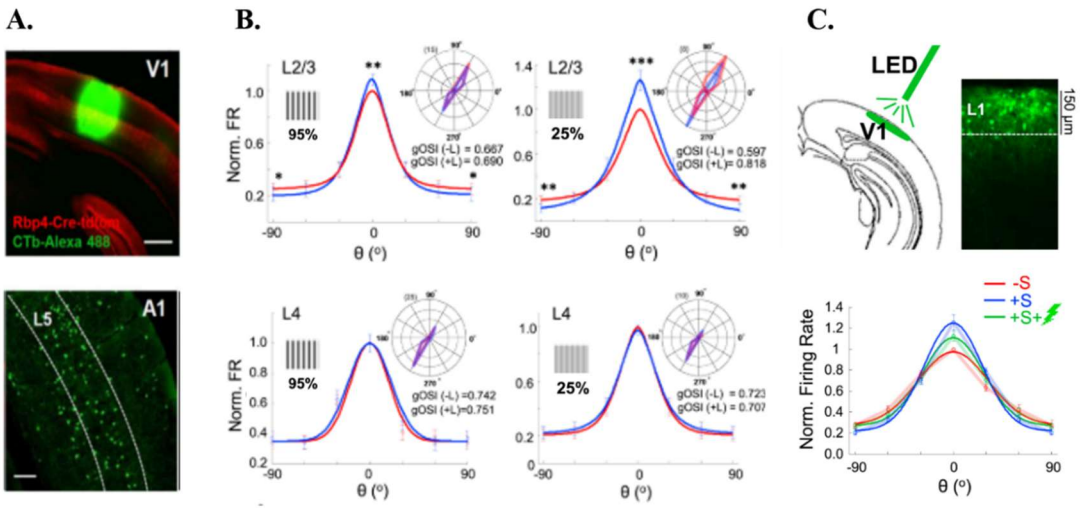

Figure 1-16. Sound-induced sharpening of orientation selectivity in V1 L2/3. ***A.*** *(Top) Retrograde labeling in V1 with CTb-Alexa 488 (green fluorescence) of an Rbp4-Cre:tdTomato (red) mouse that labeled all inhibitory neurons. (Bottom) Retrogradely labeled L5 neurons in A1.* ***B.*** *(Top) Avg. normalized evoked firing rate (FR) at different orientations (relative to the preferred orientation; Θ), presented at 95% contrast (Top Left) and 25% contrast (Top Right) for all recorded L2/3 neurons. (Inset) Polar graph plot indicating the global orientation selectivity index (gOSI) for L2/3 neurons during visual-only stimulation (-L) or visual stimulation plus optogenetic activation of A1 (+L). (Bottom) Same as (Top) for L4 neurons in V1.* ***C.*** *(Top left) Illustration of the experimental setup. Green LED light was shone on L1 neurons in V1 filled with ArchT-GFP to elicit light-induced inhibition. (Top right) expression pattern of ArchT-GFP in V1 L1. (Bottom) Avg. tuning curves (normalized evoked FR) for all recorded L2/3 neurons presented with a 25% contrast drifting grating in the visual-only (-S), visual plus sound (+S), and the visual + sound + LED stimulation (V+S + L) condition (From Ibrahim et al., 2016).*

1.5.3 Audiovisual Modulation in Mouse Primary Visual Cortex Depends on Cross-Modal Stimulus Configuration and Congruency

However, another study published the following year contradicted the results of Ibrahim and colleagues. Meijer and colleagues (2017) performed 2-photon calcium imaging in awake behaving mice during the presentation of audiovisual stimuli. They found that the features of the audiovisual stimuli, such as the temporal congruency, determined whether sounds enhanced or suppressed V1 visual responses. When an auditory tone was paired with a drifting grating, such that the auditory and visual stimuli were temporally congruent (e.g., both stimuli modulated at 1 Hz), neurons in V1 were divided between neurons enhanced and suppressed by sounds, with an equal prevalence of both (Fig 1-17A, B). However, when the audiovisual cues were temporally incongruent (visual grating = 1 Hz, auditory signal = noise burst), V1 responses in the audiovisual condition were dominated by sound-induced suppression (Fig 1-17C, D).

In addition, while visual contrast (100% or 25%) did not affect multisensory integration in V1 when the audiovisual stimulus pairs were congruent (Fig 1-17A, B), it did have a differential effect when the auditory input was temporally incongruent (e.g., noise burst) with the visual input. At 100% contrast, V1 neuronal responses were dominated by sound-induced suppression. On the contrary, when a 25% contrast visual stimulus was paired with an incongruent auditory signal, the prevalence of enhancement and suppression reached equilibrium (Fig 1-17E).

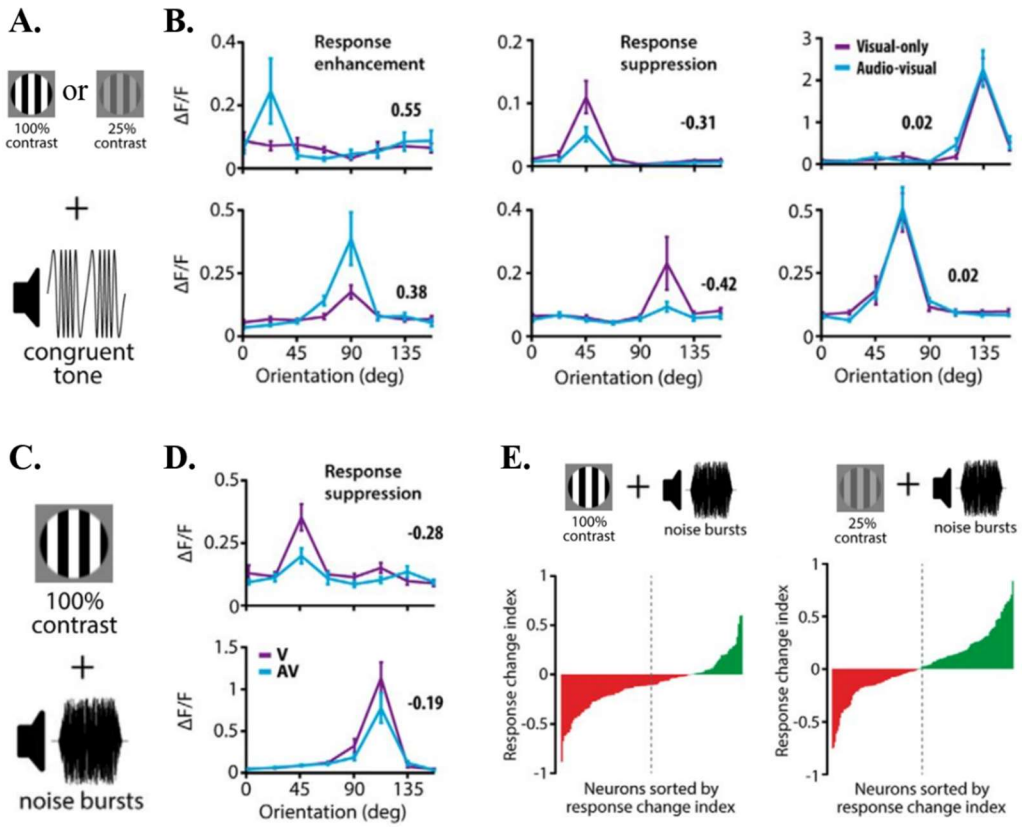

Figure 1-17. The impact of audiovisual stimulus features on multisensory integration in V1. ***A.*** *In this paradigm, either 100% or 25% contrast visual square-wave gratings of eight orientations were presented alone (V only; purple) or together with a temporally congruent tone (AV; cyan).* ***B.*** *The tuning curves of six example neurons showing sound-induced enhancement, suppression, or no difference compared to the V-only condition. The prevalence of enhancement and suppression was the same, regardless of visual contrast. The bold number represents the response change index (RCI). A positive RCI corresponds to a sound-induced response enhancement, while a negative RCI indicates a sound-induced response suppression* ***C.*** *In this paradigm, 100% visual square-wave gratings of eight orientations were paired with a temporally incongruent sound (noise burst).* ***D.*** *The tuning curves of two example neurons showing sound-induced response suppression.* ***E.*** *Histograms of sorted RCIs of all neurons in the 100% contrast + noise burst (Left) or the 25% contrast + noise burst (Right) conditions (From Meijer et al., 2017).*

In summary, the study by Meijer and colleagues suggested a more dynamic and versatile multisensory integration in V1. Specifically, the relationship between the auditory and visual stimuli (e.g., temporal congruency) dictated how V1 neurons

responded to and integrated the audiovisual information. When the unisensory inputs were temporally incongruent and accompanied by a full-contrast visual input, this resulted in a 'competitive' interaction between the multisensory inputs. This interaction was dominated by sound-driven suppression. However, when the unisensory inputs were temporally congruent, regardless of the contrast of the visual stimulus, this resulted in a heterogeneous response in L2/3 neurons.

1.5.4 A comparison of the key findings from Iurilli, Ibrahim, and Meijer

While each of the abovementioned studies provided concrete evidence for sound-induced modulation of V1 activity during visual processing, they widely differed in their conclusions on how sound modulates V1 responses. Iurilli and colleagues showed widespread sound-driven inhibition in V1 L2/3. In contrast, Ibrahim later showed that V1 neurons were primarily enhanced by sound, but only for their preferred orientation. Finally, Meijer found that V1 L2/3 neuronal responses were not strictly enhanced in the presence of sound but were also suppressed depending on the congruency of the audiovisual stimulus features.

Furthermore, all three studies provided distinct hypotheses for the mechanisms that mediated sound-induced modulation of V1 activity, during audiovisual stimulation. On the one hand, Iurilli and colleagues proposed that L5 pyramidal neurons in V1 were responsible for mediating the widespread sound-driven inhibitory effects on V1 L2/3 neurons. On the other hand, Ibrahim showed that L1 inhibitory neurons in V1 were prime candidates for the sound-driven sharpening of orientation selectivity of L2/3 neurons. Finally, Meijer's findings suggested that heterogeneous

multisensory interactions might have been mediated by “oscillatory phase-resetting”, whereby auditory inputs modified the oscillatory phase at which visual inputs arrived, therefore resetting ongoing rhythmic activity in several frequency bands. This hypothesis is supported by previous findings regarding oscillatory phase-resetting (Lakatos et al., 2008; Schroeder and Lakatos, 2009). According to this theory, the sound-induced enhancement and suppression found in Meijer’s study could be accounted for if auditory signals arrived precisely at the peak (temporally congruent) of a V1 oscillatory cycle. In turn, coincident auditory signals at the peak of the V1 oscillatory cycle would lead to enhanced V1 responses to the audiovisual stimuli. However, in the case where the auditory inputs were temporally incongruent with the visual inputs (i.e., arriving at the troughs of the oscillatory cycle), V1 response suppression would have ensued.

1.6 Goals of thesis

The studies conducted by Iurilli, Ibrahim, and Meijer provide inconsistent reports on the role of sound modulation on the visually-evoked responses of V1 neurons. Moreover, all three studies provide conflicting evidence for the mechanisms that mediate multisensory integration in V1. Due to these discrepancies, the impact of audiovisual integration on the encoding of visual information in V1, at the population level, remains unclear.

To determine the modulatory effect of sounds on V1 activity, we performed two-photon calcium imaging in the mouse V1, while simultaneously presenting auditory and visual cues. Rather than determining how sound modulates the visual

responses of individual V1 neurons, we used a population-level analysis approach. The advantage of this approach was the possibility of determining how V1 neurons collectively integrate audiovisual stimuli and subsequently transmit that information to 'higher-order' visual areas.

To that end, the goal of the first aim was to determine the modulatory effects of sound on visual processing in V1 L2/3 neurons. In particular, aim I addressed how the simultaneous presentation of one of two pure tones and one of two orthogonal drifting gratings affected the representation of orientation and direction of V1 neurons to the visual stimulus. In, the second aim I determined if the sound-induced modulation of evoked responses in V1 also impacted the performance of mice in a visually-guided behavioral task. This aim investigated whether multisensory integration confers behavioral advantages on a visual discrimination task, in which mice must discriminate between two oriented visual stimuli. Finally, the third aim was designed to determine the importance of the behavioral context on sound-induced modulation in V1. Aim III tested the hypothesis that sound modulation is flexible and can be dynamically adapted to more complex tasks through training. In the following sections, I describe each aim in detail, along with the accompanied experiments.

Chapter II: General methods

2.1 Animals

Mice were used as the animal model for all experimental procedures and testing. All procedures were approved by the Institutional Animal Care and Use Committee of Rutgers University, in compliance with the Guide for the Care and Use of Laboratory Animals (Department of Health and Human Services). In aim I, 8 week old GAD2-Cre mice (Stock no: 019022) were crossed with Ai9 reporter mice (Stock no: 007909) that expressed Td-tomato in GABAergic neurons. All mice used in aim I had a homozygous mutation and were bred in the animal facility at Rutgers University. In both aims II and III, wild type C57BL/6 mice were purchased from Jackson laboratories and bred in-house. Both male and female mice were used for all experiments.

2.2 Stereotaxic Surgery

2.2.1 Head bar implant

Head bars were implanted in all mice used in aims I, II, and III. This procedure was required to restrict head movements during *in vivo* calcium imaging in awake behaving mice. First, mice were anesthetized by placing them in a Plexiglas™ sealed chamber and exposing them to 5% isoflurane mixed with oxygen. Once the mice were anesthetized, as determined by: a non-response to a tail and paw pinch, deep and slow breathing, and immobility, the fur above the skull was shaved. Next, the mice were placed in a stereotaxic apparatus fitted with an anesthesia mask. The mask was placed

over their snout and the flow of isoflurane was reduced to a maintenance rate of 2%. To prevent deep anesthesia and excessive recovery times, isoflurane flow was progressively reduced to 1.25% during surgery. Periodically, the level of anesthesia was determined by monitoring the absence of: the tail and paw pinch reflex, whisker movement, and eye blinking.

Blunt ear bars were used to hold the head firmly in the stereotaxic apparatus and the angle of the head was adjusted to attain flat skull. Body temperature was maintained at 37-38^{o} Celsius with a homeothermic heating pad. Desiccation of the eyes were avoided by applying artificial tear ointment. Carprofen [5mg/kg], a nonsteroidal anti-inflammatory drug, was administered subcutaneously prior to any incisions. To minimize infection at the incision site immediately above the skull, iodine was liberally applied to the skin and any excess was removed with 75% ethanol. To further reduce the chances of post-operative infection, all tools were autoclaved prior to surgery and submerged in a hot bead sterilizer (Fine Science Tools), before coming in contact with the mice. Lastly, small injections of lidocaine were administered subcutaneously around the incision site above the skull as a local analgesic.

After mice were placed in the stereotaxic, the scalp was incised using autoclaved surgical scissors and the surface of the skull was cleaned of the fascia using sterile scalpel cotton swabs. The trapezius muscle was carefully separated and pushed back from the interparietal and occipital bones using a sterile scalpel. This step was especially important to prevent the development of aseptic inflammation involving bone loss and eventually brain damage. Next, small incisions were made on the

parietal and interparietal bones. These small grooves on the bone were made to increase the strength of the bond between the skull and the glue holding the head bar.

Once these small incisions were made, a thin layer of cyanoacrylate glue was spread along the skull surface where the head bar would be implanted. A light, custom made, stainless steel, and omega shaped plate was lightly pressed against the skull until a bond was achieved (~1 min). The head bar was carefully placed ~5 mm anterior from the craniotomy and the area marked for the coverslip, to ensure the objective did not damage the head bars during imaging. The head bar was then reinforced with dental acrylic. In addition, a ring made of dental acrylic was placed along the edges of the skull and the head bar, which created a small cup around the surface of the skull. This was necessary for subsequent procedures such as virus injections, coverslip implants, and calcium imaging, which required the skull to be submerged in artificial cerebral spinal fluid (i.e. cortex buffer) or water. The dental acrylic was left to dry and fully harden for ~10 min.

The head bars remained in place for several months provided excessive pressure was not applied to it. In the rare occasion when the head bar came loose or fell off, that animal was immediately euthanized by $C0_2$ asphyxiation and cervical dislocation. All mice habituated quickly to the head bars and typically resumed normal behavior within 48 hours post-surgery.

2.2.2 Adeno-associated virus injections

The virus injections were performed on mice used in aims I, II, and III. To investigate the changes in activity that occurred in V1 neurons during multisensory

integration, genetically encoded calcium sensor GCAMP6f was injected in V1. These genetically encoded calcium indicators was delivered to cells via the adeno-associated virus (AAV). In order to express GCAMP6f in the primary visual cortex, a craniotomy was made at the following coordinates AP: -4.48 mm, ML: +2.50 mm, DV: -3.00 mm from dura. Once the bone above V1 was removed, 2 µl of viral stock solution AAV1.Syn.GCaMP6f.WPRE.SV40 (Penn Vector Core) was injected at a rate of 35 nl per minute using a glass pipette (100-micron diameter tip) connected to a nanoliter injector.

After the virus was injected, the craniotomy was closed using a glass coverslip. For 5 days following surgery, mice were administered antibiotics (amoxicillin, 0.25 mg/ml) in their drinking water. Mice typically returned to normal behavior within 12 hours after surgery. To ensure there was sufficient GCAMP6f expression in V1 neurons prior to the imaging experiments, fluorescent expression levels were measured using 2-photon microscopy-based calcium imaging 2 to 3 weeks after virus injections.

2.2.3 Coverslip implant surgery

This procedure was performed on all mice that participated in aims I, II, and III. Two-photon microscopy required unimpeded access to the cortical surface in order to record neural activity. To accomplish this a 3 mm diameter circular coverslip was placed in the craniotomy after the head bar was implanted. The coverslip served as a transparent ‘window’ for imaging *in vivo* neuronal activity in cortex.

For this procedure a 3.1 mm craniotomy was made above V1. The bone was drilled until a thin layer of bone (~50 microns thick) remained above dura. Next, bone debris was carefully removed using sterile cortex buffer (135 mM NaCl, 5 mM KCl, 5 mM HEPES, 1.8 mM CaCl2 and 1 mM MgCl2, pH 7.4). Once the craniotomy site was cleared of debris, the remaining bone layer above cortex was left to soak in cortex buffer for 20 minutes. This softened the bone such that it reduced the chances of inadvertently breaking the dura, prevented dural hemorrhages and edema. After 20 minutes, the last layer of bone was removed using sterile forceps under cortex buffer immersion. To prevent cortical desiccation, the craniotomy was constantly covered with cortex buffer or Gelfoam™ (sterile compressed sponge) soaked in cortex buffer.

Finally, a 3 mm coverslip was placed in the craniotomy and glued to the surrounding bone using cyanoacrylate adhesive. The skull and the rim of the coverslip along with the head bar implant were covered and held in place with dental acrylic. As mentioned above, mice were allowed to recover for 5 days following surgery. However, normal behavior usually resumed within 48 hours of surgery.

2.3 *In vivo* calcium imaging

Whenever a neuron fires an action potential intracellular and synaptic calcium levels increase (Borst and Sakmann, 1998). This property of neuronal firing was exploited using a genetically-expressed calcium sensor GCAMP6f in order to visualize the activity of neurons with high spatial and temporal resolution. For this procedure, *in vivo* calcium imaging was performed in mice injected with GCAMP6f in V1 and set aside for recovery.

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 two-dimensional movement (Dombeck et al. 2007). During three daily 20-min sessions, the mice were acclimated to the spherical treadmill by fixing their head bar to a post, which held the mice on the apex of the spherical treadmill. Once there was sufficient neuronal expression of GCAMP6f in V1 (~ 2 to 3 weeks), during which mice in aims II and III were trained to perform on a Go/NoGo visual discrimination task or a Go/NoGo cross-modal discrimination task, imaging could begin.

On the day of the recording, mice were placed on a spherical treadmill under a two-photon microscope. Next, audiovisual stimuli were presented while the neurons in L2/3 of V1 were recorded. During a 2-photon scanning session the ball movement was tracked by an infrared (IR) camera taking pictures of the ball at 30 Hz. Eye motion was also monitored at 15 Hz with the use of a second IR camera that imaged the reflection of the eye from an infrared dichroic mirror. Functional imaging was performed at 15 frames/s using a resonant scanning two-photon microscope (Neurolabware) powered by a titanium-sapphire Ultra II laser (Coherent) set at 920 nm.

In addition, scanning mirrors in the Neurolabware microscope were located in a chamber hermetically sealed, bringing the scanning hum below the room ambient noise [< 68 A-weighted decibels (dBA)]. The laser beam was focused 200 μm below the cortical surface using a 16x magnification, 0.8-NA Nikon water-immersion objective. The objective was tilted 30° such that the objective lens was parallel to the

dura surface. Laser power was kept below 70 mW. Frames (512 x 796 pixels) were acquired using a software Scanbox developed by Neurolabware.

Two-photon microscopy-based calcium imaging was a non-invasive procedure and therefore did not affect the animal's performance in the behavioral task (aims II and III) or during passive viewing of audiovisual stimuli (aim I). At the end of the recording session, mice were placed back in their cages and multiple imaging sessions were performed over multiple days without any adverse health effects.

Chapter III

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

McClure Jr. and Polack, 2019

3.2 Background

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. Whereas 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. 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 (Ibrahim et al., 2016; Iurilli et al., 2012; Meijer et al., 2017). However, these studies yielded contradictory results.

In one study, whole cell recordings performed in anesthetized mice showed that short 50-ms 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 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 layer 2/3 (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 had so far never been assessed. In the present study, 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.

3.3 Hypothesis

Previous studies have shown that sound modulates specific response properties of V1 neurons, namely individual neurons' orientation tuning properties (Iurilli et al., 2012; Ibrahim et al., 2016, Meijer et al., 2017). However, the outcome of sound modulation on V1 population responses remained unexplored. We hypothesized that sound differentially modulates populations of V1 neurons for oriented stimuli that matches their preferred orientation compared to orthogonal orientations. Furthermore,

we hypothesized that sound also modulates the responses of direction selective V1 neurons.

3.4 Materials and methods

In the following study, two-photon calcium imaging was performed in awake mice to compare the representation of orientation and direction of drifting gratings of V1 L2/3 neurons.

3.4.1 Animals

For details on the animals used in this experiment, see section 2.1 of Chapter II.

3.4.2 Stereotaxic surgery

For technical details on all surgical procedures performed for this aim including head bar implants, virus injections, and coverslip implants refer to section 2.2 of Chapter II.

3.4.3 *In vivo* calcium imaging

For technical details see section 2.3 of Chapter II.

3.4.4 Audiovisual stimuli

A gamma-corrected 40-cm diagonal liquid crystal display (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 (The 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/deg, contrast: 75%, duration: 3 s, intertrial interval: 3 s). Auditory stimuli consisted of the presentation of one of two sine-wave pure tones (10 kHz and 5 kHz, 78 dBA; duration: 3 s). The background noise (room heat and air conditioning airflow, laser cooling fans, and spherical treadmill airflow) was 69 dBA.

Each audiovisual trial resulted from the random combination of one of the two pure tones with one of the two drifting gratings (4 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; Fig. 19B). 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° and randomly permuted; Fig. 3-1D). 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/deg, contrast: 75%, duration: 1.5 s, intertrial interval: 3 s).

This experimental design limited comparisons across orientations for individual neurons but favored analyses at the population level by making possible the gathering of a larger data set of unimodal and audiovisual trials while still characterizing the orientation and direction tuning of the imaged neurons. Because scanning was not synced to the stimuli, 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. The photodiode 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) and recorded with the software WinEDR (John Dempster, University of Strathclyde). Imaging sessions started by recording 1,000 frames with the green and red channels. The red channel was used to exclude GABAergic neurons from analysis.

3.4.5 EEG electrode implants

Adult (2 - 3 months old) male and female C57BL/6 mice were implanted with head bars (see section 2.2.1). 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; model 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.

3.4.6 EEG recordings

EEG electrode signals were preamplified (model AD4177-4; Analog Devices) at the head of the animal, fed into a four-channel EEG amplifier (model 1700 differential AC amplifier; A-M Systems), filtered between 1.0 and 5 kHz, and digitized at 1 kHz (NiDAQ; National Instruments) along with signals allowing synchronization with visual stimuli, locomotion, and an eye tracking system.

3.4.7 Data analysis

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

Imaging data preprocessing. 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 semiautomatic 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- to 5 μm 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, defined as $dF/F = (F - F_0)/F_0$, was calculated with F_0 defined as the

median of the raw fluorescence measured during every intertrial interval. Trials were then sorted by stimulus.

The mean dF/F measured during a 1.5-s intertrial period immediately preceding each visual stimulation was subtracted from the dF/F measured during the trial. We then 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 makes 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 the possibility that constantly active neurons could be considered as responding neurons, the response amplitude was expressed as a 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-s presentation of the 12 different drifting gratings used to construct the tuning curve (Fig. 3-1D). 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, and 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 because 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 (RV and RAV, 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; values between 0 and -1 indicate a negative modulation or suppression of activity. When R_{AV} and R_{Vonly} are positive, an SMI of 0.1 corresponds to a response increase of 22% in the audiovisual context and an SMI of 0.2 corresponds to a response increase of 50% in the audiovisual context. Conversely, SMIs of -0.1 and -0.2 correspond to a suppression of 18% and 33%, respectively.

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 x 0.375) to satisfy Parseval's Theorem (where N represents the number of points of the electrocorticography 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.

3.4.8 Statistics

Statistical significance was calculated using ANOVAs or Kruskal-Wallis one-way ANOVAs on ranks in 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 Tukey's 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 α level 0.05 and were defined as the 2.5 and 97.5 percentiles 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 conditions with visual-only stimuli (V_{only}) and audiovisual stimuli at 10 kHz (AV_{10kHz}) and 5 kHz (AV_{5kHz}). We then 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} - Vonly; AV_{10kHz} - AV_{5kHz}). The difference between groups was considered significant if the values of the CI boundaries were of same sign.

3.5 Results

We imaged the activity of 3,355 neurons (22 recording sessions in 14 animals) located in 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 (Fig 3-1A). Calcium imaging was performed while the presentation of unimodal (visual only) and bimodal (audiovisual) blocks of 30 or more trials was alternated (Fig 3-1B).

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 s, temporal frequency: 2 Hz, spatial frequency: 0.04 cycle/deg, contrast: 75%; Fig 3-1B). Orientations were chosen orthogonal to activate two very distinct neuronal populations in V1 with minimal response overlap (Fig 3-1C). During audiovisual blocks, the test stimuli were paired with one of two pure tones (low pitch: 5 kHz, or high pitch: 10 kHz; Fig 3-1B, C). At the end of each recording session, we assessed the tuning for orientation and direction of all the imaged neurons by presenting a series of drifting gratings of 12 evenly spaced orientations (gratings of 6 distinct orientations traveling in 2 opposite directions; Fig 3-1D). 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/deg), 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 (Fig. 3-1C, 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). Because interneurons account solely for intracortical 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 (Fig. 3-1E, 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.

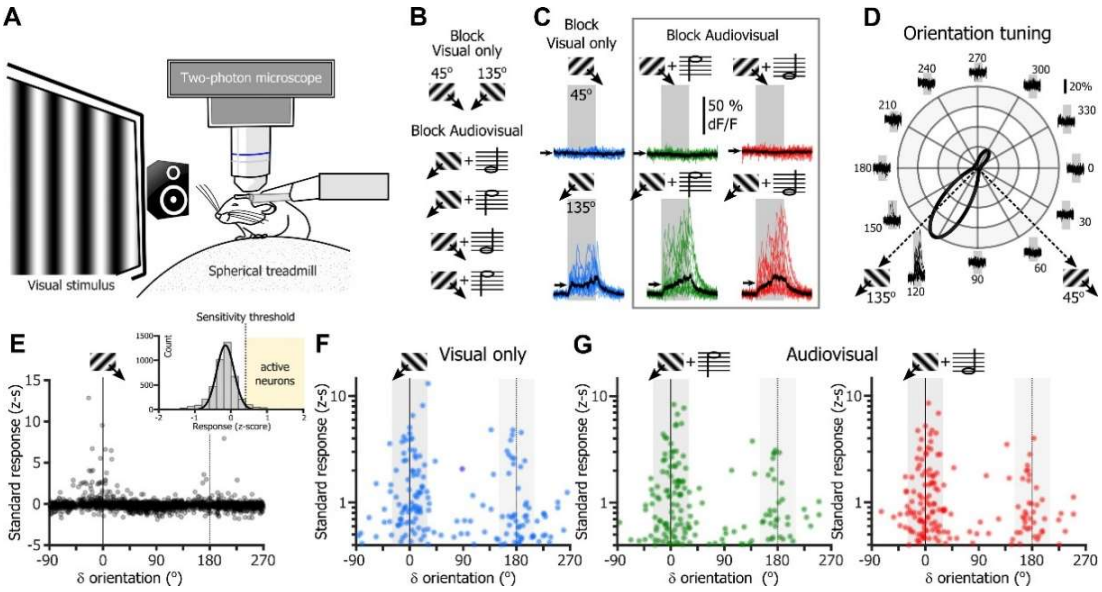

Figure 3-1. Responses of the primary visual cortex layer 2/3 neuronal population to drifting gratings in visual-only and audiovisual contexts. ***A:*** *experimental setup.* ***B:*** *stimuli of the unimodal block and audiovisual blocks.* ***C:*** *variation of the fractional fluorescence (dF/F) of the GCaMP6f signal of the same neuron during the presentation of the test stimuli 45° (top) and 135° (bottom) in the unimodal (blue traces) and multimodal contexts (with 10-kHz tone: green traces; with 5-kHz tone: red traces). Overlapped colored traces show 15 consecutive presentations of the same stimulus, and black traces are the median response. Small black arrows indicate the mean across the duration of the stimulus (gray boxes; duration: 3 s) 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 12 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 (6 orientation) for all excitatory neurons in the database (n = 2,376 neurons). Inset: distribution of the response amplitudes. Gaussian fit: r2 = 0.99; one-tailed t-test with a = 0.05: 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 6 orientation in the unimodal context.* ***G:*** *same representation as in F when a 10-kHz tone (left) or a 5-kHz tone (right) is presented simultaneously with the 135° drifting grating. z-s = z score.*

The presentation of the test stimulus caused a negative response in most of the imaged neurons (-0.119 ± 0.005 z score; t-test: $P < 0.001$; Fig. 3-1E, inset). A minute fraction of neurons displayed large, positive responses to the presentation of the test stimulus (Fig.3-1E; 1.5% of cells with responses > 2 z score). Because the distribution of responses across the imaged population was fitted by a Gaussian curve ($r^2 > 0.99$; Fig. 3-1E, 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 $\alpha = 0.05$; Fig. 3-1E, inset, 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°; Fig. 3-1F) or the two audiovisual contexts (135° test stimulus + 5- or 10-kHz tone; Fig. 3-1G) 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%; x^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.

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

The distributions of preferred orientations in the unimodal and audiovisual contexts (Fig. 3-2A) 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, whereas the circular means of the three distributions pointed similarly toward the orientation of the test stimulus (Fig. 3-2A, 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 (Fig. 3-2B). We found that while having similar orientations (Fig. 3-2C, top), the circular mean vectors were significantly longer in the audiovisual contexts than in the unimodal context (Fig. 3-2C, middle; $P < 0.05$). Moreover, the concentration of the circular distribution (k parameter of the von Mises distribution for circular data) was significantly greater in the audiovisual than in the unimodal context (Fig. 3-2C, bottom; $P < 0.05$). 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 (Fig. 3-2C).

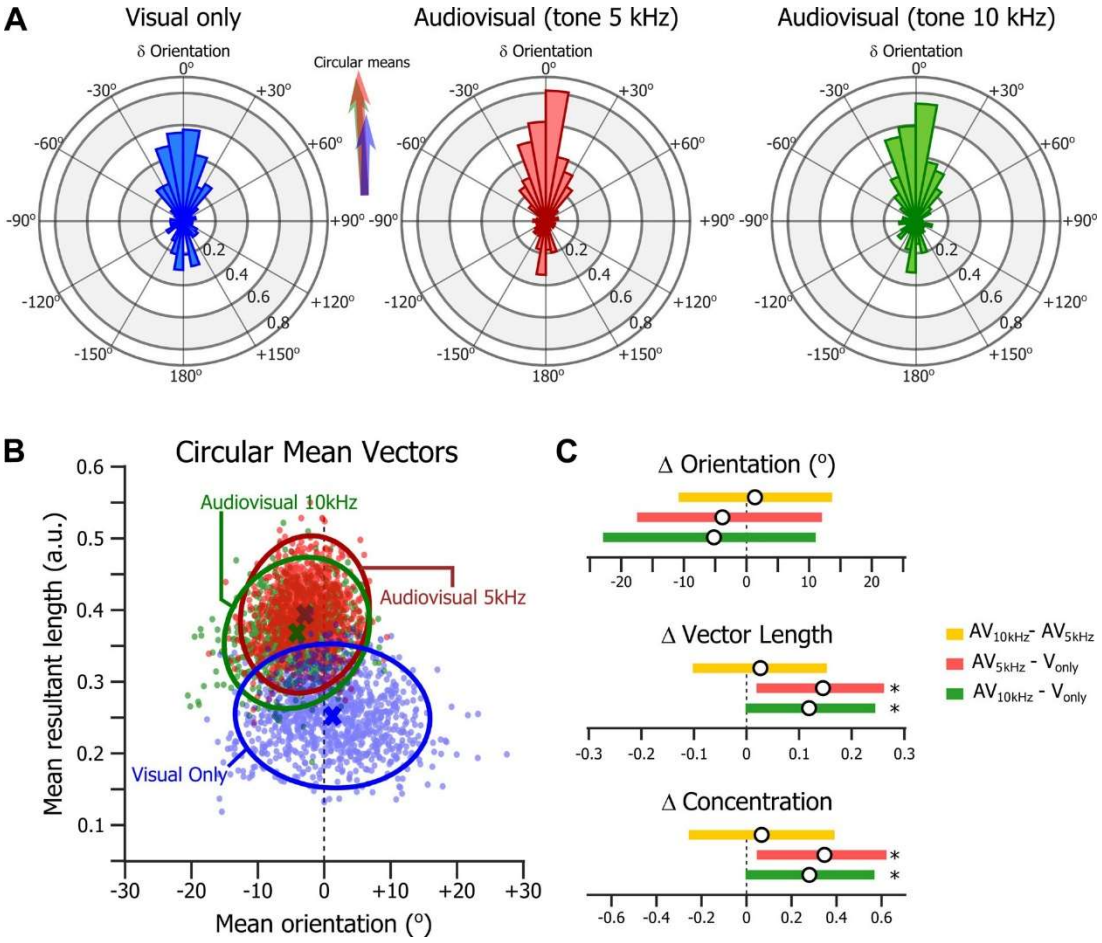

Figure 3-2. Orientation of the population of primary visual cortex layer 2/3 neurons active in the visual-only and audiovisual contexts. ***A:*** *polar histograms representing the distribution (as a probability density function) of the preferred orientations of the active neurons in the visual-only (blue; left) and audiovisual contexts (green: 10-kHz tone, right; red: 5-kHz tone, middle). Inset: circular means of the 3 distributions (scale: 2-fold larger than the polar plots' scale).* ***B:*** *plot of the mean resultant length (in a.u.; arbitrary units) and mean orientation 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 contexts (green dots: 10-kHz tone; red dots: 5-kHz tone). Ovals indicate the probability density functions at the level 0.05 computed from a 2-dimensional Gaussian fit.* ***C:*** *means and confidence intervals of the difference in orientation (Δorientation; top), mean resultant length (Δvector length; middle), and concentration (Δconcentration; bottom) between the unimodal (Vonly) and audiovisual (AV_{5kHz} and AV_{10kHz}) contexts. *P < 0.05, significant difference.*

3.5.2 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 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 in the opposite direction (opposite neurons: preferred orientation >150° from the test stimulus orientation; n = 156; Fig. 3-3A, C). The sound modulation of matching neurons was significantly greater than that of opposite and orthogonal neurons (Fig. 3-3, A and C; Kruskal-Wallis test: $x^2 = 21.3$, $P = 2.4 \times 10^{-5}$; 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 (56.1%) of matching neurons were enhanced (two-sided sign test: P = 0.03), whereas a majority (64.7%) of opposite neurons were suppressed (two-sided sign test: P = 0.0003).

Because the question of whether sound modulation in V1 follows the principle of inverse effectiveness (stating that multisensory enhancement is most prominent

when individual unimodal inputs are weak; Gleiss and Kayser 2012; Serino et al. 2007) 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 score) than spontaneous activity in at least 50% of the trials; Fig. 3-3B]. Here, too, sound modulation of visual responses differed across the three groups of neurons (Kruskal Wallis test: $x^2 = 17.9$, $P = 1.3 \times 10^{-4}$). However, in this case, matching neurons were not significantly modulated (two-sided sign test: $P = 0.30$), whereas orthogonal and opposite neurons were both strongly negatively modulated (two-sided sign test: $P = 0.03$ for both groups; Fig. 3-3B, C).

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 (Fig. 3-3D). A majority (53.1%) of matching neurons inactive in the unimodal context were potentiated by sound (two-sided sign test: $P = 0.03$), whereas 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; Fig. 3-3D, left). Moderately active and highly active orthogonal (Kruskal-Wallis test: $x^2 = 14.9$, $P = 0.006$; Dunn-Sidak post hoc comparison: $P_{inactive\text{-}active} = 0.06$, $P_{inactive\text{-}high} = 0.003$, $P_{active\text{-}high} = 0.47$; Fig. 3-3C, middle) and opposite neurons (Kruskal-Wallis test: $x^2 = 39.8$, $P = 2.3 \times 10^{-9}$; Dunn-Sidak post hoc comparison: $P_{inactive\text{-}active} = 1.3 \times 10^{-6}$, $P_{inactive\text{-}high} = 0.0002$, $P_{active\text{-}high} = 0.68$; Fig. 3-3C, right) 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.

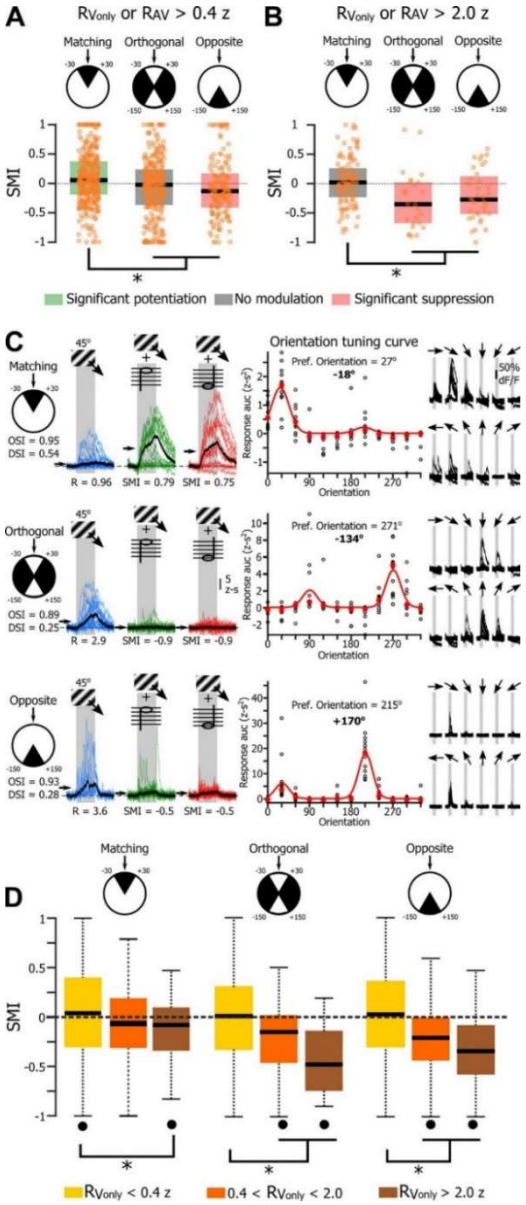

Figure 3-3. Sound modulation as a function of primary visual cortex layer 2/3 neurons' orientation tuning. ***A:*** *box plots (median and 25–75 percentile box) and individual data points of the SMI of active neurons matching the orientation of the test stimulus (left) or matching orthogonal orientations and opposite directions to that of the test stimulus (middle and right, respectively). Colored boxes indicate medians that significantly differ from 0 (2-sided sign test with P < 0.05). *P = 2.0 × 10—5, Kruskal-Wallis test followed by Dunn-Sidak multiple comparisons (Pmatching-ortho = 0.001, Pmatching-opposite = 0.0001, and Portho-opposite = 0.50).* ***B:*** *similar representation as in A when the activity threshold was brought to 2.0 z score. *P = 1.0 × 10—4, Kruskal-Wallis test followed by Dunn-Sidak multiple comparisons (Pmatching-ortho = 0.001, Pmatching-opposite = 0.009, Portho-opposite = 0.81).* ***C:*** *example of a moderately active (0.4 z < response < 2.0 z) matching neuron potentiated by sound (top), a highly active (response > 2.0 z) orthogonal neuron suppressed by sound (middle), and a highly active opposite neuron suppressed by sound (bottom).*

Figure 3-3 cont. *Left, blue, green, and red traces show overlap of the fractional fluorescence (dF/F) expressed in z score of the intertrial activity of 15 randomly selected trials in the visual-only, audiovisual with a 10-kHz tone, and audiovisual with a 5-kHz tone contexts, respectively. Black traces are the median response of all trials recorded in that context. Black arrows indicate the mean across the duration of the stimulus (gray boxes; duration: 3 s) of the median response. Middle, a tuning curve of the neuron was obtained using a resampling-based Bayesian method on the summed %dF/F of individual trials over the duration of the stimulus (or a.u.c, area under the curve). Black circles are values of the summed % dF/F of individual trials, and red circles are the medians. The angular distance (6 orientation) between the neuron's preferred orientation and the orientation of the test stimulus is indicated in bold type. Right, overlap of the calcium signal (in % dF/F) of all the trials used to establish the tuning curve. The matching and orthogonal neurons were recorded simultaneously.* ***D:*** *SMI 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. *P < 0.05, significant difference between groups (Kruskal-Wallis test followed by a Dunn-Sidak multiple comparison). ●P < 0.05, median significantly different from zero (two-sided sign test). DSI, direction selectivity index; OSI, orientation selectivity index.* R_{Vonly} *and* R_{AV}*, responses measured in the visual-only and audiovisual contexts, respectively.*

3.5.3 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 (Fig. 3-4). We found a positive linear correlation between sound modulation and the OSI in matching neurons (Fig. 3-4A, left; 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 (Fig. 3-4A), not in orthogonal (P = 0.95; Fig. 3-4B) and opposite neurons (P = 0.90; Fig. 3-4C).

Next, we tested whether the relation between orientation selectivity and sound modulation in matching neurons was expressed differentially depending on activity levels (Fig. 3-4D, E). 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: $P < 0.001$ and $P < 0.05$, respectively). Inactive matching neurons that were highly orientation selective were significantly more potentiated than moderately active and highly active neurons with similar orientation tuning (Fig. 3-4D, left; Kruskal-Wallis test: $x^2 = 35.4$, $P = 2.1 \times 10^{-8}$; Dunn-Sidak post hoc comparison: $P_{inactive\text{-}active} = 0.02$, $P_{inactive\text{-}high} = 1.3 \times 10^{-8}$, $P_{active\text{-}high} = 0.02$). On the other hand, sound significantly suppressed matching neurons that were less orientation selective (OSI < 0.75) regardless of their responsiveness (Fig. 3-4D, middle and right).

Finally, the SMI of inactive and active matching neurons highly orientation selective for the test stimulus orientation were significantly different from the SMI of moderately selective and poorly selective neurons (inactive neurons: Kruskal-Wallis test: $x^2 = 40.9$, $P = 1.3 \times 10^{-9}$; Dunn-Sidak post hoc comparison: $P_{high\ OSI\text{-}moderate\ OSI} < 0.0001$, $P_{high\ OSI\text{-}low\ OSI} < 0.0001$, $P_{moderate\ OSI\text{-}low\ OSI} = 0.21$; active neurons: Kruskal-Wallis test: $x^2 = 19.75$, $P = 5.1 \times 10^{-5}$; Dunn-Sidak post hoc comparison: $P_{high\ OSI\text{-}moderate\ OSI} < 0.001$, $P_{high\ OSI\text{-}low\ OSI} = 0.0006$, $P_{moderate\ OSI\text{-}low\ OSI} = 0.20$), whereas the SMI of highly active matching neurons was similar across OSI (Kruskal-Wallis test: $x^2 = 4.17$, $P = 0.12$; Fig. 3-4D).

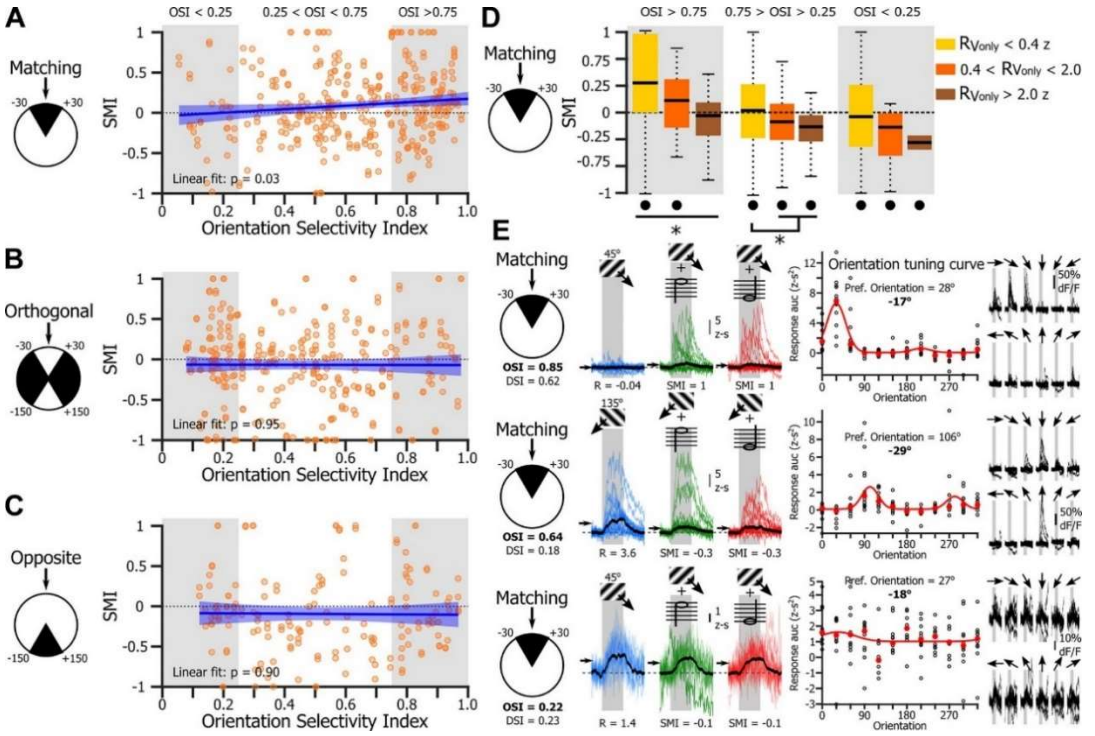

Figure 3-4. Sound modulation as a function of primary visual cortex layer 2/3 neurons' orientation selectivity. ***A:*** *plot of the sound modulation index (SMI) of the matching neurons as a function of their orientation selectivity indexes (OSI). Blue line is the linear fit, and shaded blue areas are confidence bounds of the linear fit (linear fit: r2 = 0.01; F statistic vs. constant model: 4.8; P = 0.03).* ***B:*** *same representation as in A for neurons matching orientations orthogonal to the orientation of the test stimulus (linear fit: r2 = 1 × 10-5; F statistic vs. constant model: 0.004; P = 0.95).* ***C:*** *same representation as in A for neurons matching the direction opposite to that of the test stimulus (linear fit: r2 = 1 × 10-4; F statistic vs. constant model: 0.02; P = 0.90).* ***D:*** *SMI of matching neurons with low (OSI < 0.25), moderate (0.25 < OSI < 0.75), and high orientation selectivity (OSI > 0.75). RVonly and RAV are responses measured in the visual-only and audiovisual contexts, respectively. *P < 0.05, significant difference between groups (Kruskal-Wallis test followed by a Dunn-Sidak multiple comparison). ●P < 0.05, median significantly different from zero (two-sided sign test).* ***E:*** *example of a highly orientation selective (OSI > 0.75) matching neuron potentiated by sound (top), a moderately orientation selective (0.25 < OSI < 0.75) matching neuron suppressed by sound (middle), and a poorly orientation selective (0.25 > OSI) matching neuron suppressed by sound (bottom). Left, blue, green, and red traces show overlap of the fractional fluorescence (dF/F) expressed in z score of the intertrial activity of 15 randomly selected trials in the visual-only, audiovisual with a 10-kHz tone, and audiovisual with a 5-kHz tone contexts, respectively. Black traces are the median response of all trials recorded in that context. Black arrows indicate the mean across the duration of the stimulus (gray boxes; duration: 3 s) of the median response.*

Figure 3-4 cont. *Middle, a tuning curve of the neuron obtained using a resampling-based Bayesian method on the summed % dF/F of individual trials over the duration of the stimulus (or a.u.c, area under the curve). Black circles are values of the summed % dF/F of individual trials, and red dots are the medians. The angular distance (6 orientations) between the neuron preferred orientation and the orientation of the test stimulus is indicated in bold type. Right, overlap of the calcium signal (in % dF/F) of all the trials used to establish the tuning curve. DSI, direction selectivity index.*

Next, we examined whether sound modulation was expressed differentially depending on the neurons' direction selectivity (Fig. 3-5). No linear correlation between sound modulation and DSI was found for matching and orthogonal neurons (P = 0.73 and P = 0.58, respectively; Fig. 3-5, A and B). However, a negative correlation was found for opposite neurons (linear fit: P = 0.015; Fig. 3-5C). 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 (Fig. 3-5D, E).

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 test: $x^2 = 15.2$, P = 0.0005; Dunn-Sidak post hoc comparison: $P_{inactive\text{-}active} = 0.01$, $P_{inactive\text{-}high} = 0.02$, $P_{active\text{-}high} = 0.67$; Fig. 3-5D, left) or moderately direction-selective neurons (0.25 > DSI < 0.75; Kruskal-Wallis test: $x^2 = 27.1$, $P = 1.3 \times 10^{-6}$; Dunn-Sidak post hoc comparison: $P_{inactive\text{-}active} < 0.0001$, $P_{inactive\text{-}high} = 0.01$, $P_{active\text{-}high} = 0.99$).

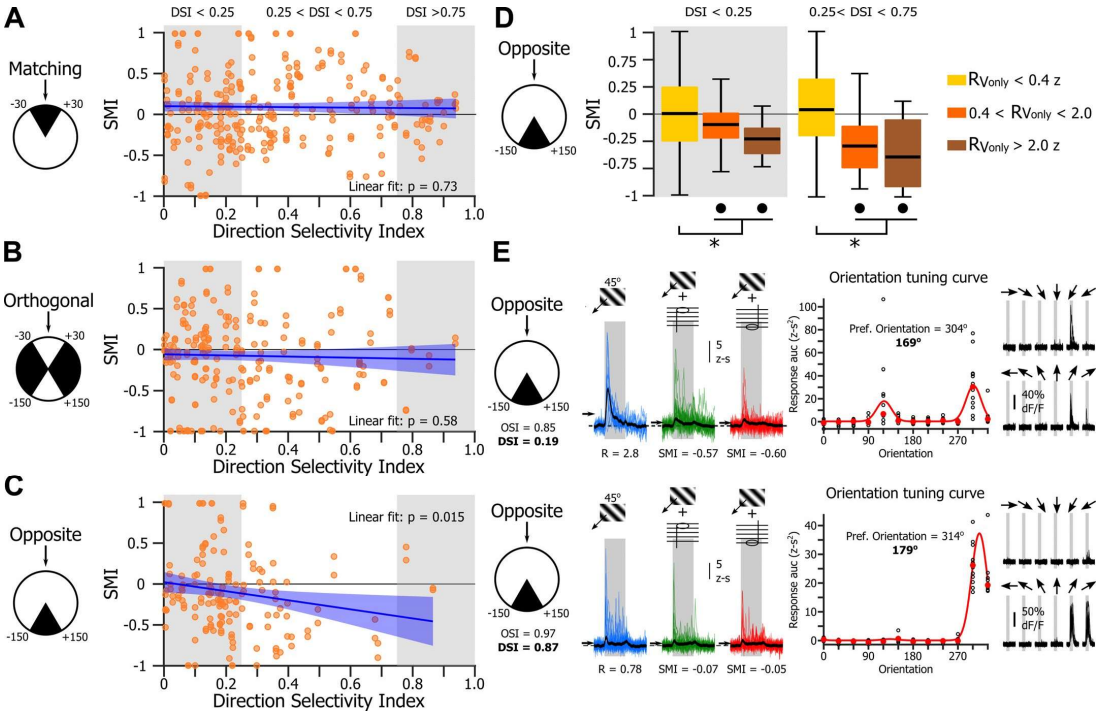

Figure 3-5. Sound modulation as a function of primary visual cortex layer 2/3 neurons' direction selectivity. ***A:*** *plot of the sound modulation index (SMI) as a function of the neuron direction selectivity indexes (DSI) for neurons matching the test stimulus orientation. Blue line is the linear fit, and shaded blue areas are confidence bounds of the linear fit (linear fit: r2 = 3 × 10—4; F statistic vs. constant model: 0.1; P = 0.73).* ***B:*** *same representation as in A for neurons matching orientations orthogonal to that of the test stimulus (linear fit: r2 = 1 × 10—4; F statistic vs. constant model: 0.3; P = 0.58).* ***C:*** *same representation as in A for neurons matching the direction opposite to that of the test stimulus (linear fit: r2 = 0.04; F statistic vs. constant model: 6.0; P = 0.015).* ***D:*** *SMI of opposite neurons with low (DSI < 0.25) and moderate direction selectivity (0.25< DSI < 0.75). Highly direction-selective neurons (DSI > 0.75) were not included because the count was too low. RVonly and RAV are responses measured in the visual-only and audiovisual contexts, respectively. *P < 0.05, significant difference between groups (Kruskal-Wallis test followed by a Dunn-Sidak multiple comparison). ●P < 0.05, median significantly different from zero (two-sided sign test).* ***E:*** *example of an opposite neuron with low direction selectivity (DSI < 0.25) suppressed by sound (top) and an opposite neuron with moderate direction selectivity (0.25 < DSI < 0.75) suppressed by sound (bottom). Left, blue, green, and red traces show overlap of the fractional fluorescence (dF/F) expressed in z score of the intertrial activity of 15 randomly selected trials in the visual-only, audiovisual with a 10-kHz tone, and audiovisual with a 5-kHz tone contexts, respectively. Black traces are the median response of all trials recorded in that context. Black arrows indicate the mean across the duration of the stimulus (gray boxes; duration: 3 s) of the median response.*

Figure 3-5 cont. *Middle, a tuning curve of the neuron was obtained using a resampling-based Bayesian method on the summed % dF/F of individual trials over the duration of the stimulus (or auc, area under the curve). Black circles are values of the summed % dF/F of individual trials, and red circles are the medians. The angular distance (6 orientations) between the neuron preferred orientation and the orientation of the test stimulus is indicated in bold type. Right, overlap of the calcium signal (in % dF/F) of all the trials used to establish the tuning curve. OSI, orientation selectivity index.*

3.5.4 Locomotion and Arousal are Similar in Visual-Only and Audiovisual Contexts

Because locomotion and arousal were both previously found to modulate the response of V1 neurons to visual stimuli (McGinley et al. 2015; Niell and Stryker 2010; Polack et al. 2013; 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 (Fig. 3-6A) 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: $x^2 = 0.4$, $P = 0.81$; Fig. 3-6B). Similar results were found for the imaging experiments presented above (Kruskal Wallis test: $x^2 = 0.8$, $P = 0.66$). Moreover, we found that although pupil size was larger during locomotion vs. no- locomotion trials (two-way ANOVA; effect of locomotion: $F = 43.8$, $P = 4 \times 10^{-9}$), there was no difference in pupil diameter between the unimodal and bimodal contexts in either stationary or locomotion trials (two-way ANOVA; effect of stimulus modality: $F = 0.95$, $P = 0.39$; Fig. 3-6C). Consistent with these

results, gamma power did not differ between the two contexts in V1 (two-way ANOVA; effect of stimulus modality: F = 0.43, P = 0.60; Fig. 3-6D) and in the vicinity of the A1 (two-way ANOVA; effect of stimulus modality: F = 0.22, P = 0.80; Fig. 3-6E).

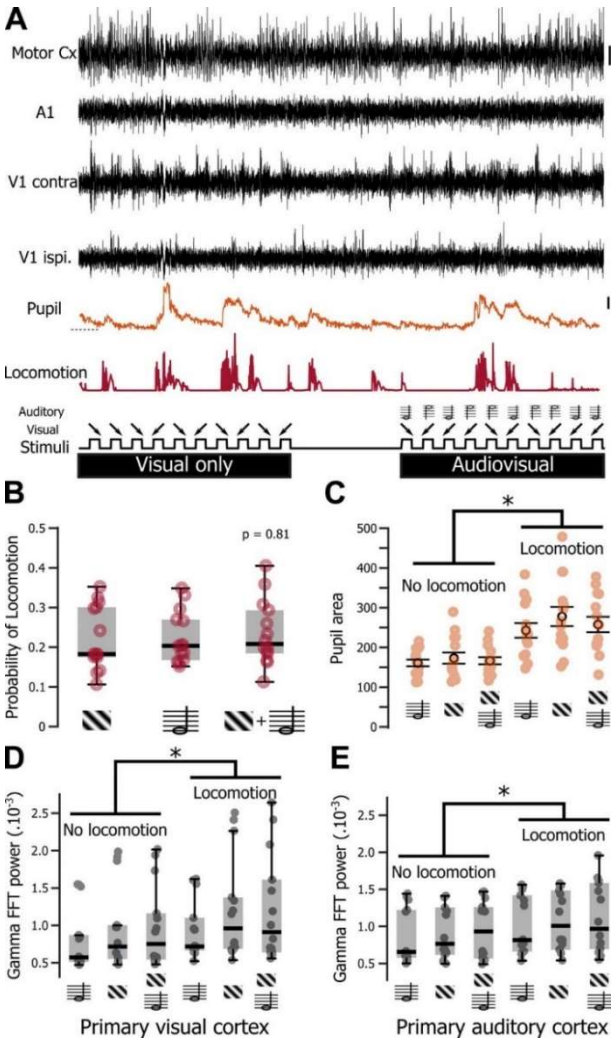

Fig. 3-6. Brain states during visual presentation in the visual and audiovisual contexts. ***A:*** *example of a 4-channel EEG recording (top traces) of the motor cortex (Motor Ctx), primary auditory cortex (A1), primary visual cortex ipsilateral (V1 ipsi.), and primary visual cortex contralateral (V1 contra.) made in a head-fixed mouse placed on a spherical treadmill. Recordings were performed while blocks of visual and audiovisual cues were presented (bottom trace) and pupil size and locomotion were monitored (orange and red traces).* ***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 pupil area [means ± SE (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. Two-way ANOVA; effect of locomotion: F = 43.8, *P = 4 × 10—9; effect of stimulus modality: F = 0.95, P = 0.39; interaction: P = 0.8.* ***D:*** *plots of the power of the gamma band (FFT, fast Fourier transform) measured in the contralateral V1 to the visual stimulus as a function of the stimulus modality and the presence or absence of locomotion. Two-way ANOVA on normal scores (Van der Waerden normalization); effect of loco- motion: F = 11.1, *P = 0.001; effect of stimulus modality: F = 0.43, P = 0.60; interaction: P = 0.97.* ***E:*** *same representation as in D for A1. Two-way ANOVA on normal scores (Van der Waerden normalization); effect of loco- motion: F = 4.7, *P = 0.03; effect of stimulus modality: F = 0.22, P = 0.80; interaction: P = 0.96.*

3.6 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 with 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 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 (McGinley et al., 2015; Niell and Stryker, 2010; Polack et al., 2013; Vinck et al., 2015).

Our findings provide several advances on the previous studies that have so far documented sound modulation in V1 at the individual neuron level (Ibrahim et al., 2016; Iurilli et al., 2012; 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 et al. (2016), which showed a sharpening of the tuning curve of V1 neurons by sound. This finding was debated, because Meijer et al. (2017) were unable to reproduce it. 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 (Gleiss and Kayser, 2012; Meredith and Stein, 1983, 1986; Serino et al., 2007). 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 saturate 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 pioneering 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 A1, such as broadband noise, white noise, or frequency-modulated tones (Ibrahim et al., 2016; Iurilli et al., 2012; 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 (because 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. The relatively high background noise of our recording setup might lead us to underestimate sound modulation carried by A1 projections in V1, by reducing the amplitude of the response of A1 neurons to pure tones (Liang et al., 2014). Further experiments will be necessary to determine if the signal sent by A1 to V1 carries information on the frequency and strength 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 previously shown to control direction selectivity of V1 L2/3 neurons (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 L4 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 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 (Ibrahim et al., 2016; Xu and Callaway, 2009), 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 (L5), which in turn could mediate the inactivation of supragranular neurons (L2/3), 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. Regardless of its origin, the depolarization induced by sound in V1 neurons is likely to have the largest impact in highly orientation selective neurons for which responses to the matching unimodal visual stimulus remain mostly subthreshold. Indeed, because the action potential threshold transforms small fluctuations in membrane potential into large fluctuations in firing rate (Carandini, 2004), a small depolarization induced by sound is likely to improve dramatically the probability and amplitude of the responses to the matching visual stimulus in these 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, a V1 neuron's preferred orientation is stable across contrasts (Alitto and Usrey, 2004) and temporal frequency (Moore et al., 2005) and is 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, temporal frequency tuning. and on 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 by 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 (Gleiss and Kayser, 2014; Lippert et al., 2007; Odgaard et al., 2004) and discrimination thresholds (Vroomen and de Gelder, 2000) as well as to decrease the reaction times (Gielen et al., 1983; Hershenson, 1962; Posner et al., 1976) during object perception tasks in multimodal 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).

Chapter IV

4.1 Pure tones improve the representation of oriented stimuli in V1 and facilitate angular discrimination

4.2 Background

In the first aim, we demonstrated that the evoked responses of V1 neurons tuned to the orientation and direction of the visual stimulus are potentiated in the audiovisual context, while the responses of neurons tuned for orthogonal orientations or the opposite direction are suppressed (McClure and Polack, 2019). This modulation of V1 activity led to a refinement of the representation of the visual stimulus in the multisensory context. In this aim, we address the question of whether the improvement in the neural representation of orientation, by concomitant auditory stimuli, underpins an actual improvement in perceptual discrimination.

Several past studies have suggested that the refinement of the representation of a visual stimulus could impact visual perception. One theory, known as the reverse hierarchy theory, and developed by Hochstein and Ahissar (2002) posits that the information available at the lowest level of the sensory processing hierarchy constrains the resultant behavior. According to this theory, since orientation selectivity is an emergent property of V1, and that V1 is required for orientation discrimination, at least in mice (Glickfeld et al., 2013), the representation of orientation in V1 constrains orientation perception. Under this theoretical framework, the sound-induced sharpening of the representation of orientation in V1 should lead to a decrease in the angular distance at which mice can perceptually discriminate between two oriented

stimuli, by favoring the separation of the populational representations. In the cat and monkey, the orientation sensitivity in V1 neurons has been correlated with the ability to perceptually discriminate between different stimuli (Bradley et al., 1987; Vogels and Orban, 1990). For example, the discrimination threshold for orientation discrimination of V1 neurons of anesthetized cats corresponds to the known psychophysical performance of cats (Bradley et al., 1987). However, because anesthesia is known to alter the activity patterns of V1 neurons and disrupt V1 response dynamics across cortical layers (Aasebo et al., 2017; Sellers et al., 2015), those results need to be validated in behaving animals.

In this aim, the goal was to determine if the sound-induced improvement of the representation of orientation, observed in aim I, can be associated with an improvement in discrimination ability. To achieve this, I trained mice to perform a Go/NoGo orientation discrimination task using two oriented drifting gratings and then gradually reduced the angular distance (AD) between the two stimuli. I determined the limit of the mice' visual discrimination ability and tested for the possible effects of coincident pure tones on the mice' performance. I then analyzed the activity of the neurons located in the layer 2/3 of V1 to link any sound-induced alteration of the representation of orientations with an improved discrimination ability.

4.3 Materials and methods

Two-photon calcium imaging was performed in awake mice during a Go/NoGo visual discrimination task to test the impact of sound on visual perception.

4.3.1 Animals

For technical details on the animals used in this experiment, see section 2.1 of Chapter II.

4.3.2 Stereotaxic surgery

For technical details on all surgical procedures performed in this experiment including head bar implants, virus injections, and coverslip implants, refer to section 2.2 of Chapter II.

4.3.3 *In vivo* calcium imaging

For technical details see section 2.3 of Chapter II

4.3.4. 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 (The MathWorks) using the Psychtoolbox (Brainard 1997).

A Go/NoGo visual discrimination task was implemented to test the ability of mice to discriminate between two visual stimuli of decreasing angular distances in unimodal and multimodal contexts. Prior to training, mice were water

restricted for five days. Water consumption was limited to 1 ml per day, corresponding to 35% of ad libitum consumption. The goal of water restriction was to reduce their weight by 15 %, which was sufficient to create adequate motivation during the behavioral training, without being detrimental to their health. Mice were monitored daily to ensure their weight did not drop by more than 20% of their pre-training starting weight. When necessary, mice were given supplemental water (0.3 ml to 1.0 ml) beyond their daily limit to help them maintain a healthy weight. During training, mice were administered water during 1-hour/day training sessions.

Following water restriction, mice were trained to discriminate between a 45^{o} (Go) and a 135^{o} (NoGo) visual stimulus with the following parameters: temporal frequency: 2 Hz, spatial frequency: 0.04 cycle/deg, contrast: 75%, duration: 3 sec, intertrial interval: 3 sec. For the Go trials the mice were trained to lick during the last second of the trial (response window) in order to receive a 0.01 ml water droplet; recorded as a 'Hit' response. If the mice failed to lick during the response window, the trial was considered a 'Miss' and a 9 second timeout was imposed. Unlike the Go trials, a successful NoGo trial occurred if there was a no lick response during the response window and was recorded as a 'Correct Rejection'. In addition, if the mice incorrectly licked during the one second response window of a NoGo trial, the trial was considered a 'False alarm' and the mice received a 9 second timeout. A timeout was defined as the absence of visual and auditory stimulation and was added to the 3 second intertrial interval, totaling 12 seconds. Water droplets were only delivered during 'Hit' trials.

Once mice performed at an expert level (> 90% correct response for 300 trials), we tested them on their ability to discriminate between the Go and the NoGo cue at smaller angular distances, in the absence or presence of sound. Each day mice were tested on one of the following angular distances between the Go and NoGo drifting gratings 45°, 30°, 25°, 20°, and 15°. In the audiovisual context, visual stimuli were paired with one of two pure sinewave tones (5 or 10 kHz). The orientation of the Go signal remained the same (45°) across all imaging sessions in both the visual-only and audiovisual conditions. However, the NoGo signal was incrementally reduced from the training orientation (135°) by one of the angles listed above. Finally, at the end of each recording session, the orientation tuning of all recorded neurons was assessed by presenting 12 repetitions of 12 orientations, evenly spaced by 30° (Fig 4-1).

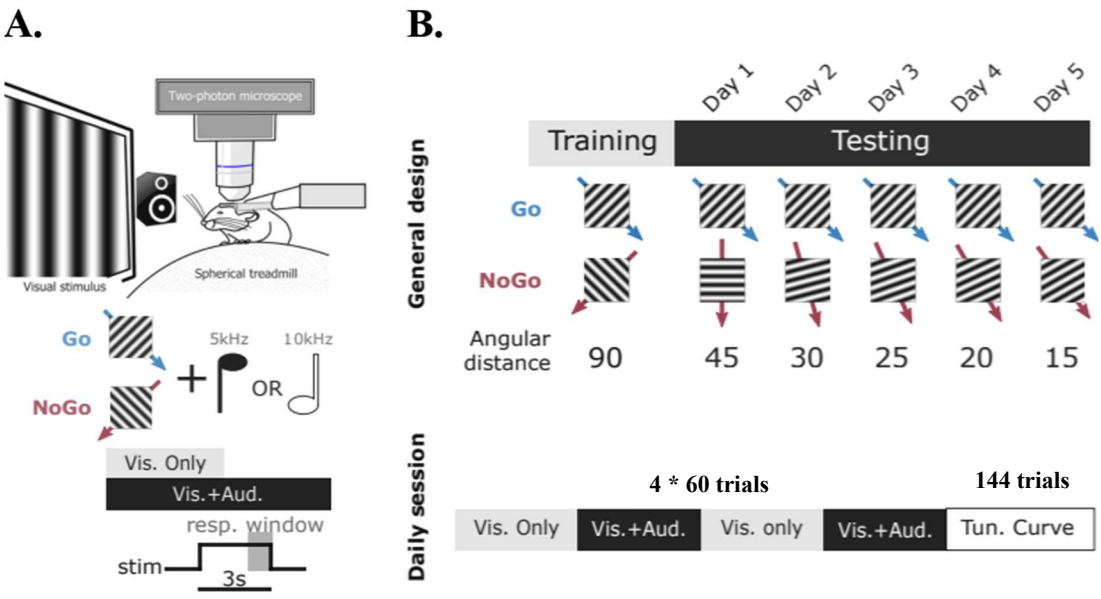

Figure 4-1. Go/NoGo visual discrimination task. ***A.*** *Illustration of the experimental setup (top panel) along with the individual visual and auditory stimuli presented during visual-only and audiovisual trials, as well as the time course of a typical trial (middle and bottom panels, respectively).* ***B.*** *(Top panel) general design of the training and testing phase of the visual discrimination task. During training, mice were presented with orthogonal visual stimuli consisting of Go (45º) or NoGo (135º) drifting gratings. Mice were rewarded only when they licked for the Go visual stimulus during the last second of the stimulus presentation (resp. window). During the testing phase, mice were tested on their ability to discriminate between the Go and NoGo visual cues presented at incrementally smaller angular distances, over the course of five sessions. (Bottom panel) daily sessions consisted of block presentations of visual only and audiovisual cues totaling 240 trials. At the end of each recording session the tuning curve was measured for each recorded neuron by presenting 12 different oriented drifting gratings separated by 30º.*

4.3.5 Data analysis

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

Behavioral data analysis. For each trial the following behavioral measures were recorded and stored in a Microsoft SQL server database located on a local server

(HPE ProLiant ML350 Gen10), for each animal: the stimulus presented (Go or NoGo unimodal or audiovisual cue), behavioral response (lick or no lick), locomotion, and eye-tracking. The behavioral performance averaged across all mice was synthesized as the mean proportion of trials for each stimulus condition (Go Vis. Only (GoV_o), Go Vis. + Aud. (GoAV), NoGo Vis. Only ($NoGoV_o$), NoGo Vis. + Aud. (NoGoAV), in which the mice made a correct response (Hit or Correct Rejection). The behavioral analysis for each mouse was synthesized using the D' statistic (Einstein et al., 2017).

D', also known as the sensitivity index, is a statistical measure of how well a subject can detect and respond to a signal against background noise or another signal. D' is defined as the standardized difference between the 'Hit' rate and the 'False alarm' rate and is calculated as z(Hit/(Hit + Miss)) – z(FA/(FA + CR)), where z indicates the z-transform of the 'Hit' and 'FA' rates (Stanislaw and Todorov, 1999). When the Hit rate = the FA rate, D' is 0 and is indicative of chance performance, whereas a behavioral performance with a D' greater than 1.75 has a probability less than 0.001 to be driven by chance (Einstein et al., 2017).

Neural data analysis. Raw calcium-imaging signals, recorded during the visual discrimination task and orientation tuning, were preprocessed (see section 3.3.7, paragraphs 1 and 2). Spiking activity was inferred from the fractional fluorescence (dF/F) signal of each neuron, using a deconvolution algorithm (Deneux et al., 2016). This algorithm was benchmarked on actual data and was shown to be the most performant and best adapted for our signal/needs (Berens et al., 2018). The response of each neuron for each trial (including the preceding and following intertrial

intervals) was stored as fractional fluorescence (dF/F) and inferred action potential discharge in the SQL database.

To determine the impact of sound on the representation of orientation in V1 in the unimodal or audiovisual condition, a resampling method was applied to the deconvolved calcium-imaging data. In particular, 250 neurons were randomly selected from the entire population of recorded neurons across all animals and for each stimulus condition (GoVisOnly, GoAV, NoGoVisOnly, and NoGoAV). Next, a random trial was selected for each of the 250 neurons, allowing us to reconstruct random population trials from the entire dataset of recorded neurons. To elucidate how V1 represents the orientation of visual stimuli during the task, we needed to quantify the amount of evoked activity that represents each orientation. To accomplish that, we sorted each neuron, from the random subpopulation, according to their preferred orientation. The spiking activity was then smoothed across orientations by performing a sliding Gaussian-weighted average (width = 30^{o}). This reconstruction of the orientation representational space allowed us to visualize and quantify the trial-by-trial average number of spikes per time point for each orientation, termed hereafter the spiking density. This procedure was repeated 1000 times as a means of reproducing a typical V1 trial (Fig 4-2).

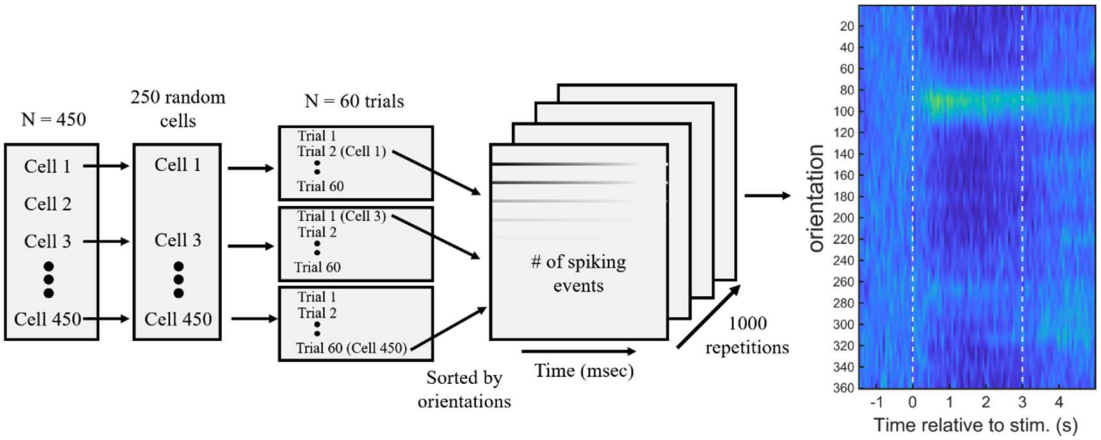

Figure 4-2. Schematic of the resampling method for the generation of population PSTHs. *In all four stimulus conditions (GoVo, GoAV, NoGoVo, NoGoAV), for each of the five angular distances tested (45º, 30º, 25º, 20º, and 15º) 250 random V1 neurons were selected from the whole population of V1 neurons recorded across all animals (e.g. N = 450). Next, a single random trial was selected from the total available trials (e.g. N = 60) for each of the 250 neurons. Each of the neurons along with their randomly chosen trial was sorted and grouped based on their preferred orientation (width = 30º). The resultant matrix consisted of the number of spiking events recorded for each of the 250 selected neurons and trials at a given time point. This procedure was repeated 1,000 times and a Gaussian-weighted average was computed and plotted as population PSTHs.*

Once population PSTHs were computed, additional analyses were performed to compare the populational orientation representations of the Go and NoGo stimuli, in the visual only and audiovisual contexts. First, I summed the spiking density, for each orientation, for the first two seconds after stimulus onset, allowing me to visualize the profile of V1 activation in the orientation domain. The location of the peaks (i.e. the orientations represented by the most activated neural subpopulation(s)) were extracted from these profiles by selecting the portions of values that were above 1.645 SD of the whole profile values. Trials where none of the activation profile values exceeded the threshold were not included in this analysis. Next, for every trial I extracted the orientation at the maximum. Finally, I compared the strength of the

representations by comparing the amplitude at the peak(s) extracted from the mean profile. To account for the occurrences where the activation profiles contained more than one significant peak, I repeated the amplitude measure for every peak (i.e. every discontinuous portion of the mean profile that exceeded the threshold).

4.3.6 Statistics

Behavioral data. A z-test on binomial proportions was used to determine if sounds affected visual discrimination for the entire cohort of mice imaged. For this test, I compared the proportion of correct responses between the unimodal and audiovisual conditions for each angular distance (e.g., 45º, 30º, 25º, 20º, and 15º). However, for the analysis of individual animals' performance in the visual-only versus the audiovisual conditions across the five angular distances tested, a permutation test (randomization test) was performed. For this, the behavioral responses for the unimodal and audiovisual context, for each animal, were randomly permuted 10,000 times, allowing me to compute the likeliness to observe, by chance, the measured difference in D' between the visual-only and audiovisual conditions. All statistical tests were two-tailed, and significance was judged at the 0.05 α level.

Neural data. To compare the orientation(s) at the maximum, between stimulus conditions, a Wilcoxon signed-rank test was performed. As statistical tests on large samples (1000 repetitions) are likely to yield significant differences even if the said difference is unrelatedly small, the Wilcoxon test was accompanied by the Cohen's U3 measure of effect size (Hentschke & Stüttgen, 2011). Ninety-five percent confidence intervals (CI) of Cohen's U3 were used to compare the effect size between

conditions. This specific non-parametric measure of effect size was chosen because of the bi-modality and skewness of the orientation distributions. Classically, a U3 of 0.50 ± 0.1, indicates the absence of a significant effect. Values between 0.1 and 0.2, apart from 0.5, indicate a small effect. A medium effect size ranged from 0.2 to 0.3, while a large effect size ranged from 0.3 to 0.5 (Hentschke and Stüttgen, 2011). Additionally, we performed t-tests along with Hedge's g measure for effect size to compare the amplitudes at the peaks of the mean activation profiles. We used T-test and Hedge's g because the distribution of the amplitude values followed a Gaussian distribution. A Hedge's g effect size is generally interpreted as follows: no significant effect if $g < 0.2$, small: ± 0.2, medium: ± 0.5, large: ± 0.8 (Hentschke and Stüttgen, 2011).

4.4 Results

In this study, the main objectives were: [1] to determine if the representation of orientation in V1 was linked to orientation discrimination, [2] to characterize the minimal angular distance between two oriented stimuli that the mice can discriminate, and [3] to establish if the sound-induced modulation of the representation of orientation in V1 also impacted orientation discrimination by reducing the minimal angular distance between the two visual stimuli. To accomplish those objectives, mice were trained to discriminate between two orthogonal drifting gratings using a Go/NoGo paradigm, before testing their ability to discriminate smaller and smaller angular distances between the Go and NoGo signals (Fig. 4-1B). Two-photon calcium imaging was performed to record the activity of V1 L2/3 neurons while mice executed the orientation discrimination task.

4.4.1 Behavioral performance as a function of the angular distance between the Go and NoGo visual cues

Since the orientation discrimination task involved a binary outcome (lick or no lick response), behavioral performance was based on the proportion of trials for which mice licked or withheld licking during the presentation of the Go and NoGo stimuli, respectively. When the larger angular distance between the two visual cues was tested (45^{o} angular distance), trained mice correctly licked for a water reward on the vast majority of the Go trials and refrained from licking for water during NoGo trials (Fig 4-3; Top- and bottom-left panels, respectively). On the contrary, when the mice were tested with an angular distance of 15^{o} between the two cues, the number of errors i.e. absence of lick response for Go trials ('Miss') and lick response for NoGo cue ('False alarm'; Fig 4-3; Top- and bottom-right panels, respectively) increased considerably to the point that there was no difference in the probability of lick response for the Go and NoGo trials (Fig 4-3; Top-right vs. bottom-right panels). Therefore, as expected, the behavioral performance was affected by the difficulty of the task. The range of angular distances used allowed me to characterize V1 processing when the two cues were easily distinguishable (45^{o}), and when the two signals were too similar to enable discrimination (15^{o}), leading to a random response.

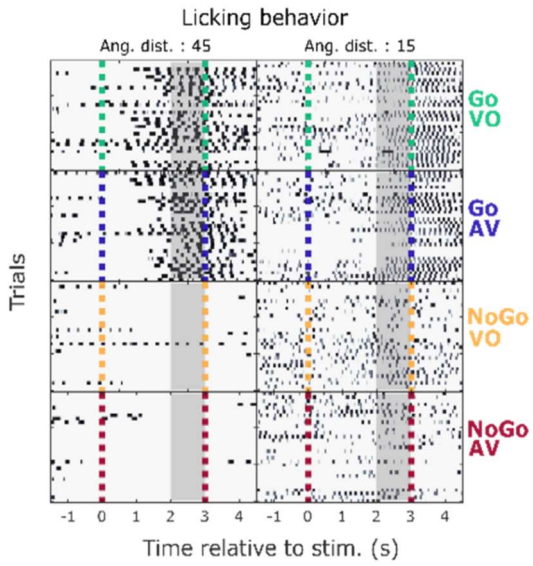

Figure 4-3. Lick responses of a single mouse during the Go/NoGo orientation discrimination task. *The raster plot depict the accumulated lick responses of a single mouse for the GoVo (Green), GoAV (Blue), NoGoVo (Yellow), and the NoGoAV (Red) orientations in the visual only and audiovisual conditions, when the Go and NoGo stimuli were separated by 45º (left panels) and 15º (right panels). The gray area indicates the response window. Comparisons in lick responses between the Go and NoGo orientations were only made during the response window.*

4.4.2 Sound-induced modulation of the behavioral performances as a function of the angular distance between the Go and NoGo cues

For the next analysis I determined for every mice the proportion of correct responses to the presentation of the Go and the NoGo cues in the unimodal condition and the audiovisual condition. We found that [1] the proportion of correct responses decreased for both the Go and the NoGo cues when the angular distance between the Go and NoGo cue was smaller than 30º (Fig 4-4), and [2] sound significantly improved the behavioral performance. Indeed, in the audiovisual context, the proportion of correct responses to the Go orientation was increased when the angular distance between the two cues was 30º (z-test on binomial proportions; Go orientation at 30º angular distance; $p = 0.035$) and 25º (Go orientation at 25º angular distance; $p = 0.0013$). For the NoGo stimuli, the audiovisual context improved the proportion of correct responses when the angular distance between the two cues was 25º (NoGo orientation at 25º angular distance; $p = 1.79 \times 10^{-6}$), and 20º (NoGo orientation at 20º

angular distance; p = 0.006). The most substantial sound-induced improvement in behavioral performance occurred, therefore, for the 25° angular distance for both the Go and NoGo orientations. For 20° and 15°, mice were behaving at chance levels (i.e., were not able to discriminate between the Go and NoGo orientations), with a bias toward licking as evidenced by a probability below 0.5 to withhold licking when the NoGo cue was presented.

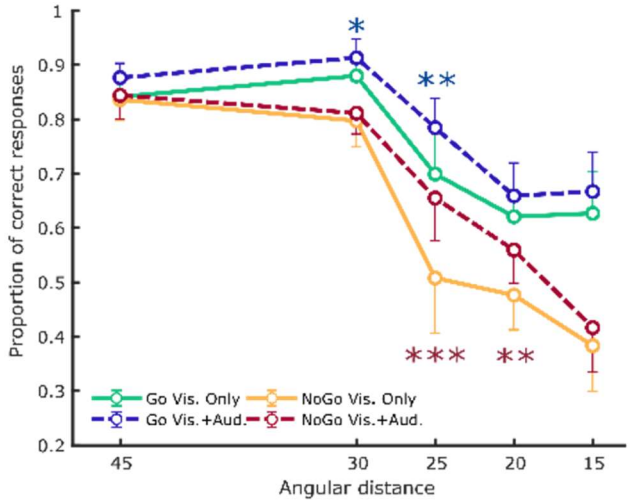

Figure 4-4. Sound-induced effects on population performance in an orientation discrimination task. *The mean proportion of correct responses for mice tested on their ability to discriminate between Go and NoGo orientations of a given angular distance. Performance was measured and compared for the unimodal and audiovisual conditions. * Z-test on binomial proportions (p < 0.05).*

To confirm my results measured at the population level, I estimated the impact of sound on the behavioral performance of each mouse by computing the D' (see section 4.3.6) in the unimodal and audiovisual conditions. Sounds significantly improved orientation discrimination at 25° angular distance in 6 out of 9 of the mice tested (Fig 4-5; permutation test: $VO_{mouse1} - AV_{mouse1}$; p-value = 0.002; $VO_{mouse2} - AV_{mouse2}$; p-value = 0.008; $VO_{mouse3} - AV_{mouse3}$; p-value = 0.006; $VO_{mouse4} - AV_{mouse4}$; p-value = 0.004; $VO_{mouse5} - AV_{mouse5}$; = 0.050; $VO_{mouse6} - AV_{mouse6}$; = 1.000×10^{-5}). Sound-induced improvements in orientation discrimination was also found at 45° for 2 out of 11 mice ($VO_{mouse1} - AV_{mouse1}$; p-value = 0.020; $VO_{mouse4} - AV_{mouse4}$; p-value = 0.020), at 30° for 2 out of 11 mice ($VO_{mouse5} - AV_{mouse5}$; p-value = 0.005; VO_{mouse7}

– AV_{mouse7}; p-value = 0.030), and at the 20° for 2 out of the 8 mice tested (VO_{mouse3} – AV_{mouse3}; p-value = 0.010; VO_{mouse5} – VA_{mouse5}; p-value = 0.009). No differences in behavioral performance were found between the visual-only and audiovisual context in individual mice at the 15° angular distance. Furthermore, at 15° angular distance, the mice performed at chance as indicated by a D' of 0.

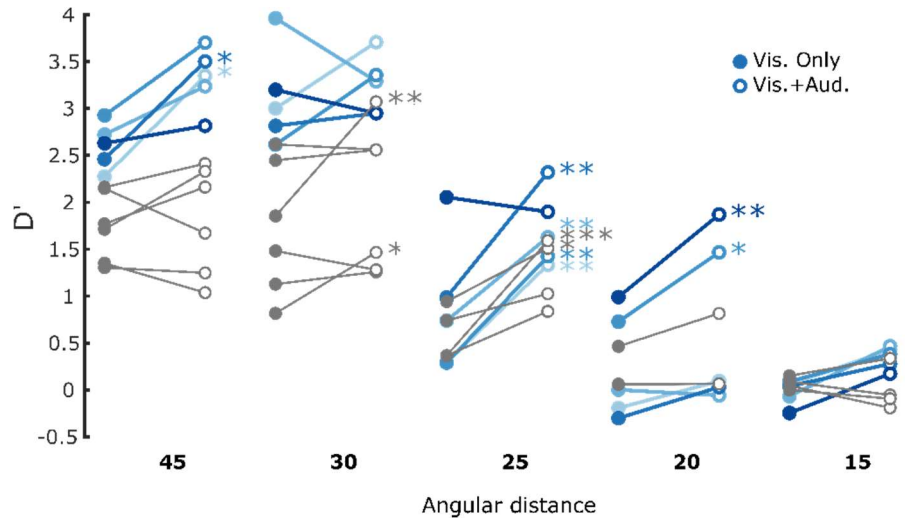

Figure 4-5. Sound-induced effects on individual mice performance in the orientation discrimination task. *Individual behavioral performances, measured as the D'prime statistic, for the five angular distances tested in the unimodal and audiovisual condition. A permutation test revealed the largest impact of sound for individual mice (n = 6 of 9) at the 25° angular distance. Sound-induced improvements were also found at the 45°, 30°, and 20° angular distances. * Permutation test (P < 0.05).*

Taken together, both the population and individual behavioral performance in the Go/NoGo orientation discrimination task showed that an angular distance between the Go and NoGo stimuli of 25° was the smallest angular distance that the mice could discriminate between the Go/NoGo orientations.

4.4.3 Representation of the orientation in V1 of a Go and NoGo stimuli distant by 45º

To assess V1 population representation of oriented stimuli with or without auditory cues during the visual discrimination task, I generated population PSTHs (see methods); an analysis that allowed me to evaluate and visualize neural population activity as a function of the neurons' preferred orientation. When the angular distance between the Go (45º) and the NoGo (90º) stimuli was 45º, a robust distinction between the V1 population representations of the two stimuli could be observed (Fig 4-6A). The Go visual stimulus activated a population of neurons representing orientations surrounding that of the Go signal (Fig 4-6A; Top panels), while the NoGo stimulus activated a separate subpopulation of neurons comprising units representing two non-overlapping sets of orientations (Fig 4-6A; Bottom panels). Go and NoGo evoked responses started less than 1 second after the onset of the stimulus presentation and lasted throughout its duration. In the presence of sound, neither representations of the Go and the NoGo orientations were altered (Fig 4-6A; Left vs. right panels).

In order to further define the orientations that were represented by the activated ensembles of V1 neurons, we summed for each orientation of the PSTHs the activity of the first two seconds after stimulus onset, generating activation profiles in the orientation domain. The two second time window was chosen because it most likely corresponds to the time during which V1 activity is relevant for the animal's response (Resulaj et al., 2018). These activation profiles were used to extract parameters such as the orientation and the amplitude of the single or multiple peaks of response profiles

and provided a method for reading out the orientation and strength of the population response.

Hence, we calculated the location of the maximum of the activation profile for each reconstructed trial and compared the resultant distributions in the unimodal and audiovisual contexts (4-6C). During the presentation of the Go stimulus in the unimodal condition, the maxima were distributed bimodally with a median of 41°. The distribution obtained for the audiovisual Go was similar with a median of 43°. Both medians were not significantly different (Wilcoxon Test; $p = 0.223$). For the unimodal and audiovisual NoGo presentations, distributions were also similar: most of the trials had a maximum around 130°, while a relatively smaller portion of trials were close to 100° (unimodal NoGo vs. audiovisual NoGo, median = 131° and 131°, $p = 0.566$). In the unimodal and audiovisual contexts, the representation of the Go and NoGo signals were absolutely different as assessed by the effect size (Unimodal Go vs Unimodal NoGo: $p = 1.60 \times 10^{-123}$, Cohen's U3 for effect size = 0.941; Audiovisual Go vs Audiovisual NoGo: $p = 2.04 \times 10^{-120}$, U3 = 0.935).

We compared the strength of the orientation representation using the amplitude of the activation at the different peaks of the activation profile. For the unimodal NoGo condition, the activity at the 132° peak was significantly higher than at 103° peak (mean amplitude ± 95% CI; 103°: 0.264 ± 0.0234; 132°: 0.527 ± 0.0331; $p = 1.07 \times 10^{-35}$; Hedge's $g = 0.569$). Similarly, in the audiovisual NoGo condition, the most strongly activated subpopulation of neurons was representing 132° (mean amplitude ± 95% CI; 103°: 0.211 ± 0.0257; 132°: 0.471 ± 0.0320; $p = 3.526 \times 10^{-34}$; $g = 0.555$). No significant difference was found when comparing the amplitudes at the maximum

between Go and NoGo (Unimodal Go vs Unimodal NoGo: $0.522 \pm 2.68 \times 10^{-2}$ vs $0.527 \pm 3.31 \times 10^{-2}$, $p = 0.794$; Audiovisual Go vs Audiovisual NoGo: $0.517 \pm 2.68 \times 10^{-2}$ vs $0.471 \pm 3.19 \times 10^{-2}$, $p = 3.39 \times 10^{-2}$, $g = 9.49 \times 10^{-2}$) or Unimodal and Audiovisual contexts (Unimodal Go vs Audiovisual Go: $p = 0.790$; Unimodal NoGo vs Audiovisual NoGo: $p = 1.72 \times 10^{-2}$, $g = 0.106$).

To summarize, the presentation of the Go and the NoGo stimuli separated by 45°, generated two distinct profiles of activation, which was in line with the behavioral results that indicated optimal discrimination sensitivity. These clearly segregated representations were mediated by the activation of non-overlapping populations of neurons. The presence of auditory stimuli did not change the location of the V1 response in the orientation representational space nor the magnitude of the representations. Interestingly, the NoGo stimulus mostly activated a population tuned around 132°; an orientation that did not correspond to the actual stimulus presented (90°), but was close to the orientation of the NoGo stimulus during the training phase (135°).

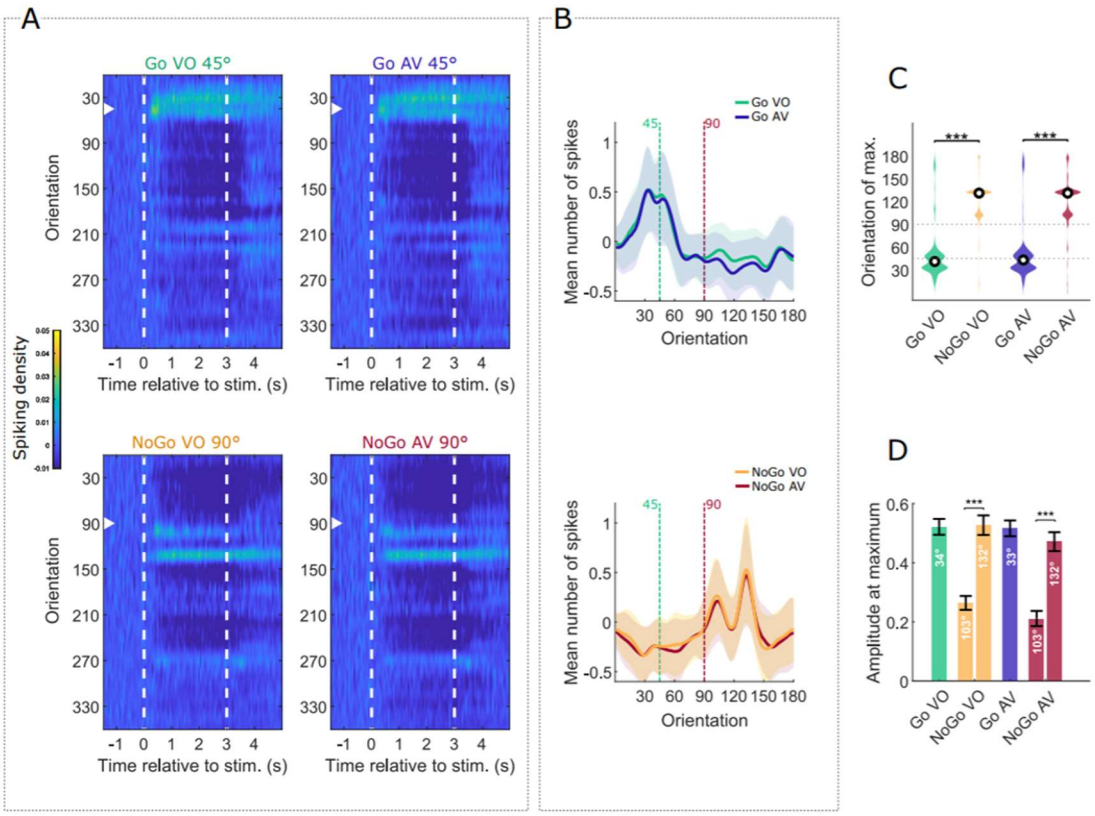

Figure 4-6. V1 representation of orientation at 45º AD with and without auditory stimulation. ***A.*** *Population PSTHs for all four stimulation conditions. Every line corresponds to the spiking activity representing different visual orientations. Warmer colors represent an increase in spiking density compared to the baseline activity. Dashed white lines indicate the starting and ending times of the stimuli. The white arrowheads indicate the orientation of the visual stimulus considered in each plot.* ***B.*** *Profiles of V1 summed activation in the orientation representational space during the first two seconds post stimulus onset, for the Go VO and AV (top) and NoGo VO and AV stimulation contexts (bottom). Solid lines show the average over the 1000 reconstructed trials. Shaded areas show the standard deviation. Dashed green and red lines indicate the orientation of the Go and NoGo stimuli respectively.* ***C****. Distributions of maximum location in the orientation domain.* ***D.*** *Amplitude (mean ± SEM) at the peaks of the activation profiles from B. Numbers in white indicate the orientation of the peak.*

4.4.4 Representation of the orientation in V1 of a Go and NoGo stimuli distant by 30°

The results for the angular distance of 30° between the Go (45°) and the NoGo (75°) drifting grating similarly revealed distinct representations of the two stimuli in V1 L2/3 (Fig 4-7A; Top- and bottom-left panels). Neither representations of the Go and the NoGo orientations were affected by the presence of auditory stimuli (Fig 4-7A; Top and bottom right panels). The activation profiles evoked by the Go stimulus (45°) presented a single peak of activity centered on the Go stimulus orientation (Fig 4-7A; Top). The NoGo stimulus presentation (75°) generated a broader activation profile, whose peak was centered away from the stimulus orientation (Fig 4-7A; Bottom). For both stimuli, the unimodal and audiovisual profiles overlapped for the majority of the orientations.

The median of the single trial orientation profile maximum for the Go stimuli was 49° and 50° in the unimodal and audiovisual conditions, respectively. Those two unimodal distributions were not significantly different (Wilcoxon test; $p = 0.152$). For the unimodal NoGo and audiovisual conditions, the single trial orientation profile maxima showed similar unimodal but broader distributions: in the unimodal NoGo trials the median was 84°, while in the audiovisual NoGo condition the median was 87° ($p = 1.05 \times 10^{-2}$, U3 = 0.550). Both Go and NoGo distributions were different, with comparable effect size for the unimodal and audiovisual contexts (Unimodal Go vs Unimodal NoGo: $p = 4.20 \times 10^{-7}$, U3 = 0.689; Audiovisual Go vs Audiovisual NoGo: $p = 2.68 \times 10^{-7}$, U3 = 0.667).

When comparing the amplitude of the responses of the neurons activated by the visual stimulus in the unimodal Go (orientation at the maximum = 42°) and the unimodal NoGo (88°) conditions, we found that the unimodal Go peak was significantly higher than the unimodal NoGo peak (unimodal Go vs unimodal NoGo: $0.190 \pm 2.02 \times 10^{-2}$ vs $0.094 \pm 2.01 \times 10^{-2}$; $p = 6.62 \times 10^{-11}$, $g = 0.293$). Similarly, the audiovisual Go peak at 43° was significantly higher than the audiovisual NoGo peak at 89° (audiovisual Go vs audiovisual NoGo: 0.219 ± 0.0222 vs 0.134 ± 0.0199; $p = 3.16 \times 10^{-8}$, $g = 0.248$). The presence of the auditory stimuli did not change the magnitude of the Go representation (unimodal Go vs audiovisual Go: $p = 5.87 \times 10^{-2}$), and trivially increased the amplitude at the peak for the NoGo condition (unimodal NoGo vs audiovisual NoGo: $p = 5.51 \times 10^{-3}$, $g=0.124$).

As found when angular distance between the Go and NoGo neurons was 45°, the representation of the Go and NoGo stimuli separated by 30° were clearly distinct and the presence of auditory stimuli did not alter their representations. Yet, the Go stimulus tended to be more strongly and accurately represented than the NoGo, as the amplitude of the population response was greater and recruited a population of neurons more sharply tuned to the visual stimulus presented. The activity in the NoGo profiles of activation straddled the actual stimulus orientation (75°) but was centered around 90°, an orientation corresponding to the NoGo of the previous day of recording.

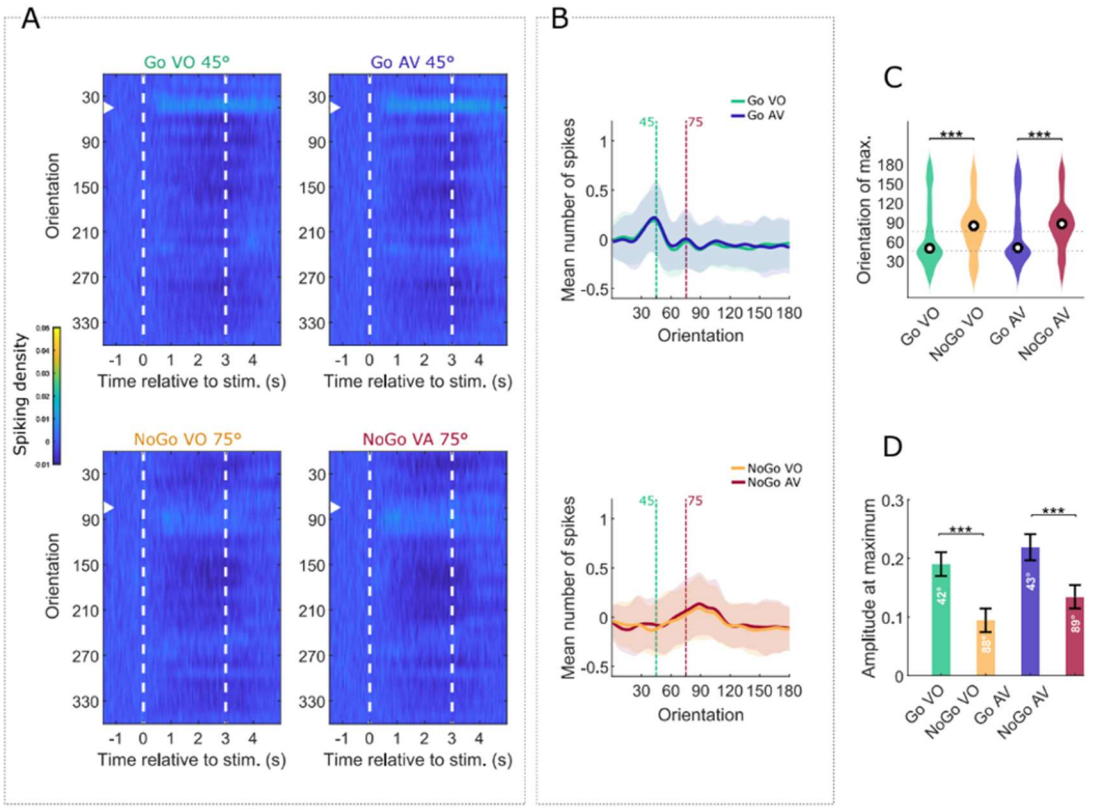

Figure 4-7. V1 representation of orientation at 30º AD with and without auditory stimulation. ***A.*** *Population PSTHs for all four stimulation conditions. Every line corresponds to the spiking activity representing different visual orientations. Warmer colors represent an increase in spiking density compared to the baseline activity. Dashed white lines indicate the starting and ending times of the stimuli. The white arrowheads indicate the orientation of the visual stimulus considered in each plot.* ***B.*** *Profiles of V1 summed activation in the orientation representational space during the first two seconds post stimulus onset, for the Go VO and AV (top) and NoGo VO and AV stimulation contexts (bottom). Solid lines show the average over the 1000 reconstructed trials. Shaded areas show the standard deviation. Dashed green and red lines indicate the orientation of the Go and NoGo stimuli respectively.* ***C****. Distributions of maximum location in the orientation domain.* ***D.*** *Amplitude (mean ± SEM) at the peaks of the activation profiles from B. Numbers in white indicate the orientation of the peak.*

4.4.5 Representation of the orientation in V1 of a Go and NoGo stimuli distant by 25°

We anticipated that the representation of the two stimuli would start to overlap when the angular distance between the Go and NoGo signal fell below 30°. Indeed, 30° is the angle corresponding to the mean half-width of the orientation tuning curves of V1 L2/3 neurons (Niell and Stryker, 2008). As expected, we found a partial overlap between the ensembles activated by the Go (45°) and NoGo (70°) stimuli (Fig 4-8A; Top- and bottom-left panels). The activation profiles for the Go stimulus (45°) presented a single peak of activity close to the Go stimulus orientation (Fig 4-8B; Top), whereas, the activation profile generated by the NoGo stimulus presentation (70°) showed two distinct peaks, one centered on 90° and the other, close to 45°, overlapping the representation of the Go cue (Fig 4-8B; Bottom). In the presence of auditory stimuli, the NoGo representation was noticeably altered. While the two activated subpopulations representing 45° and 90° were still significantly activated, they were modulated in opposite directions (Fig 4-8B; Bottom): sound had a suppressive effect on the population of neurons representing the Go (45°) orientation, and enhanced the response of the subpopulation representing 90°. Strikingly, the actual NoGo orientation (70°) was not represented in the activation profile.

At the single trial level, the orientation profile maximum for the Go stimulus was distributed unimodally in the unimodal and audiovisual conditions, with medians of 49° and 48°, respectively. Although the distributions were different between the unimodal and audiovisual conditions, the effect size was small (Wilcoxon test: $p = 6.11 \times 10^{-5}$; U3 = 0.372, U3 95% CI = [0.343- 0.5]). For the NoGo cue, the maxima of

the single trial orientation profiles presented bimodal distributions in both the unimodal and audiovisual conditions. In the unimodal condition, a fraction of trials represented an orientation of 90º, while a majority of trials indicated 50º (median of all trials 56º). Contrarily, in the audiovisual condition, the majority of the trials represented the orientation of 90°, while the other trials indicated 50° (median of all trials 85°). Both distributions were significantly different (unimodal NoGo vs. audiovisual NoGo, $p = 4.27 \times 10^{-12}$, U3 = 0.655). In addition, Go and NoGo distributions were different in the unimodal condition (unimodal Go vs unimodal NoGo: $p = 1.02 \times 10^{-20}$, U3 = 0.785, 95% CI = [0.740 – 0.811]) and in the audiovisual condition as well, but with a larger effect size (audiovisual Go vs audiovisual NoGo: $p = 2.22 \times 10^{-51}$, U3 = 0.890, 95% CI = [0.870 – 0.910]).

When comparing the amplitude of the peaks of activity within the representation of the unimodal NoGo cues, we found that the amplitude of the 50º peak was significantly higher than the second peak at 88º (mean amplitude ± 95% CI; 50º: 0.655 ± 0.0252; 88º: 0.508 ± 0.0245; $p = 4.30 \times 10^{-16}$; g = 0.367, i.e., small effect size). On the contrary, in the audiovisual condition, the activity evoked by the NoGo cue at the 50º peak was lower than the one evoked at 87º (50º: 0.378 ± 0.0222; 87º: 0.684 ± 0.0264; $p = 2.40 \times 10^{-63}$; g = 0.779, i.e., large effect size). When comparing the amplitudes at the peak of representation of the Go and NoGo signals, a difference of negligible effect size was found in the unimodal and audiovisual conditions (unimodal Go vs unimodal NoGo: $0.711 \pm 2.57 \times 10^{-2}$ vs $0.655 \pm 2.52 \times 10^{-2}$, $p = 2.32 \times 10^{-3}$, g = 0.136; audiovisual Go vs audiovisual NoGo: $0.632 \pm 2.47 \times 10^{-2}$ vs $0.684 \pm 2.64 \times 10^{-2}$, $p = 4.39 \times 10^{-3}$, g = 0.128). Finally, sound had no effect on the magnitude

of the Go and NoGo representations (unimodal Go vs audiovisual Go: $p = 1.57 \times 10^{-5}$ $g = 0.194$; unimodal NoGo vs audiovisual NoGo: $p = 0.109$).

In contrast to larger angular distances, the representations in V1 of a Go and NoGo cue spaced by 25° activated overlapping ensembles of neurons. This was mainly true when the NoGo cue was presented, as the presentation of this stimulus activated a group of neurons that corresponded to the representation of the Go stimulus. The other activated subpopulation represented the orientation of the NoGo stimulus presented in the previous imaging session (NoGo drifting grating 85º). Finally, the neurons representing the actual NoGo orientation (70º) were poorly responsive. The presence of sound influenced the representation. In the unimodal condition, the most active population represented the Go orientation. In the audiovisual condition, the prevailing representation was that of the NoGo cue. Hence, the contrast between the two activated populations increased in the audiovisual condition. This result is compatible with the sound-induced improvement on behavioral performance when the angular distance between the Go and NoGo signals was 25º (Fig 4-4).

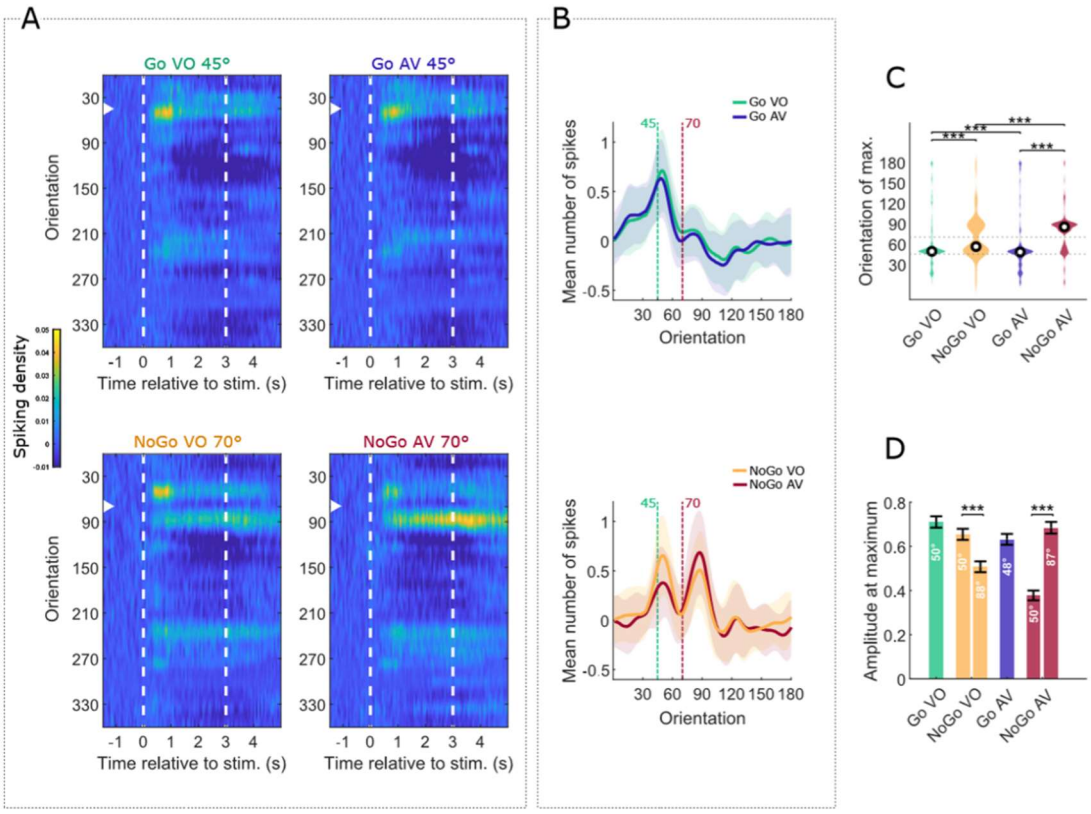

Figure 4-8. V1 representation of orientation at 25º AD with and without auditory stimulation. ***A.*** *Population PSTHs for all four stimulation conditions. Every line corresponds to the spiking activity representing different visual orientations. Warmer colors represent an increase in spiking density compared to the baseline activity. Dashed white lines indicate the starting and ending times of the stimuli. The white arrowheads indicate the orientation of the visual stimulus considered in each plot.* ***B.*** *Profiles of V1 summed activation in the orientation representational space during the first two seconds post stimulus onset, for the Go VO and AV (top) and NoGo VO and AV stimulation contexts (bottom). Solid lines show the average over the 1000 reconstructed trials. Shaded areas show the standard deviation. Dashed green and red lines indicate the orientation of the Go and NoGo stimuli respectively.* ***C****. Distributions of maximum location in the orientation domain.* ***D.*** *Amplitude (mean ± SEM) at the peaks of the activation profiles from B. Numbers in white indicate the orientation of the peak.*

4.4.6 Representation of the orientation in V1 of a Go and NoGo stimuli distant by 20º

The results for the angular distance of 20º between the Go (45°) and NoGo (65°) stimuli mostly overlapped (Fig 4-9A; Top and bottom left panels). The activation profile of the Go stimulus presented a single peak of activity close to the Go stimulus orientation (Fig 4-9B; Top), whereas the activation profile of the NoGo stimulus was broader, encompassing the area activated by the Go and extending towards the NoGo orientation (Fig 4-9B; Bottom). For both stimuli, the unimodal and audiovisual profiles overlapped for the majority of the orientations.

The single trial orientation profile maxima in response to the presentation of the Go stimuli were distributed unimodally with a median of 53° in both the unimodal and audiovisual conditions. The distributions were not significantly different ($p = 0.397$). For the NoGo cue, the single trial orientation profiles maxima showed similar unimodal but broader distributions: the unimodal NoGo trials had a median of 60º, while in the audiovisual NoGo condition the median was 63º ($p = 0.187$). For both the Go and NoGo distributions, no significant differences were found (unimodal Go vs unimodal NoGo: $p = 0.268$; audiovisual Go vs audiovisual NoGo: $p = 0.142$). The amplitude at the peaks of activity evoked by the Go and NoGo cues were significantly different, but of negligible effect size (unimodal Go vs audiovisual NoGo: $0.254 \pm 2.61 \times 10^{-2}$ vs $0.176 \pm 2.26 \times 10^{-2}$, $p = 9.70 \times 10^{-6}$, $g = 0.198$; audiovisual Go vs audiovisual NoGo: $0.200 \pm 2.29 \times 10^{-2}$ vs $0.148 \pm 2.19 \times 10^{-2}$, $p = 0.001$, $g = 0.143$). Sound had no effect on the magnitude of the representations (unimodal Go vs

audiovisual Go: $p = 0.136$; unimodal NoGo vs audiovisual NoGo: $p = 0.077$; fig 4-9D).

Hence, when the angular distance between the Go and NoGo cues was 20°, the subpopulation representing the Go cue was fully recruited by the NoGo cue. Auditory stimuli had no effect on the representations of the Go and NoGo visual stimuli, nor on their magnitude. This result is compatible with the poor behavioral performance and the absence of sound modulation on orientation discriminability at this angular distance (Fig 4-4).

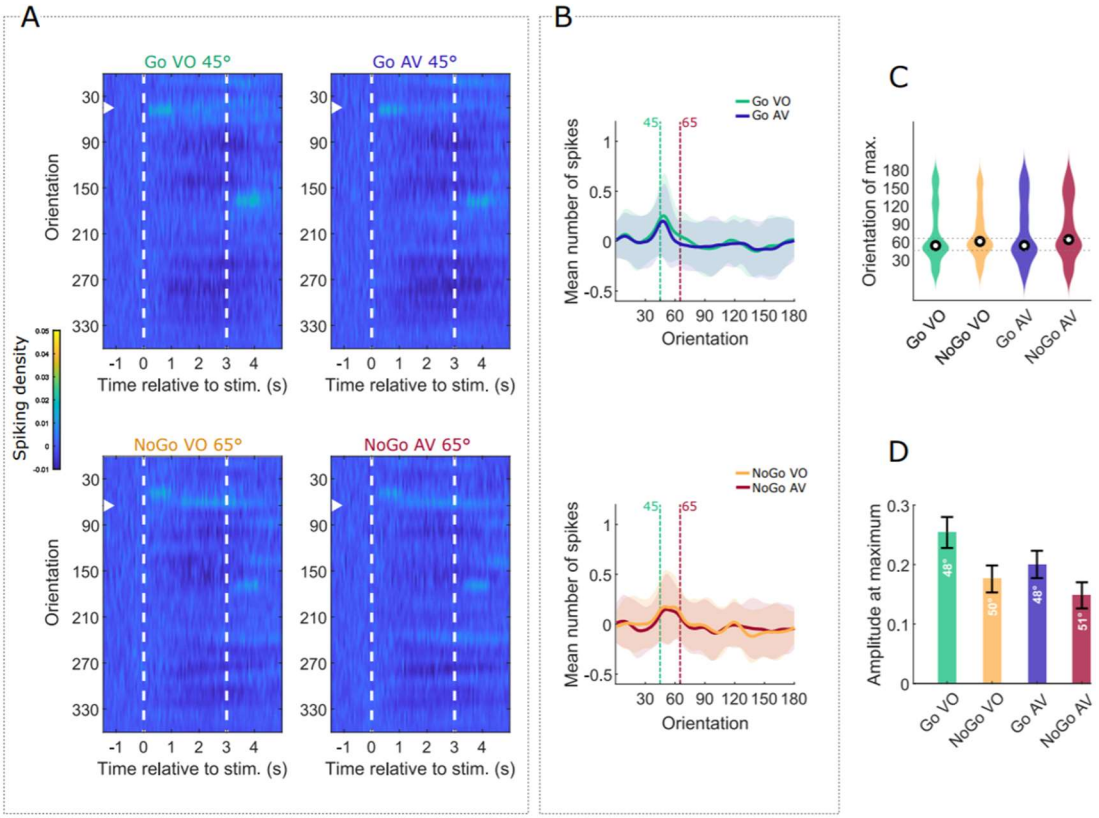

Figure 4-9. V1 representation of orientation at 20º AD with and without auditory stimulation. ***A.*** *Population PSTHs for all four stimulation conditions. Every line corresponds to the spiking activity representing different visual orientations. Warmer colors represent an increase in spiking density compared to the baseline activity. Dashed white lines indicate the starting and ending times of the stimuli. The white arrowheads indicate the orientation of the visual stimulus considered in each plot.* ***B.*** *Profiles of V1 summed activation in the orientation representational space during the first two seconds post stimulus onset, for the Go VO and AV (top) and NoGo VO and AV stimulation contexts (bottom). Solid lines show the average over the 1000 reconstructed trials. Shaded areas show the standard deviation. Dashed green and red lines indicate the orientation of the Go and NoGo stimuli respectively.* ***C****. Distributions of maximum location in the orientation domain.* ***D.*** *Amplitude (mean ± SEM) at the peaks of the activation profiles from B. Numbers in white indicate the orientation of the peak.*

4.4.7 Representation of the orientation in V1 of a Go and NoGo stimuli distant by 15°

The results found for an angular distance of 15° were consistent with the results obtained at 20°AD: the Go and NoGo representations overlapped, making them qualitatively indistinguishable. The activation profiles broadly overarched Go and NoGo orientations (Fig 4-10; Top and bottom panels). There was no significant difference between the distributions of single-trial orientation profile maxima between the unimodal Go and unimodal NoGo responses (unimodal Go median = 52°, unimodal NoGo median = 57°, $p = 5.60 \times 10^{-4}$, U3 = 0.572), but there was a significant yet small difference between audiovisual GoAV and audiovisual NoGoAV (audiovisual Go median = 55°, audiovisual NoGo median = 62°, $p = 2.00 \times 10^{-3}$, U3 = 0.601), reflecting a higher number of trials whose maximum was close to 60° in the audiovisual condition. The distribution of the orientation profile maxima of the Go cue in the unimodal and audiovisual conditions were not differentiable (p= 0.264), nor were the distributions of the NoGo cue in the unimodal and audiovisual contexts (p= 0.179). The amplitudes at the peak were not significantly different between Go and NoGo in the unimodal and audiovisual (unimodal Go vs. unimodal NoGo: $0.290 \pm 2.32 \times 10^{-2}$ vs. $0.248 \pm 2.32 \times 10^{-2}$, $p = 1.17 \times 10^{-2}$, g = 0.113; audiovisual Go vs audiovisual NoGo: $0.289 \pm 2.13 \times 10^{-2}$ vs $0.250 \pm 2.24 \times 10^{-2}$, $p = 1.23 \times 10^{-2}$, g = 0.112), and sound had no effect on the magnitude of the representations (unimodal Go vs audiovisual Go: p = 0.951; unimodal NoGo vs audiovisual NoGo: p = 0.912; Fig 4-10D).

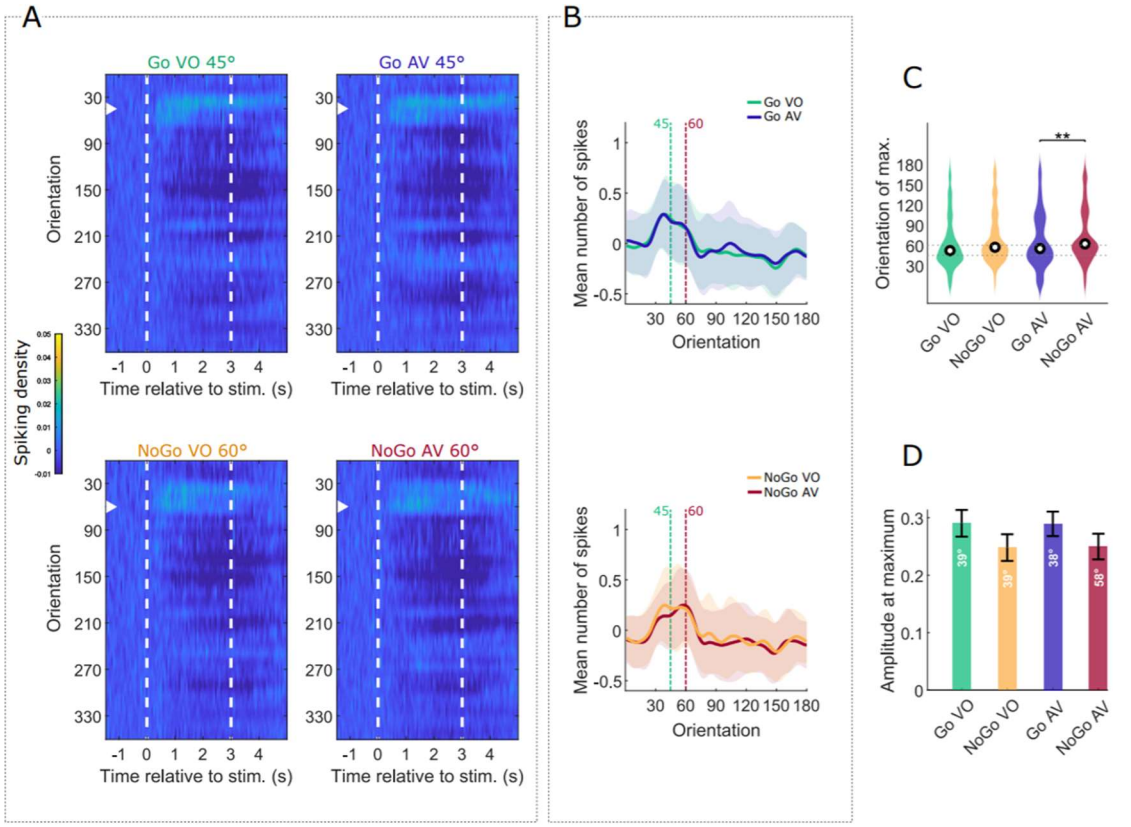

Figure 4-10. V1 representation of orientation at 15º AD with and without auditory stimulation. ***A.*** *Population PSTHs for all four stimulation conditions. Every line corresponds to the spiking activity representing different visual orientations. Warmer colors represent an increase in spiking density compared to the baseline activity. Dashed white lines indicate the starting and ending times of the stimuli. The white arrowheads indicate the orientation of the visual stimulus considered in each plot.* ***B.*** *Profiles of V1 summed activation in the orientation representational space during the first two seconds post stimulus onset, for the Go VO and AV (top) and NoGo VO and AV stimulation contexts (bottom). Solid lines show the average over the 1000 reconstructed trials. Shaded areas show the standard deviation. Dashed green and red lines indicate the orientation of the Go and NoGo stimuli respectively.* ***C.*** *Distributions of maximum location in the orientation domain.* ***D.*** *Amplitude (mean ± SEM) at the peaks of the activation profiles from B. Numbers in white indicate the orientation of the peak.*

4.5 Discussion

The main objective for aim II was to investigate the link between population representation in V1 and visual perception by testing the hypothesis that the sound-induced sharpening of the representation of orientations in V1 can impact angular discrimination.

To accomplish this, the ability of mice to discriminate between two oriented visual stimuli was tested using a Go/NoGo visual discrimination task, in which the angular distance (AD) separating the two stimuli was gradually reduced across daily sessions. Behavioral testing was accompanied by *in vivo* calcium imaging in V1. The behavioral results allowed me to determine the limits of angular discrimination of the mice in the unimodal and audiovisual contexts (Fig 4-4 and Fig 4-5). For angular distances between the Go and NoGo signal equal to or greater than 30º, mice performed almost optimally. In the neural data, the Go and NoGo orientations were represented by non-overlapping neural populations. For large angular distances, sound did not influence the behavioral performance nor the representation of orientations in V1. For angular distances between the Go and NoGo signal equal to or smaller than 20º, the same mice performed close to chance, indicating that mice could not discriminate between the two cues. In line with these results, the representation of the Go and NoGo stimuli almost fully overlapped. Coincident auditory stimuli had no significant impact on V1 representations or on perceptual discrimination. The angular distance of 25º corresponded to the angular discrimination threshold for the task (Fig 4-4 and 4-5). At this specific level of difficulty only, the addition of pure tones along with the visual stimuli consistently improved the behavioral performance.

Accordingly, the overlap between V1 representation of the Go and NoGo orientations was significantly reduced in the audiovisual context. Our findings support the hypothesis that visual discrimination occurs when the neuronal representation in the cortex are non-overlapping.

4.5.1 Effect of learning in V1: the presence of attractors in the representations

One of the most striking findings in this study was the discrepancy between the orientation of the visual cue and the orientation that was represented by neural populations in V1. For example, when the angular distance between the two cues was 45°, the NoGo stimulus (90°) activated mostly the neurons representing 135°. Interestingly, 135º corresponded to the orientation of the NoGo signal used during training. Similarly, when the angular distance between the two cues was 30º, the population activity was centered on 90° (orientation of the NoGo cue of the preceding session), instead of 75°. Finally, when the angular distance between the two cues was 25°, the NoGo (70°) activated neurons representing orientations at 45° (corresponding to the orientation of the Go cue) and 85°, the orientation of the NoGo cue of the previous session. These results strongly suggests that V1 tends to represent expected or behavior-relevant orientations more strongly than actual orientations. Further analyses are needed to explore this phenomenon in our dataset and further studies are needed to understand its substrate. Some evidence suggests that at least a part of it is due to plasticity, as opposed to top down, task-related online modulation. First, the represented NoGo was that of the previous day, which can only be a consequence of learning and plasticity occurring either in V1 or in a structure that modulates V1 in an

orientation-specific way. Second, this shift of the representation towards the learned Go or NoGo was observed during the tuning curve trials that were performed after the task itself, suggesting that the effect is either indeed due to long term plasticity (over a day) or at least it persists for 15 minutes after the task (Fig 4-11; Poort et al., 2015; Jurjut et al., 2017; Pakan et al., 2018)

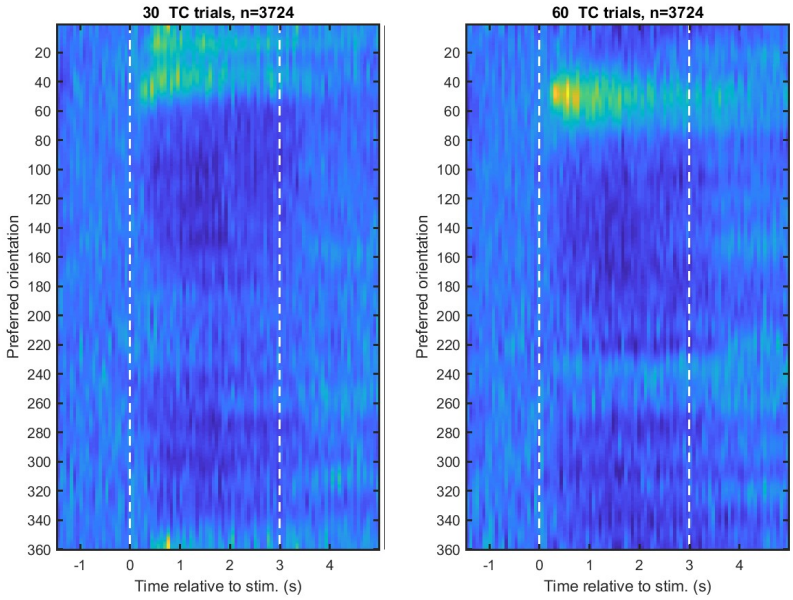

Figure 4-11. The Effects of the Go/NoGo stimuli on V1 representation of orientation in tuning curve trials. ***(Left panel)*** *Population PSTHs for 30º tuning curve trials. Every line corresponds to the spiking activity representing different visual orientations. Warmer colors represent an increase in spiking density compared to the baseline activity. Dashed white lines indicate the starting and ending times of the stimuli.* ***(Right panel)*** *Population PSTHs for 60º tuning curve trials.*

4.5.2 Linking evoked activity in V1 and perceptual orientation discrimination

In future analyses, other assessments will be used to further elucidate the causal link existing between V1 activity and orientation discriminability: [1] the

quantification of the broadness of the peaks in the activation profile as an assessment of representation precision, [2] the second to second analysis of the mice licking response during the task, along with the associated representation of the orientation in V1 neurons. Another strategy will involve performing the same analytical procedures used in this aim after sorting the trials by response types (e.g. Hit, Miss, CR, and FA) in order to link specific activation within V1 populations, that represent a certain orientation, with the likelihood of making a particular decision among the four available outcomes. For example, we could measure for all the trials where the mice make a correct rejection (CR) and when the angular distance between the Go and NoGo signals is 25^{o}, if the neural populations that represent the 90^{o} orientation (as seen in Fig 4-8B; Bottom) prevail more (i.e. is represented more) compared to the other population of neurons that represent the Go orientation (Fig 4-8B; Bottom). In conclusion, these analyses will provide additional information on the link between V1 representation and mouse discrimination.

Chapter V

5.1 Adaptation to the behavioral goal of the representation of the orientation and direction of the visual stimuli in a cross-modal discrimination task

5.2 Background

The previous chapter investigated the effects of auditory tones on the representation of orientation and direction of the visual stimulus by V1 L2/3 neurons, and its' impact on the perception of orientation. However, both studies demonstrated the effects of sounds on visual processing in mice passively receiving the auditory cues. Therefore, the expression of sound modulation in V1 when mice are engaged in an audiovisual task in which both the visual and the auditory stimuli contain relevant behavioral information (e.g., information on reward retrieval), remains to be determined. As shown in the introduction of this thesis, there is a growing stream of evidence that early visual areas in rodents, non-human, and human primates receive multimodal information. However, the role of this modulation remains elusive. The most prominent hypothesis is that neurons in early visual cortices are contextually modulated by sounds to facilitate perception, cognition, and behavior (Petro et al., 2017). In mice and anesthetized cats, it was found that sound modulation in V1 depends on the ethological characteristics of the audiovisual inputs (Deneux et al., 2019; Meijer et al., 2017; Chanauria et al., 2019). However, there is no direct evidence, to my knowledge, that sound modulation in V1 can adapt visual processing to the behavioral goals (context-dependent modulation), by favoring the processing of a specific type of information over another.

5.3 Hypothesis

For this aim, I tested the hypothesis that the representation of orientation and direction in V1 is dynamically adapted to the behavioral goal by enhancing the representation of valued information and suppressing the representation of non-pertinent or conflicting information.

5.4 Materials and methods

I performed two-photon calcium imaging in awake mice while executing a Go/NoGo audiovisual discrimination task. The task was designed to determine the representation in V1 of the same oriented stimuli when the visual cues were presented alone, or in competition with auditory information.

5.4.1 Animals

For technical details on the mice used in this experiment, see section 2.1 of Chapter II.

5.4.2 Stereotaxic surgery

For technical details on all surgical procedures performed in this experiment including head bar implants, virus injections, and coverslip implants, refer to section 2.2 of Chapter II.

5.4.3 *In vivo* calcium imaging

For technical details see section 2.3 of Chapter II.

5.4.4. Audiovisual stimuli

A gamma-corrected 40-cm diagonal LCD monitor was placed 30 cm from the eye contralateral to the recording site 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 (The MathWorks) using the Psychtoolbox (Brainard 1997).

To investigate how cross-modal detection influences V1 processing, mice were trained to perform in a cross-modal discrimination task in which both the auditory and visual stimuli were indicators of the possibility to obtain a water reward. Prior to behavioral training, mice were water-restricted for 5 days as described previously in section 4.3.4 of Chapter IV.

Following water restriction, mice were first trained to perform in a Go/NoGo visual discrimination task. During the Go trials mice were shown a 45º drifting grating with the following properties (temporal frequency: 2 Hz, spatial frequency: 0.04 cycle/deg, contrast: 75%, duration: 3 sec, intertrial interval: 3 sec), whereas in the NoGo trials mice were shown a 135º drifting grating with the same properties. For Go trials, water delivery was triggered when the mice successfully licked for water during the response window. On the contrary, for NoGo trials, a lick response during the response window (CR) triggered a timeout. The intertrial interval (ITI) following a correct trial was 3 seconds. When the response was incorrect (no lick response for a Go trial or a lick response for a NoGo trial), a 9 sec timeout was implemented (+ the 3 sec ITI).

Upon successful training on the Go/NoGo visual discrimination task (> 90% correct response over 300 trials), mice were trained on a similar Go/NoGo auditory discrimination task. During Go auditory trials, mice were presented with a 10 kHz tone (78 dBA; duration: 3 sec). In NoGo trials, mice were presented with a 5 kHz pure tone (78 dBA; duration: 3 sec). Reward schedules were the same as outlined in aim II on the Go/NoGo visual discrimination task.

In the cross-modal discrimination task, mice that were previously trained for the visual discrimination task and the auditory discrimination task were presented with either visual-only or auditory-only cues for the first 30 trials, followed by the simultaneous presentation of auditory and visual cues. The block of unimodal cues indicated the rewarded modality of the following audiovisual bout. For example, if unimodal visual cues were presented prior to the audiovisual task, then the water reward was paired with the value (Go or NoGo) of the visual cues presented in audiovisual block. The same was true for the auditory priming blocks (Fig 5-1). During audiovisual stimulus presentations, trials were either congruent, in which both the visual and the auditory signals matched in value (Go/Go or NoGo/NoGo), or they were conflicting (signal mismatch between visual and auditory cues, Go/NoGo or NoGo/Go). Reward was always contingent on the primed modality (Fig 5-1).

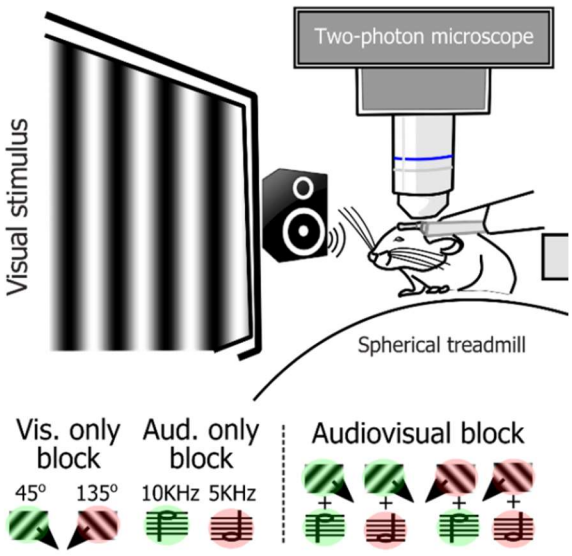

Figure 5-1. Go/NoGo cross-modal discrimination task. *(Top). Schematic of the two-photon calcium imaging set up. Visual and auditory stimuli was presented contralateral to the hemisphere where GCAMP6f was injected in V1 (left hemisphere). (Bottom) Stimulus presentation scheme. Visual and auditory stimuli were presented in visual-only, auditory-only, or audiovisual blocks. Go visual and auditory stimuli are shaded in green. NoGo visual and auditory stimuli are shaded in red.*

5.4.5 Data analysis

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

Acquisition and analysis of behavioral data. Behavioral data was recorded in WinEDR (John Dempster, University of Strathclyde), a data collection software used to track and record the following information related to the behavioral task: licking activity, the response window, and task outcome (e.g. water delivery and timeout). All data was recorded in seconds. For each trial, across all mice, the following behavioral measures were stored in a Microsoft Server SQL database located on a local server (HPE ProLiant ML350 Gen10 Solution): stimulus presented, behavioral response (i.e. lick or no lick), locomotion, and eye tracking.

To determine that the mice were taking into account the value (Go or NoGo) of both the visual and auditory cues during the cross-modal discrimination task, we compared their probability of licking for water for the conflicting trials when the mice

were primed for the visual cue at the start of the audiovisual task and when the mice were primed for the auditory cue at the start of the audiovisual task. To compare the differences in the probability of licking for water under the 8 stimulus combinations: Vis_{Go}, Vis_{NoGo}, Aud_{Go}, Aud_{NoGo}, $Vis_{Go} + Aud_{Go}$, $Vis_{NoGo} + Aud_{NoGo}$, $Vis_{Go} + Aud_{NoGo}$, and $Vis_{NoGo} + Aud_{Go}$, the percentage of trials in which the animal made the correct response was divided by the total number of trials presented, and averaged across animals.

The impact of cross modal discrimination on the response (lick onset during Go trials and lick offset during NoGo trials) was quantified using the average lick trace on a single trial recorded for each animal during the 8 different stimulus presentations. The mean trace across trial was fitted by a sigmoid function and defined by a baseline value, a ramp, a half rise time value, and a max/plateau value. The half rise times for each stimulus condition were averaged across animals.

To test the potential effects of arousal and locomotion on behavioral performance in the unimodal and audiovisual context, the pupil area and locomotion for each stimulus condition was quantified and compared.

Neural data analysis. Raw calcium imaging signals recorded during the task, along with orientation tuning curves for each neuron was preprocessed similarly to the calcium data in aim I of chapter III (see section 3.3.7, paragraphs 1 and 2). dF/F values were then stored in a SQL database and used for training the Shallow Neural Network (SNN). SNNs are computational tools designed for pattern detection. They consist of an input layer where data is inputted, an output layer providing the classification, and one or more hidden layers. SNNs are trained to classify inputs according to two or

more target classifications. Trained SNNs provide the probability that a unique/novel input belongs to one of the categories that they were trained on and learned to recognize (Fig. 5-2C). We used those properties of SNNs to determine the orientation represented by V1 for a single trial. Indeed, when an oriented stimulus is presented in the visual field, a subpopulation of V1 neurons whose tuning properties match the properties of the stimulus are activated. Since the distribution of preferred orientation of V1 neurons is continuous (Fig. 5-2A), any oriented visual stimulus is represented in V1 by the activation of a specific subpopulation of neurons. However, orientation representations overlap due to the wide tuning width of V1 neurons in the mouse cerebral cortex (± 30º; Fig 5-2B; Niell and Stryker, 2008). Therefore, in theory, a shallow neural network trained to recognize two orientations (e.g. 30º and 60 º) will classify the pattern evoked by a stimulus equidistant of the two trained stimuli (45º) as 50% for 30º and 50% for 60º (Fig 5-2B, C). From this result we can recover the orientation indicated by the population by computing the circular mean of the SNN output (Fig 5-2E).

We therefore trained SNNs to classify the neural patterns of subpopulations of V1 neurons (n = 250 neurons) randomly selected from the neural database of imaged neurons. Each SNN was composed of an input layer of 250 neurons fully connected to a layer of 10 hidden neurons which were in turn fully connected to an output layer of 12 neurons (Fig. 5-2D). The network was trained using trials made of the random combination of the responses (mean dF/F across the visual stimulus presentation) of each V1 neurons to the presentation of one of the twelve evenly spaced oriented visual stimuli (0º, 30º, 60º, 90º, 120º, 150º, 180º, 210º, 240º, 270º, 300º, 330º) presented at

the end of each imaging session (144 trials per orientation). Training was performed with 70% of the trails, validated with 15% of the trials and tested with the remaining 15%. At the end of training, each SNN was able to accurately classify all the presented trials ($p > 0.99$). We then used the SNN to classify trials collected from the same V1 neurons when presented with oriented stimuli that the SNN was not trained to recognize (45^o and 135^o), unimodal auditory stimuli (pure tones of 5 kHz and 10 kHz), and audiovisual stimuli (Fig. 5-2E). The orientation represented by the selected population of V1 neurons was then calculated from the circular mean of the SNN output. The accuracy (precision index) of this represented orientation was then calculated as the projection of the circular mean vector onto the radial axis of the visual stimulus orientation (Fig. 5-2F). To determine the variability of the representation of the non-training stimuli across the population of imaged V1 neurons, the procedure was repeated 1,000 times, taking each time a novel combination of 250 neurons from the database of thousands of imaged V1 neurons (Fig. 5-2G).

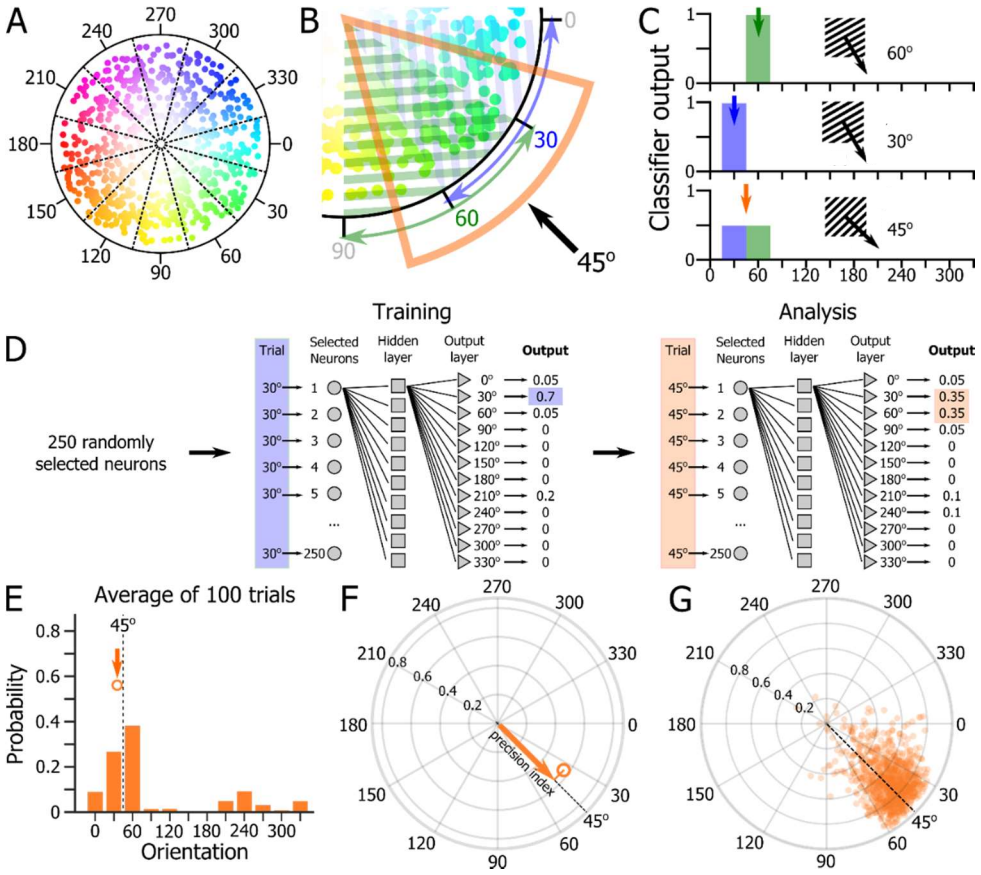

Figure 5-2. Analysis of the representation of the orientation of drifting gratings in V1 using a Shallow Neural Network (SNN). ***A.*** *Schematic representation of the distribution of preferred orientation in V1. Dots represent V1 neurons; the hue represents the neuron's preferred orientation; the hue intensity represents the amplitude of the response to the preferred stimulus.* ***B.*** *Neurons preferentially recruited by a 30º (blue hatching), 45º (orange sector) and 60º (green hatching) drifting gratings.* ***C.*** *Theoretical response of a SNN trained to discriminate the 30º drifting grating (top panel) and the 60º drifting grating (middle panel) when it is presented with the response of the same neuronal population for the presentation of a 45º drifting grating (bottom panel).* ***D.*** *Schematic representation of the SNN training and testing.* ***E.*** *Average across 100 trials of the output of a trained SNN for the presentation of a 45º drifting grating. Dot: orientation and length of the circular mean of the SNN output.* ***F.*** *Orientation index: projection of the circular mean onto the visual stimulus orientation axis.* ***G.*** *Responses to the presentation of unimodal 45º drifting gratings of 1,000 SNNs, each built from 250 neurons randomly selected from the population of imaged neurons.*

5.4.6 Statistics

Behavioral data. A normality test (Shapiro-Wilk) was always performed before each test to assess the normality of the data. When the sample data was not normally distributed, attempts were made to transform the skewness and normalize the data using a Van der Waerden normalization (rank-based normalization method). In cases where the data could not be normalized, non-parametric test were used for analysis. When appropriate, a one-way or two-way ANOVA or a Kruskal Wallis one-way ANOVA on ranks was used to determine statistical significance among groups tested. T-test (parametric) or Wilcoxon test (non-parametric) were used to test if the mice performed above chance level for each the stimulus conditions. All pairwise multiple comparisons were performed using Tukey's test (ANOVA) or Dunn-Sidak methods (Kruskal-Wallis) with α level of $P < 0.05$.

Neural data. For all neural data analyses on the outputs of the trained SNNs, a Kruskal Wallis one-way ANOVA on ranks was used to extract statistically significant differences between the stimulus groups. Follow-up pairwise multiple comparisons were performed using a Tukey-Kramer post hoc comparison with α level of $P < 0.05$.

5.5 Results

5.5.1 Behavioral performance in a cross-modal discrimination task

We designed a Go/NoGo cross-modal discrimination task in which water restricted mice compared the reward values of a visual cue and an auditory cue to guide their responses. To determine the behavioral performance when the auditory and

visual stimuli conveyed important information regarding the reward (Go signal), in the unimodal and congruent or conflicting cross-modal conditions, we compared the performance of the mice (i.e., probability of licking for water) for all the trials when the auditory and visual stimuli were presented in the unimodal and audiovisual context (Fig 5-3). In the unimodal condition, mice licked systematically whenever the Go signal was presented (auditory: 92% hit rate (n = 10 mice), $p < 0.0001$; visual: 96% hit rate (n = 10 mice), $p < 0.0001$), and avoided licking in the presence of the NoGo signal (auditory: 29% hit rate (n = 10 mice), $p < 0.0001$; visual: 32% hit rate (n = 10 mice), $p = 0.0004$). In the cross-modal condition, the performance of the mice at refraining from licking was improved when the auditory and visual NoGo signal were presented simultaneously, compared to the unimodal NoGo conditions (19% hit rate (n = 10 mice), $p_{NoGo/NoGo - Aud.\ NoGo} = 0.0100$, $p_{NoGo/NoGo - Vis.\ NoGo} = 0.0009$). We did not find an improvement in behavioral performance when the two Go signals were presented together, compared to the two unimodal conditions (98% hit rate (n = 10 mice), $p_{Go/Go - Aud.\ Go} = 0.8109$, $p_{Go/Go - Vis.\ Go} = 0.2415$), likely because mice perform almost perfectly in the unimodal contexts. When the visual and auditory visual cues were in conflict, mice clearly chose to lick (by default) (Go_{visual}/ $NoGo_{auditory}$: 79% hit rate (n = 10 mice), $p < 0.0001$; $NoGo_{auditory}/Go_{visual}$: 82% hit rate (n = 10 mice), $p < 0.0001$). Even though the mice' choice in the conflicting cross-modal context was not as ascertained as the Go choice in the unimodal contexts, this result strongly suggested that the mice responded preferentially to the Go signal.

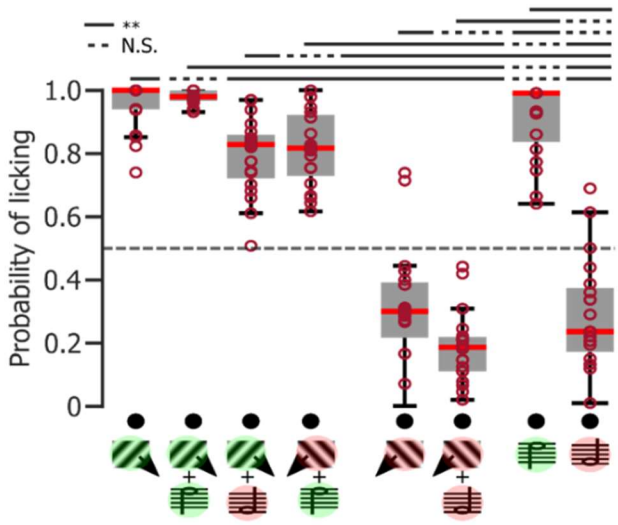

Figure 5-3. Sound modulation of behavioral performance on a cross-modal task. *Boxplots of the probability of licking for water as a function of the visual and auditory stimulus pairs presented during the cross-modal discrimination task. Solid black bars indicate significant comparisons determined by a Friedman test followed by a Tukey-Kramer pairwise comparison. Dotted lined indicates no significant comparison. Black circles for each stimulus pair indicate significant probability of licking from chance (dotted horizontal midline).*

5.5.2 Sound shapes licking onset and offset during the cross-modal behavioral task

Since combining auditory and visual cues was shown to improve decision time (Arnold et al., 2010), we determined the latency at which mice changed their licking behavior in response to the auditory and/or visual cues. The lick onset (Go response) and lick offset (NoGo response) were measured in the unimodal and audiovisual contexts. Mice started liking earlier during the audiovisual trials compared to the unimodal trials (*Friedman test followed by Tukey-Kramer post hoc comparison $p < 0.01$; Fig 5-4). In addition, lick offset (the cessation of licking) was lower in the NoGo

congruent trials compared to the NoGo unimodal visual-only trial (* Mann-Whitney U test $p < 0.05$).

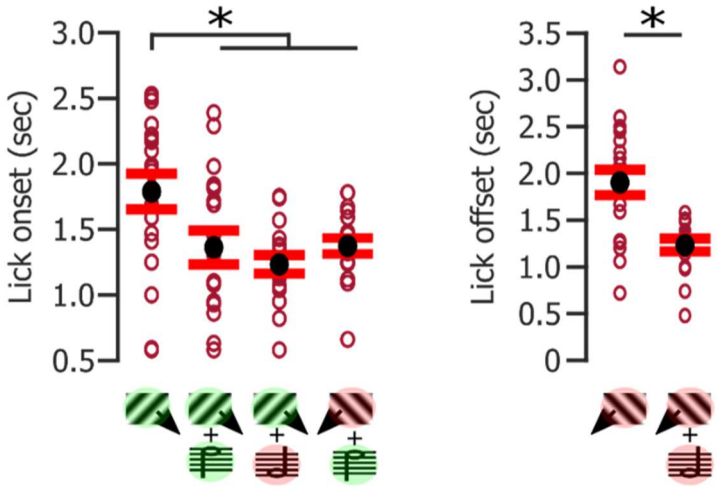

Figure 5-4. The effects of audiovisual cues on lick onset and offset during the cross-modal task. *Boxplots of the changes in lick response to the presentation of Go/NoGo unimodal and audiovisual cues. The left panel depicts the lick onset (latency to start licking for water) for the Go visual and auditory unimodal trials, the Go congruent (Vis. Go + Aud. Go), and the Go conflicting trials (Vis. Go + Aud. NoGo or Vis. NoGo + Aud. Go). The right panel depicts the lick offset (cessation of licking) for the NoGo visual trials, along with the NoGo congruent trials.*

5.5.3 Locomotion and arousal are similar in visual-only and audiovisual contexts

Because locomotion and arousal were both previously found to modulate the response of V1 neurons to visual stimuli (McGinley et al. 2015; Niell and Stryker 2010; Polack et al. 2013; Vinck et al. 2015), a test was performed to determine if the simultaneous presentation of a sound with the test stimulus triggered an increase in locomotion or arousal. We found that the probability of locomotor activity did not differ during unimodal and audiovisual blocks (ANOVA; $p = 0.12$ Fig 5-5; Left).

Similarly, the pupil size was similar for the different stimulus pairs (Friedman test; p = 0.06; Fig 5-5; Right).

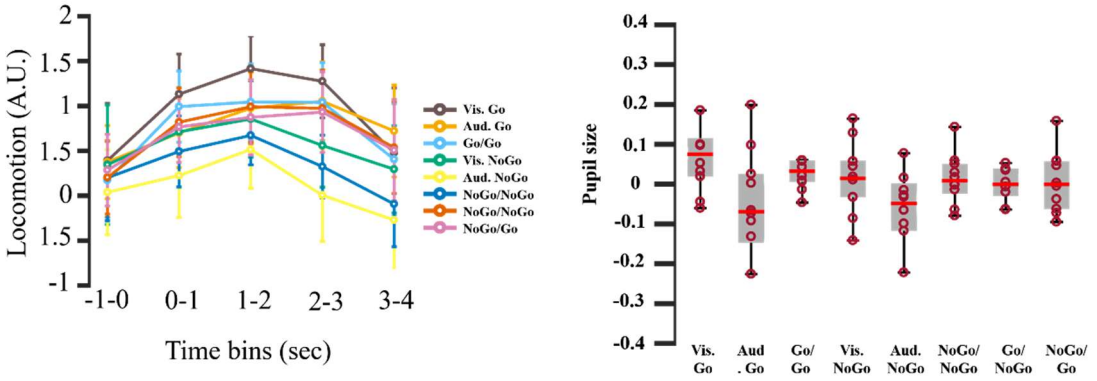

Figure 5-5. Changes in arousal and locomotion during the cross-modal discrimination task. *Left panel depicts the locomotion across mice for each of the stimulus conditions. The right panel indicates the pupil size changes during unimodal and audiovisual trials.*

5.5.4 Modulation by sound of the representation of orientation in V1 in naïve mice.

For the analysis of the neural data, our first goal was to determine if it was possible to assess a modulation of the representation of oriented stimuli in V1 using the SNN approach (see Material and Methods and Fig 5-2). We tested the validity of this new approach by determining if it was possible to identify, using SNNs, the modulation by sound of the representation of orientation in V1 (McClure and Polack, 2019; see Aim #1). We trained SNNs on the responses to the presentation of 12 oriented drifting gratings (0º, 30º, 60º, 90º, 120º, 150º, 180º, 210º, 240º, 270º, 300º, 330º) of a random selection of 250 V1 neurons chosen from same database as the one analyzed in aim I (i.e. mice not trained to perform the cross-modal discrimination task; McClure and Polack, 2019). Each trained SNN was then asked to classify the

responses of its set of neurons to the presentation of the two visual stimuli whose orientations were not included in the training orientations (45° and 135° drifting gratings). The two visual stimuli were either presented in isolation (unimodal trials) or in the presence of one of two sounds (5 kHz and 10 kHz pure tones). The two tones were also presented in isolation. The output of each SNN was transformed into a vector whose orientation indicated the orientation represented by the selected neurons and whose length represented the certainty (from 0 to 1) on this measure. To determine the variability of the representation of the visual stimulus orientation across V1 subpopulations, the process was repeated 1,000 times. In the unimodal visual and audiovisual conditions, the output vectors of the SNNs were pointing in the direction of the visual stimulus (For 45°: unimodal visual: 42.3° ± 1.0, visual + 5 kHz tone: 47.7° ± 0.8, visual + 10 kHz tone: 48.0° ± 0.7; Fig 5-6A; For 135°: unimodal visual: 105.4° ± 4.4, visual + 5 kHz tone: 123.0° ± 2.7, visual + 10 kHz tone: 127.9° ± 2.7; Fig. 5-6B). In the unimodal auditory condition, the output vector was weakly attracted to 90°, an orientation equidistant to the two test stimuli (5 kHz tone: 84.6° ± 5.0, 5 kHz tone: 89.0° ± 4.3; Fig 5-6C).

We then compared the accuracy of the representation of the visual stimulus in the unimodal and audiovisual conditions, measured as the precision index of the output vectors. We found that the precision index of the representation of the 45° stimulus was significantly smaller in the unimodal condition than in the audiovisual conditions (Kruskal-Wallis test followed by Tukey-Kramer post-hoc comparison $p_{unimodal/audiovisual} < 0.0001$). Moreover, the precision index for the representation of the 45° stimulus was similar in the two audiovisual conditions (Tukey-Kramer post-hoc comparison

$p_{audiovisual\ 5\ kHz/\ audiovisual\ 10\ kHz} = 0.37$; Fig 5-6D). We found similar results when comparing the representation of the 135° stimulus in the unimodal visual and audiovisual conditions (Kruskal-Wallis test followed by Tukey-Kramer post-hoc comparison $p_{unimodal/audiovisual} < 0.0001$, $p_{audiovisual\ 5\ kHz/\ audiovisual\ 10\ kHz} = 0.36$; Fig 5-6E).

Altogether, the SNN results replicated the previous findings that orientation is better represented in V1 of naïve mice in the presence of sound (McClure and Polack, 2019; see Aim #1). This also demonstrated that the SNN approach was able to accurately detect modulations of the representation in V1 of orientation of the visual stimulus.

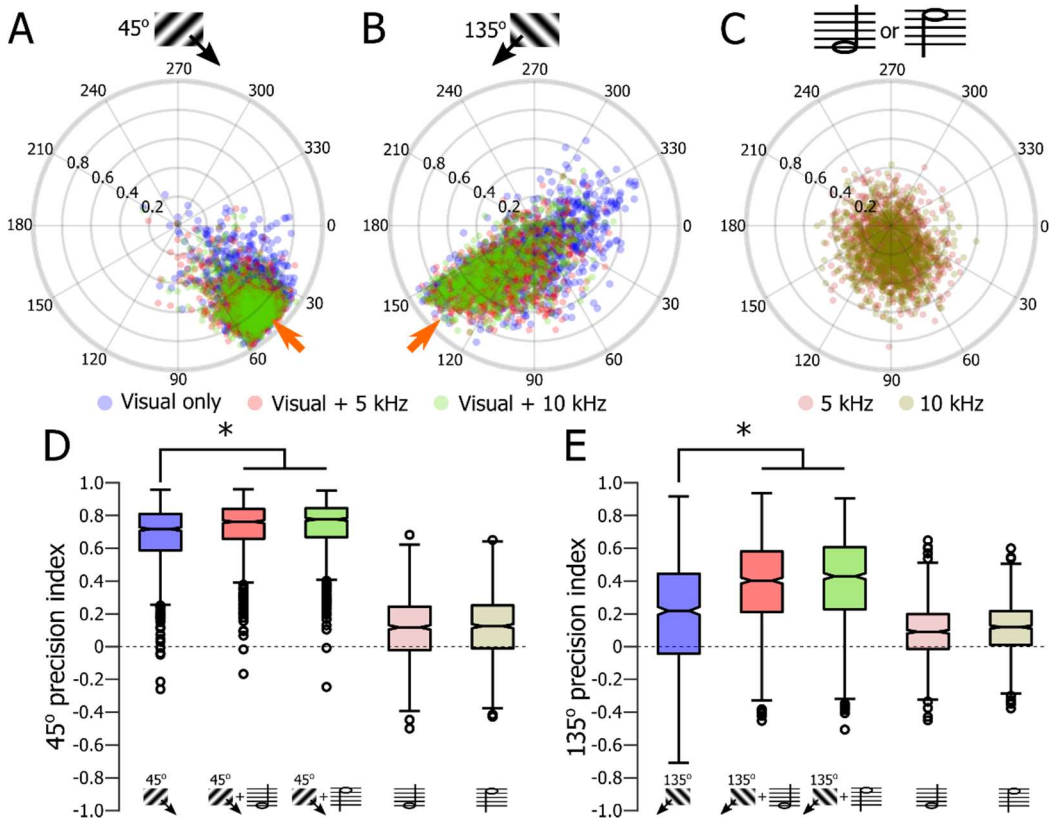

Figure 5-6. Modulation by sound of the population response of V1 in naïve mice. ***A***. *Population response of 1,000 V1 neuronal subpopulations to the presentation of a 45º drifting grating (orange arrow) in the visual only (unimodal, blue) and audiovisual (5 kHz tone, red; 10 kHz, green) contexts.* ***B***. *Same representation as in (A) the presentation of a 135º drifting grating.* ***C***. *Same representation as in (A) for unimodal auditory stimuli (5 kHz tone, dark red; 10 kHz, dark green).* ***D***. *Distribution of the 45º precision indexes of the 1,000 V1 neuronal subpopulations presented in (A) and (C). *Kruskal-Wallis test followed by Tukey-Kramer post-hoc comparison (P<0.0001).* ***E***. *Distribution of the 135º precision indexes of the 1,000 V1 neuronal subpopulations presented in (B) and (C). *Kruskal-Wallis test followed by Tukey-Kramer post-hoc comparison (P < 0.0001).*

5.5.5 Modulation by sound of the representation of orientation in V1 in mice performing a cross-modal discrimination task.

Next, we tested the hypothesis that the representation of the orientation of the visual stimuli is adapted to the behavioral goal by optimizing sensory processing to

the behavioral goal. In regards to the cross-modal discrimination task, we hypothesized that the representation of the Go signals (45^{o} drifting grating) would be improved and the representation of the NoGo signals (135^{o} drifting grating) would be attenuated in order to facilitate the detection of the Go signal in either the visual or auditory condition. Using the SNN approach, we characterized the representation of the Go and NoGo signal in mice performing the unimodal and audiovisual section of the cross-modal discrimination task (Fig 5-7A, B), as well as the representation of the two sounds in V1 (Fig 5-7C). We found that the representation of the visual Go stimulus was improved in the audiovisual context compared to the unimodal context (Fig 5-7D; Kruskal Wallis test followed by Tukey-Kramer post-hoc comparison $p < 0.0001$). More interestingly, the representation of the NoGo visual cue was degraded when it was presented in the audiovisual context (Fig 5-7E; Kruskal-Wallis test followed by Tukey-Kramer post-hoc comparison $p < 0.0001$).

Altogether, those results suggest that in a context in which the visual and auditory cues are in competition, the representations of the important stimuli (in this case the Go signal) are improved while the representation of stimuli less important for the decision (the NoGo signal) are degraded.

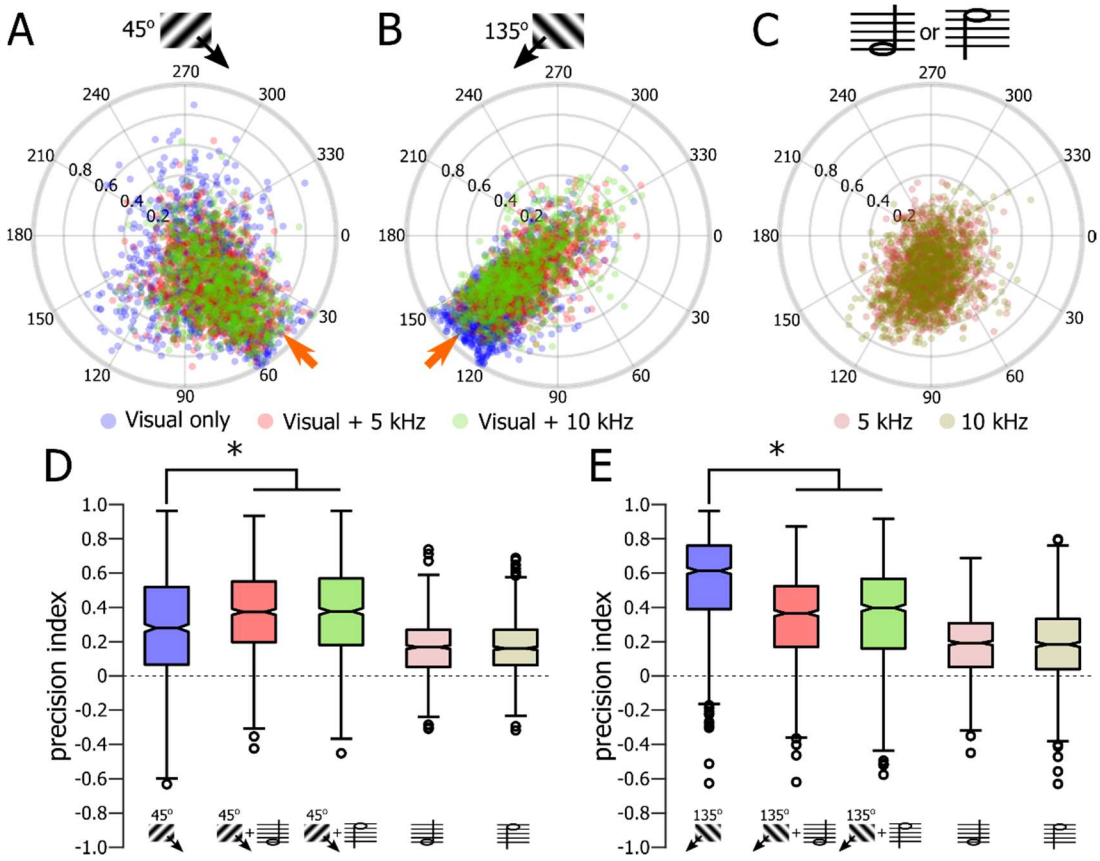

Figure 5-7. Modulation by sound of the population response of V1 in mice actively performing the task. ***A.*** *Population response of 1,000 V1 neuronal subpopulations to the presentation of a 45º drifting grating (orange arrow) in the visual only (unimodal, blue) and audiovisual (5 kHz tone, red; 10 kHz, green) contexts.* ***B.*** *Same representation as in (A) the presentation of a 135º drifting grating.* ***C.*** *Same representation as in (A) for unimodal auditory stimuli (5 kHz tone, dark red; 10 kHz, dark green).* ***D.*** *Distribution of the 45º precision indexes of the 1,000 V1 neuronal subpopulations presented in (A) and (C). *Kruskal-Wallis test followed by Tukey-Kramer post-hoc comparison ($P < 0.0001$).* ***E.*** *Distribution of the 135º precision indexes of the 1,000 V1 neuronal subpopulations presented in (B) and (C). *Kruskal-Wallis test followed by Tukey-Kramer post-hoc comparison ($P < 0.0001$).*

5.5.6 Modulation by sound of the representation of orientation in V1 in trained mice passively processing the audiovisual stimuli.

To determine if the differential modulation of the Go and NoGo visual cues found in V1 of mice performing the cross-modal discrimination task was dynamic (i.e., provided on line by 'top-down' inputs) or the result of a plasticity mechanism

occurring during training, we analyzed the representation of the Go and NoGo visual stimuli in mice trained for the task but satiated for water and therefore passively receiving the cues presented during the task. Those neuronal responses were analyzed using the SNN approach. The modulation profile in the audiovisual context of the Go and NoGo visual signal was different from that of mice actively performing in the cross-modal task. The improvement of the representation of the Go visual signal was significant in passive mice only when combined with the NoGo auditory signal (Fig 5-8A, D; Kruskal-Wallis test followed by Tukey-Kramer post-hoc comparison $p = 0.01$ and $p = 0.96$ respectively), while the representation of the NoGo visual signal was improved in both audiovisual conditions, similarly to what was shown to occur in naïve mice (Fig. 3-2; McClure and Polack, 2019). Moreover, the improvement in the presence of the Go auditory signal was greater than the improvement with the NoGo auditory signal (Fig. 5-8A, D; Kruskal-Wallis test followed by Tukey-Kramer post-hoc comparison $p_{\text{NoGo unimodal – NoGo/NoGo}} < 0.0001$, $p_{\text{NoGo unimodal – NoGo/Go}} < 0.0001$, $p_{\text{NoGo/Go – NoGo/NoGo}} < 0.0001$).

Altogether, those results suggest that the differential modulation of the Go and NoGo signals found in mice actively performing the task is likely to be a dynamic modulation. Further investigation will be necessary to determine if the difference in audiovisual representation are the result of learned associations.

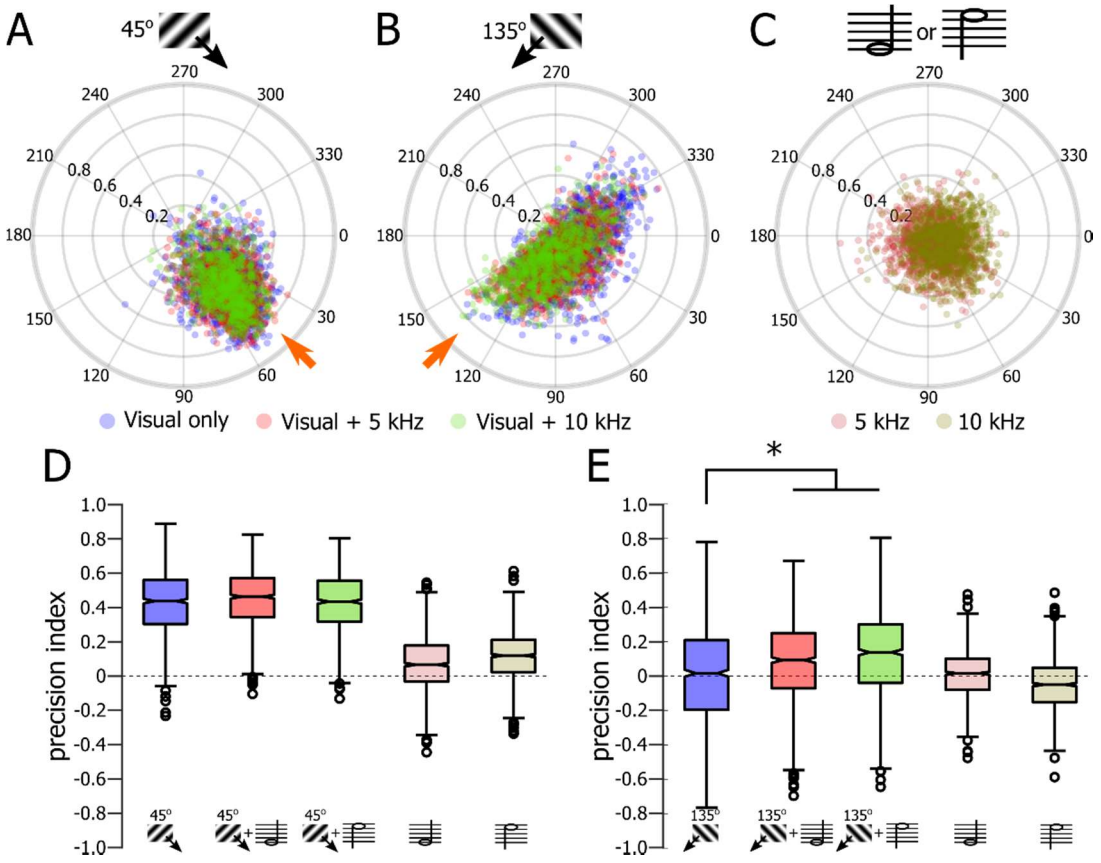

Figure 5-8. Modulation by sound of the population response of V1 in mice trained to the task but passively exposed to the auditory and visual stimuli. ***A.*** *Population response of 1,000 V1 neuronal subpopulations to the presentation of a 45° drifting grating (orange arrow) in the visual only (unimodal, blue) and audiovisual (5 kHz tone, red; 10 kHz, green) contexts.* ***B.*** *Same representation as in (A) the presentation of a 135° drifting grating.* ***C.*** *Same representation as in (A) for unimodal auditory stimuli (5 kHz tone, dark red; 10 kHz, dark green).* ***D.*** *Distribution of the 45° precision indexes of the 1,000 V1 neuronal subpopulations presented in (A) and (C).* ***E.*** *Distribution of the 135° precision indexes of the 1,000 V1 neuronal subpopulations presented in (B) and (C). *Kruskal-Wallis test followed by Tukey-Kramer post-hoc comparison ($P < 0.0001$).*

5.6 Discussion

The goal of this study was to determine if the sound modulation that was described in the first two aims was immutable or if the same sounds could lead to different modulations depending on the behavioral context. Indeed, several previous studies suggested that the modulation by sounds of V1 neurons depends on the contextual relationship between the visual and auditory stimuli (Meijer et al., 2017; Petro et al., 2017). To address this question, we presented the same series of unimodal visual, auditory, and audiovisual stimuli to mice naïve to the stimuli, mice actively using those stimuli to perform a cross-modal discrimination task, and mice trained to perform that task but passively processing those stimuli. We showed that when mice are passively viewing visual stimuli, the presence of sound refines that representation of the visual stimulus orientation in V1. However, when the mice were performing a task during which visual cues were competing with the auditory cue, the representation of the visual cue depended on the behavioral value of the visual cue: the representation of Go cues were improved while the representation of NoGo cues were degraded. Since the behavioral analysis indicated that mice favored the Go outcome over the NoGo outcome when the indications provided by the auditory and visual cues were contradictory, this result suggest that visual processing is adapted during the task to favor the transmission of the most valued signal (Go) used to accomplish the behavioral goal. This adaption of visual processing is likely a dynamic process and not the result of a plasticity as mice trained but not actively performing the task present a modulation in V1 very similar to that found in naïve animals.

5.6.1 Cross-modal integration and behavioral performance

Several evidences in the behavior of the mice during the cross-modal discrimination task indicate that mice integrate the behavioral value of the visual and auditory cue to decide their behavioral response (lick or no lick). The first evidence is that mice perform better for congruent NoGo audiovisual trials (i.e. when NoGo visual cues are paired with NoGo auditory cues) compared to NoGo unimodal trials (either visual or auditory). This result was not found for congruent Go trials as the performance of mice was not significantly improved compared to the Go unimodal trials. It is likely that this absence of improvement is mainly caused by the already high performance of the mice to respond to the visual and auditory Go cues in the two unimodal contexts, leaving little room for improvement. Interestingly, the Go congruent trials show less variability in performance across the animals tested, suggesting that congruent information improves the performance of the worst performers.

The behavioral analysis also revealed that cross-modal integration had a positive effect on the latency in which mice initiated lick for water during Go trials or stopped licking for water during the NoGo trials. These results corroborate previous findings that reaction times are reduced in an audiovisual context in different animal species including humans, monkeys, and rodents (Posner et al., 1976, Lanz et al., 2013; Komura et al., 2005, respectively). These faster reaction times are unlikely due to the discrepancy in visual and auditory propagation speeds in their respective cortices. Indeed, even if auditory information generally reaches the auditory cortex at shorter latencies than visual inputs converge in the visual cortex (Burr and Alais, 2006), the

fact that mice initiate licking earlier when the Go visual signal is paired with the NoGo auditory signal suggest that mice take into account the value of the visual cue when initiating the licking activity.

Finally, the behavioral data suggest that, during the audiovisual blocks, mice use a "search for Go signal, and lick" strategy. Indeed, when conflicting audiovisual trials are presented, mice choose to lick in more than 80% of the trials. These results are corroborated by analyses showing that mice do not attend to one modality over the other and do not adopt an aleatory choice strategy. Altogether, those result indicate that mice, during audiovisual blocks, look for Go cues in both visual and auditory signals.

5.6.2 Assessing the representation of the orientation of the visual stimulus in V1 with shallow neuronal networks (SNNs)

The use of SNNs to determine the representation of the orientation and direction of the visual stimulus integrated in V1, improves the approach used in Aim #1 (McClure and Polack, 2019). Indeed, the use of SNNs does not require setting a threshold to distinguish neurons to be included in the analysis from the others. Every neuron is given, during the training of the SNN, a weight proportional to the importance of this neuron for the classifier. This approach can be consider as more 'physiological' as feed-forward architecture of the SNNs along with the weights and biases that are assigned to each input correspond, in many ways, to how visual information is computed in the visual cortex (Suryani et al., 2016). The fact that we can reproduce the results of the study in Aim #1 using this approach is an important

first step to validate its relevance. In particular, it will be important to demonstrate that it is possible to identify the neurons having the greatest weight in the classifier (Olden and Jackson, 2002; Olden et al., 2004) to be able to investigate further the mechanisms underpinning the modulations of representation.

Chapter VI: General discussion

The experiments performed in this dissertation investigated how top-down information interacts with early visual processing, specifically in V1. I accomplished this goal by using sound modulation as a reliable and physiologically relevant model to study top-down influences in V1, compared to classical top-down modulations like attentional modulation that require the animal to be engaged in a complex behavioral task. In aim I, I addressed the overall impact of auditory stimuli (pure tones), presented concomitantly with visual drifting gratings, on the integrative properties of V1 L2/3 neurons. In aim II, I determined if the sound-induced modulation of V1 firing activity also had an impact on perceptual orientation discrimination in mice. Finally, in aim III, I investigated the flexibility of sound modulation in V1 in order to adapt visual processing to the mouse behavioral goals. Unlike previous studies that focused on sound-induced modulation on individual V1 neurons (Iurilli et al., 2012; Ibrahim et al., 2016; Meijer et al., 2017), my population-analysis approach allowed me to answer the following question: which orientation and direction of a visual stimulus are represented in V1 and subsequently sent to higher-order visual areas? I then addressed this question, by taking advantage of two-photon Ca^{2+} imaging, allowing me to measure the activity of hundreds of neurons for every recording session. I also used specific auditory tones (5 kHz and 10 kHz) with non-overlapping representations in the auditory cortex, rather than white noise, to test the specificity of sound-induced modulation of V1 activity. Similarly, in aims I and III I presented orthogonal drifting gratings (45^{o} and 135^{o}) that activated two very distinct neuronal populations in V1 with minimal response overlap, while in aim II I reduced the angular distance between

these two orientations to determine the limit of V1 orientation discriminability. In the remainder of this discussion, I discuss the principal findings of each experiment and their significance for top-down modulation of V1 activity as well as putative mechanisms that mediate such modulation in V1.

6.1 Auditory tones improve the representation of orientation and direction of visual stimuli in V1

In aim I (chapter III), I found that the evoked responses of V1 neurons tuned to the orientation and direction of the visual stimulus are potentiated in the audiovisual context while the responses of neurons tuned for orthogonal orientations or the opposite direction are suppressed, compared to the unimodal condition (Fig 3-3 and Fig 3-4). I also showed that sound modulation in V1 follows the "principle of inverse effectiveness" (Fig 3-3D), stating 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). Furthermore, I showed that the sound-induced modulatory effects in V1 were not due to changes in arousal during the audiovisual trials compared to the unimodal trials, nor changes in locomotion (Fig 3-6).

The mechanisms that underpin sound modulation of V1 representation of orientation and direction, specifically, the enhanced response of some V1 neurons tuned for the orientation of the visual stimulus and the suppressed response of other V1 neurons tuned for the orthogonal orientation, remain unclear. However, the current hypothesis states that sound modulation in V1 is due to the direct cortico-cortical projections reciprocally connecting A1 to V1. Under this hypothesis the enhancement and suppression of differently tuned neurons to a visual stimulus can be explained by

A1 to V1 synapses on excitatory and inhibitory neuronal subclasses (Iurilli et al., 2012; Zingg et al., 2014; Ibrahim et al., 2016). In one possible pathway, described by Ibrahim and colleagues (2016), sound-induced enhancement of neurons tuned to the orientation of the visual stimulus is mediated by A1 L5 direct connections with L1 interneurons in V1. Indeed, they found sound stimulation excited L1 V1 interneurons. However, when they optogenetically inactivated this pathway, they observed partial suppression of sound modulation in V1 L2/3 neurons, suggesting other inhibitory circuits may be involved (Ibrahim et al., 2016). In a different pathway, Iurilli and colleagues proposed that a subpopulation of L5 V1 neurons that are excited by A1 projections, in turn, activate local V1 L2/3 interneurons and are responsible for inhibiting excitatory L2/3 neurons (Iurilli et al., 2012). In their study, they showed sound-induced suppression of the responses of L2/3 VIP neurons, but no sound modulation was found for PV and SOM neurons (Iurilli et al., 2016). Finally, unpublished results from my experiments in aim I suggest that a majority of identified GABAergic neurons in L2/3, tuned for orientations orthogonal or direction opposite to that of the visual stimulus, are suppressed by sound (Fig. 6-1). My findings suggest that the enhanced response of excitatory neurons tuned for the orientation of the visual stimulus could be due to a decrease of cross-orientation and cross-direction inhibition. However, future studies investigating the sound-induced effects on individually labeled and identifiable interneuron subtypes are needed to determine how V1 interneurons contribute to the modulatory effects of sound on excitatory V1 neurons.

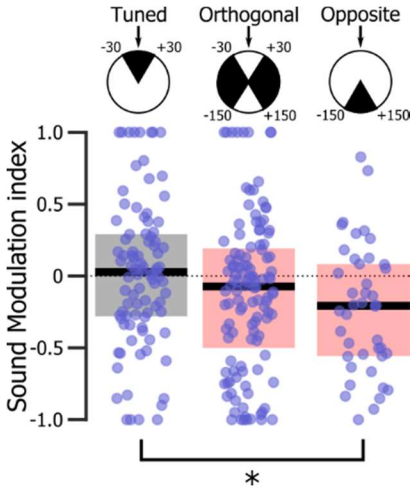

Figure 6-1: Sound modulation of V1 L2/3 GABAergic neurons. Sound modulation [defined as $(R_{AV} - R_V) / (R_{AV} + R_V)$ with R_{AV} response in the audiovisual context and R_V response in the unimodal visual context] of V1 L2/3 GABAergic neurons as a function of their tuning relative to the visual stimulus orientation. GABAergic neurons were labelled genetically with td-tomato. *: Kruskal–Wallis test ($p = 0.025$) followed by Dunn-Sidak multiple comparisons ($p_{1-2} = 0.19$, $p_{1-3} = 0.03$, $p_{2-3} = 0.50$ with 1: tuned, 2: orthogonal and 3: opposite). A majority of interneurons with preferred orientations corresponding to orientations orthogonal or opposite to the orientation of the visual stimulus were suppressed by sound (two-tailed sign test: $p = 0.03$ and $p = 0.04$, respectively).

6.2 Auditory stimuli sharpen the representation of orientation in V1 during an angular perception task and improves perceptual orientation discrimination

My results from aim I suggested that cross-modal integration in V1 could be a mechanism that underpins the well-studied improvement of visual perception in an audiovisual context (Gleiss and Kayser, 2014; Lippert et al., 2007; Vroomen and Geyser, 2000). Therefore, for my second aim, I tested this possible link in mice by determining if auditory stimuli improve mice' ability to discriminate between two oriented visual stimuli of decreasing angular distances, while I performed two-photon Ca^{2+} imaging in L2/3 of V1. First, my behavioral results revealed that for the Go/NoGo angular perception task mice were able to discriminate between two oriented stimuli up to an angular distance between the Go and NoGo cue of 25° (Fig 4-4 and Fig 4-5). Furthermore, the behavioral performances at the group and individual levels in the audiovisual condition was improved for the 25° angular distance compared to the

unimodal condition. Moreover, sound had the biggest impact on the mice performance during NoGo trials (Fig 4-4 and Fig 4-5). The sound-induced behavioral improvements, specifically for the NoGo trials, was associated with an improvement in the representation of the NoGo orientation (Fig 4-8B; Bottom). Interestingly, sound did not influence the behavioral performance nor the representation of the Go/NoGo orientations in V1 for easily discriminable (45^{o} and 30^{o}) or indistinguishable angular distances (20^{o} and 15^{o}). However, for all angular distances the V1 populations representing the NoGo orientation was biased towards the previous day's NoGo orientation. This particular finding suggested a top-down effect of learning and plasticity, occurring either in V1 or in a structure that modulates V1 in an orientation-specific way.

To parse out the effects of learning on orientation discrimination during visual perception, future studies are needed in order to implement control experiments. For the control experiments, a cohort of mice will be tested on their ability to discriminate between Go and NoGo orientations for the same angular distance (45^{o}) across multiple sessions/days. We hypothesize that for the first day of testing the V1 population representing the NoGo orientation will favor the orientation presented from the previous day. However, following consecutive days of testing at the 45^{o} distance the representation should change to the actual orientation shown and become stable with each testing day. Furthermore, additional experiments are needed to link orientation discrimination in V1 to perceptual orientation discriminability.

6.3 Sound dynamically adapts V1 to the behavioral goals

In our final experiment (aim III), the goal was to demonstrate that sound modulation in V1 is flexible and depends on the behavioral goal. To accomplish this, mice were trained to discriminate between two orthogonal visual stimuli as well as two separate auditory pure tones. Next, mice were tested on their ability to correctly lick for a water reward during congruent and incongruent pairings of the visual and auditory stimuli. During the task V1 neurons were simultaneously recording with the same calcium imaging procedure utilized in aims I and II.

The behavioral results revealed that the mice use both visual and auditory cues to perform the audiovisual discrimination task. Indeed, during the presentation of NoGo congruent trials (i.e. when NoGo visual cues are paired with NoGo auditory cues) mice performed significantly better compared to unimodal NoGo trials. The performance in Go congruent trials was not significantly improved likely because mice already respond almost perfectly to both the Go unimodal and Go congruent trials. While mice were less prone to licking during ambiguous trials (Go paired with NoGo), they licked by default well above chance level. Overall, the behavioral results showed that mice do not attend to one modality specifically, but instead assess the visual and auditory cues looking for a Go signal. The neural data overall showed that the representation of Go cues was improved while the representation of NoGo cues were degraded. When we investigated the V1 representation of the Go and NoGo visual cues in trained mice satiated for water and passively viewing the unimodal and audiovisual cues, we found the same pattern of sound modulation as the naïve mice,

suggesting that NoGo cue suppression in trained mice is dynamic, and is specific to the context of the task (behavioral goal).

6.4 Innovation of the thesis experiments

The innovation of this thesis lies in: [1] the establishment of a link between sound-induced modulation of the representation of orientation in V1 and orientation discrimination during an angular perception task. To our knowledge, no studies have directly investigated the link between the representation of V1 populations and orientation discrimination abilities in mice. In aim II we successfully established a behavioral paradigm such that we could determine the mice limit of discriminability. We also showed, for the first time, that naïve auditory stimuli improves the mice performance at their orientation discrimination threshold. [2] The establishment of a personalized behavioral paradigm conceived to elicit in mice the cross-modal comparison of the value (reward vs. no reward) of an auditory cue and a visual cue. The cross-modal discrimination task allowed us to test the adaptability of auditory modulation in V1. [2] The use of two-photon functional imaging to record a large number of neurons in V1 during the passive viewing of audiovisual stimuli, the Go/NoGo angular perception task, and during the cross-modal discrimination task. As the encoding of visual information was found to be sparse in the visual cortex (Olshausen and Field, 1997), it was necessary to be able to record large neuronal populations. This type of technique is not available in human research and is difficult to implement with non-human primates. [3] The use of genetically encoded calcium sensors allowing chronic recordings. This technique made possible the recording of

neural activity in V1 from day-after-day, increasing the yield of the recording sessions and therefore making possible a more detailed description of visual processing in V1 during cross-modal perception. [4] The use of transgenic mice expressing a fluorescent protein td-tomato in all interneurons. This mouse strain (Gad2Cre-Ai9) allowed for the identification of interneurons during the recording sessions in Aim I.

Altogether, carrying out this thesis proposal will provide novel evidence on the cellular and network mechanisms of cross-modal interaction at the level of primary sensory cortices and contribute to our understanding of the impact of crossmodal inputs on early sensory cortical processing.

6.5 Disorders related to alterations in cross-modal perception

Two neurological diseases, Alzheimer's disease (AD) and schizophrenia, have been associated with a reduced capacity in cross-modal integration. In particular, these patients have a difficult time processing sensory information, leading them to poor performances at tasks involving face recognition, motion processing and visual attention (Possin, 2010; Marco et al., 2011; Hornix et al., 2018). For example, Alzheimer patients show pronounced audiovisual deficits during speech perception (Delbeuck et al., 2007). They report a smaller McGurk effect compared to normal controls. On the other hand, Schizophrenic patients show less audiovisual integration during adolescence and early adulthood compared to healthy patients (Pearl et al., 2009). One hypothesis for the pathophysiology of these diseases is that the functional connectivity between higher-order cortical areas and sensory cortices is altered, resulting in an abnormal top-down control of sensory processing. Indeed, the altered

contextual modulation of V1 responses (Seymour et al., 2013) and the need for longer stimulus durations to detect visual stimuli (Butler and Javitt, 2005) in schizophrenic patients could all be potentially explained by deficiencies in top-down control of visual processing. Therefore, a better understanding of the physiological mechanisms underpinning top-down cross-modal interactions at different stages of sensory processing would provide valuable insights on the pathogenesis of these two diseases.

List of References

Aasebo IEJ, Lepperod ME, Stavrinou M, Nokkevangen S, Einevoll G, Hafting T, Fyhn M (2017) Temporal Processing in the Visual Cortex of the Awake and Anesthetized Rat. eNeuro 4.

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

Alonso JM, Yeh CI, Weng C, Stoelzel C (2006) Retinogeniculate connections: A balancing act between connection specificity and receptive field diversity. Prog Brain Res 154:3-13.

Arnold DH, Tear M, Schindel R, Roseboom W (2010) Audio-visual speech cue combination. PLoS One 5:e10217.

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

Balaram P, Young NA, Kaas JH (2014) Histological features of layers and sublayers in cortical visual areas V1 and V2 of chimpanzees, macaque monkeys, and humans. Eye Brain 2014:5-18.

Baldwin MK, Kaskan PM, Zhang B, Chino YM, Kaas JH (2012) Cortical and subcortical connections of V1 and V2 in early postnatal macaque monkeys. J Comp Neurol 520:544-569.

Barraclough NE, Xiao D, Baker CI, Oram MW, Perrett DI (2005) Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. J Cogn Neurosci 17:377-391.

Bartlett EL (2013) The organization and physiology of the auditory thalamus and its role in processing acoustic features important for speech perception. Brain Lang 126:29-48.

Bendor D, Wang X (2008) Neural response properties of primary, rostral, and rostrotemporal core fields in the auditory cortex of marmoset monkeys. J Neurophysiol 100:888-906.

Berens P (2009) CircStat: A MATLAB toolbox for circular statistics. Journal of Statistical Software 31:1-21.

Berens P et al. (2018) Community-based benchmarking improves spike rate inference from two-photon calcium imaging data. PLoS Comput Biol 14:e1006157.

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

Borst JG, Sakmann B (1998) Calcium current during a single action potential in a large presynaptic terminal of the rat brainstem. J Physiol 506 (Pt 1):143-157.

Bradley A, Skottun BC, Ohzawa I, Sclar G, Freeman RD (1987) Visual orientation and spatial frequency discrimination: a comparison of single neurons and behavior. J Neurophysiol 57:755-772.

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.

Bulkin DA, Groh JM (2006) Seeing sounds: visual and auditory interactions in the brain. Curr Opin Neurobiol 16:415-419.

Burr D, Alais D (2006) Combining visual and auditory information. Prog Brain Res 155:243-258.

Butler PD, Javitt DC (2005) Early-stage visual processing deficits in schizophrenia. Curr Opin Psychiatry 18:151-157.

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.

Cappe C, Morel A, Barone P, Rouiller EM (2009) The thalamocortical projection systems in primate: an anatomical support for multisensory and sensorimotor interplay. Cereb Cortex 19:2025-2037.

Carandini M, Churchland AK (2013) Probing perceptual decisions in rodents. Nat Neurosci 16:824-831.

Chanauria N, Bharmauria V, Bachatene L, Cattan S, Rouat J, Molotchnikoff S (2019) Sound Induces Change in Orientation Preference of V1 Neurons: Audio-Visual Cross-Influence. Neuroscience 404:48-61.

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.

Dai J, Wang Y (2012) Representation of surface luminance and contrast in primary visual cortex. Cereb Cortex 22:776-787.

De Valois RL, Yund EW, Hepler N (1982) The orientation and direction selectivity of cells in macaque visual cortex. Vision Res 22:531-544.

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

Delbeuck X, Collette F, Van der Linden M (2007) Is Alzheimer's disease a disconnection syndrome? Evidence from a crossmodal audio-visual illusory experiment. Neuropsychologia 45:3315-3323.

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

Deneux T, Kaszas A, Szalay G, Katona G, Lakner T, Grinvald A, Rozsa B, Vanzetta I (2016) Accurate spike estimation from noisy calcium signals for ultrafast three-dimensional imaging of large neuronal populations in vivo. Nat Commun 7:12190.

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.

Donohue SE, Appelbaum LG, Park CJ, Roberts KC, Woldorff MG (2013) Cross-modal stimulus conflict: the behavioral effects of stimulus input timing in a visual-auditory Stroop task. PLoS One 8:e62802.

Eggermont JJ (2013) On the similarities and differences of non-traumatic sound exposure during the critical period and in adulthood. Front Syst Neurosci 7:12.

Einstein MC, Polack PO, Tran DT, Golshani P (2017) Visually Evoked 3-5 Hz Membrane Potential Oscillations Reduce the Responsiveness of Visual Cortex

Neurons in Awake Behaving Mice. J Neurosci 37:5084-5098.

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

Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex 1:1-47.

Foster KH, Gaska JP, Nagler M, Pollen DA (1985) Spatial and temporal frequency selectivity of neurones in visual cortical areas V1 and V2 of the macaque monkey. J Physiol 365:331-363.

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.

Glickfeld LL, Histed MH, Maunsell JH (2013) Mouse primary visual cortex is used to detect both orientation and contrast changes. J Neurosci 33:19416-19422.

Goodale MA, Milner AD (1992) Separate visual pathways for perception and action.

Trends Neurosci 15:20-25.

Goodale MA, Meenan JP, Bulthoff HH, Nicolle DA, Murphy KJ, Racicot CI (1994) Separate neural pathways for the visual analysis of object shape in perception and prehension. Curr Biol 4:604-610.

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.

Hanson L, Sethuramanujam S, deRosenroll G, Jain V, Awatramani GB (2019) Retinal direction selectivity in the absence of asymmetric starburst amacrine cell responses. Elife 8.

Heffner RS (2004) Primate hearing from a mammalian perspective. Anat Rec A Discov Mol Cell Evol Biol 281:1111-1122.

Hentschke H, Stuttgen MC (2011) Computation of measures of effect size for neuroscience data sets. Eur J Neurosci 34:1887-1894.

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

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

Hirokawa J, Sadakane O, Sakata S, Bosch M, Sakurai Y, Yamamori T (2011) Multisensory information facilitates reaction speed by enlarging activity difference between superior colliculus hemispheres in rats. PLoS One 6:e25283.

Hochstein S, Ahissar M (2002) View from the top: hierarchies and reverse hierarchies in the visual system. Neuron 36:791-804.

Hornix BE, Havekes R, Kas MJH (2019) Multisensory cortical processing and dysfunction across the neuropsychiatric spectrum. Neurosci Biobehav Rev 97:138-151.

Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol 160:106-154.

Hubel DH, Wiesel TN (1968) Receptive fields and functional architecture of monkey striate cortex. J Physiol 195:215-243.

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.

Ji W, Gamanut R, Bista P, D'Souza RD, Wang Q, Burkhalter A (2015) Modularity in the Organization of Mouse Primary Visual Cortex. Neuron 87:632-643.

Jurjut O, Georgieva P, Busse L, Katzner S (2017) Learning Enhances Sensory Processing in Mouse V1 before Improving Behavior. J Neurosci 37:6460-6474.

Kaas JH, Hackett TA (1999) 'What' and 'where' processing in auditory cortex. Nat Neurosci 2:1045-1047.

Kaas JH, Hackett TA, Tramo MJ (1999) Auditory processing in primate cerebral

cortex. Curr Opin Neurobiol 9:164-170.

Khibnik LA, Tritsch NX, Sabatini BL (2014) A direct projection from mouse primary visual cortex to dorsomedial striatum. PLoS One 9:e104501.

Khoei MA (2014) Motion-based position coding in the visual system: a computational study. In: Neurosciene France: Aix-Marseille Universite´.

Kim EJ, Juavinett AL, Kyubwa EM, Jacobs MW, Callaway EM (2015) Three Types of Cortical Layer 5 Neurons That Differ in Brain-wide Connectivity and Function. Neuron 88:1253-1267.

Komura Y, Tamura R, Uwano T, Nishijo H, Ono T (2005) Auditory thalamus integrates visual inputs into behavioral gains. Nat Neurosci 8:1203-1209.

Lakatos P, Karmos G, Mehta AD, Ulbert I, Schroeder CE (2008) Entrainment of neuronal oscillations as a mechanism of attentional selection. Science 320:110-113.

Langers DR, van Dijk P (2012) Mapping the tonotopic organization in human auditory cortex with minimally salient acoustic stimulation. Cereb Cortex 22:2024-2038.

Lanz F, Moret V, Rouiller EM, Loquet G (2013) Multisensory Integration in Non-Human Primates during a Sensory-Motor Task. Front Hum Neurosci 7:799.

Liang F, Bai L, Tao HW, Zhang LI, Xiao Z (2014) Thresholding of auditory cortical representation by background noise. Front Neural Circuits 8:133.

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

Lodish H, Berk A, Zipursky SL, al. e (2000) Sensory Tranduction In: Molecular Cell Biology 4th Edition Edition. New York: W.H. Freeman

Luttke CS, Perez-Bellido A, de Lange FP (2018) Rapid recalibration of speech perception after experiencing the McGurk illusion. R Soc Open Sci 5:170909.

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

Marco EJ, Hinkley LB, Hill SS, Nagarajan SS (2011) Sensory processing in autism: a review of neurophysiologic findings. Pediatr Res 69:48R-54R.

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.

Masland RH (2001) Neuronal diversity in the retina. Curr Opin Neurobiol 11:431-436.

Masland RH (2012) The neuronal organization of the retina. Neuron 76:266-280.

McClure JP, Jr., Polack PO (2019) Pure tones modulate the representation of orientation and direction in the primary visual cortex. J Neurophysiol 121:2202-2214.

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.

McGurk H, MacDonald J (1976) Hearing lips and seeing voices. Nature 264:746-748.

Medathati NVK, Neumann H, Masson G, Kornprobst P (2016) Bio-Inspired Computer Vision: Towards a Synergistic Approach of Artificial and Biological Vision

Computer Vision and Image Understanding 150.

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) Spatial factors determine the activity of multisensory neurons in cat superior colliculus. Brain Res 365:350-354.

Michael-Titus A, Revest P, Shortland P (2010) The Visual System In: The Nervous System, 2 Edition, pp 121-140: Churchill Livingstone.

Milner AD, Goodale MA (2008) Two visual systems re-viewed. Neuropsychologia 46:774-785.

Mishkin M, Ungerleider LG (1982) Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. Behav Brain Res 6:57-77.

Mitani A, Itoh K, Nomura S, Kudo M, Kaneko T, Mizuno N (1984) Thalamocortical projections to layer I of the primary auditory cortex in the cat: a horseradish peroxidase study. Brain Res 310:347-350.

Moore BD, Alitto HJ, Usrey WM (2005) Orientation Tuning, But Not Direction Selectivity, Is Invariant to Temporal Frequency in Primary Visual Cortex. Journal of Neurophysiology 94:1336-1345.

Movshon JA, Thompson ID, Tolhurst DJ (1978) Spatial and temporal contrast sensitivity of neurones in areas 17 and 18 of the cat's visual cortex. J Physiol 283:101-120.

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.

Ohki T, Gunji A, Takei Y, Takahashi H, Kaneko Y, Kita Y, Hironaga N, Tobimatsu S, Kamio Y, Hanakawa T, Inagaki M, Hiraki K (2016) Neural oscillations in the temporal pole for a temporally congruent audio-visual speech detection task. Sci Rep 6:37973.

Olden JD, Jackson DA (2002) Illuminating the “black box”: a randomization approach for understanding variable contributions in artificial neural networks. Ecological Modelling 154:135-150.

Olden JD, Joy MK, Death RG (2004) An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. Ecological Modelling 178:389-397.

Olshausen BA, Field DJ (1997) Sparse coding with an overcomplete basis set: a strategy employed by V1? Vision Res 37:3311-3325.

O'Malley CD, Clarke E (1961) The discovery of the auditory ossicles. Bull Hist Med 35:419-441.

Pakan JM, Francioni V, Rochefort NL (2018) Action and learning shape the activity of neuronal circuits in the visual cortex. Curr Opin Neurobiol 52:88-97.

Parker AJ, Newsome WT (1998) Sense and the single neuron: probing the physiology of perception. Annu Rev Neurosci 21:227-277.

Pearl D, Yodashkin-Porat D, Katz N, Valevski A, Aizenberg D, Sigler M, Weizman A, Kikinzon L (2009) Differences in audiovisual integration, as measured by McGurk phenomenon, among adult and adolescent patients with schizophrenia and age-matched healthy control groups. Compr Psychiatry 50:186-192.

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.

Poort J, Khan AG, Pachitariu M, Nemri A, Orsolic I, Krupic J, Bauza M, Sahani M, Keller GB, Mrsic-Flogel TD, Hofer SB (2015) Learning Enhances Sensory and Multiple Non-sensory Representations in Primary Visual Cortex. Neuron 86:1478-1490.

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

Possin KL (2010) Visual spatial cognition in neurodegenerative disease. Neurocase 16:466-487.

Ray TA, Roy S, Kozlowski C, Wang J, Cafaro J, Hulbert SW, Wright CV, Field GD, Kay JN (2018) Formation of retinal direction-selective circuitry initiated by starburst amacrine cell homotypic contact. Elife 7.

Remington L (2012) Visual Pathway. In: Clinical Anatomy and Physiology of the Visual System, 3rd Edition Edition, p 588: Butterworth-Heinemann.

Resulaj A, Ruediger S, Olsen SR, Scanziani M (2018) First spikes in visual cortex enable perceptual discrimination. Elife 7.

Reynolds RP, Kinard WL, Degraff JJ, Leverage N, Norton JN (2010) Noise in a laboratory animal facility from the human and mouse perspectives. J Am Assoc Lab Anim Sci 49:592-597.

Roederer JG (2009) Sound Vibrations, Pure Tones, and the Perception of Pitch. In: The Physics and Psychophysics of Music: An Introduction, pp 22-75. New York, NY: Springer US.

Rubio-Garrido P, Perez-de-Manzo F, Porrero C, Galazo MJ, Clasca F (2009) Thalamic input to distal apical dendrites in neocortical layer 1 is massive and highly convergent. Cereb Cortex 19:2380-2395.

Scannell JW, Blakemore C, Young MP (1995) Analysis of connectivity in the cat cerebral cortex. J Neurosci 15:1463-1483.

Schiller PH, Finlay BL, Volman SF (1976) Quantitative studies of single-cell properties in monkey striate cortex. I. Spatiotemporal organization of receptive fields. J Neurophysiol 39:1288-1319.

Schroeder CE, Lakatos P (2009) Low-frequency neuronal oscillations as instruments of sensory selection. Trends Neurosci 32:9-18.

Self MW, van Kerkoerle T, Super H, Roelfsema PR (2013) Distinct roles of the cortical layers of area V1 in figure-ground segregation. Curr Biol 23:2121-2129.

Sellers KK, Bennett DV, Hutt A, Williams JH, Frohlich F (2015) Awake vs. anesthetized: layer-specific sensory processing in visual cortex and functional connectivity between cortical areas. J Neurophysiol 113:3798-3815.

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

Seymour K, Stein T, Sanders LL, Guggenmos M, Theophil I, Sterzer P (2013) Altered contextual modulation of primary visual cortex responses in schizophrenia. Neuropsychopharmacology 38:2607-2612.

Shams L, Kamitani Y, Shimojo S (2000) Illusions. What you see is what you hear. Nature 408:788.

Shastry BS (1998) Light, sight and fight for insight. Biochimie 80:339-341.

Sherman SM, Guillery RW (1998) On the actions that one nerve cell can have on another: distinguishing "drivers" from "modulators". Proc Natl Acad Sci U S A 95:7121-7126.

Siemann JK, Muller CL, Bamberger G, Allison JD, Veenstra-VanderWeele J, Wallace MT (2014) A novel behavioral paradigm to assess multisensory processing in mice. Front Behav Neurosci 8:456.

Sincich LC, Horton JC (2002) Divided by cytochrome oxidase: a map of the projections from V1 to V2 in macaques. Science 295:1734-1737.

Smith PH, Uhlrich DJ, Manning KA, Banks MI (2012) Thalamocortical projections to rat auditory cortex from the ventral and dorsal divisions of the medial geniculate nucleus. J Comp Neurol 520:34-51.

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 93:940-953.

Stanislaw H, Todorov N (1999) Calculation of signal detection theory measures.

Behav Res Methods Instrum Comput 31:137-149.

Stein BE, Stanford TR (2008) Multisensory integration: current issues from the perspective of the single neuron. Nat Rev Neurosci 9:255-266.

Stein BE, Stanford TR, Rowland BA (2014) Development of multisensory integration from the perspective of the individual neuron. Nat Rev Neurosci 15:520-535.

Stein BE, London N, Wilkinson LK, Price DD (1996) Enhancement of perceived visual intensity by auditory stimuli: a psychophysical analysis. J Cogn Neurosci 8:497-506.

Suryani D, Doetsch P, Ney H (2016) On the Benefits of Convolutional Neural Network Combinations in Offline Handwriting Recognition.

Tees RC (1999) The effects of posterior parietal and posterior temporal cortical lesions on multimodal spatial and nonspatial competencies in rats. Behav Brain Res 106:55-73.

Tsukano H, Horie M, Ohga S, Takahashi K, Kubota Y, Hishida R, Takebayashi H, Shibuki K (2017) Reconsidering Tonotopic Maps in the Auditory Cortex and Lemniscal Auditory Thalamus in Mice. Front Neural Circuits 11:14.

Van Essen DC, Anderson CH, Felleman DJ (1992) Information processing in the primate visual system: an integrated systems perspective. Science 255:419-423.

Van Horn SC, Erisir A, Sherman SM (2000) Relative distribution of synapses in the A-laminae of the lateral geniculate nucleus of the cat. J Comp Neurol 416:509-520.

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.

Vogels R, Orban GA (1990) How well do response changes of striate neurons signal differences in orientation: a study in the discriminating monkey. J Neurosci 10:3543-3558.

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.

Wolf G (2004) The visual cycle of the cone photoreceptors of the retina. Nutr Rev 62:283-286.

Woods TM, Recanzone GH (2004) Visually induced plasticity of auditory spatial perception in macaques. Curr Biol 14:1559-1564.

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

Yang W, Carrasquillo Y, Hooks BM, Nerbonne JM, Burkhalter A (2013) Distinct balance of excitation and inhibition in an interareal feedforward and feedback circuit of mouse visual cortex. J Neurosci 33:17373-17384.

Yau JM, DeAngelis GC, Angelaki DE (2015) Dissecting neural circuits for multisensory integration and crossmodal processing. Philos Trans R Soc Lond B Biol Sci 370:20140203.

Zhou X, Merzenich MM (2008) Enduring effects of early structured noise exposure on temporal modulation in the primary auditory cortex. Proc Natl Acad Sci U S A 105:4423-4428.

Zingg B, Hintiryan H, Gou L, Song MY, Bay M, Bienkowski MS, Foster NN, Yamashita S, Bowman I, Toga AW, Dong HW (2014) Neural networks of the mouse neocortex. Cell 156:1096-1111.