2021

Neuronal Codes And Circuits Underlying Audiovisual Integration

Aaron Williams
University of Pennsylvania

Neuronal Codes And Circuits Underlying Audiovisual Integration

Abstract

We rely on interactions between our sensory systems to help us communicate with each other and navigate our world. This multisensory integration can improve the accuracy of the sensory systems involved. However, many questions remain on how multisensory integration, specifically audiovisual integration, is mediated within the brain. In Chapter 2, we tested whether sound improves visual processing in the primary visual cortex. We found that both individual and populations of neurons encoded visual stimuli better with simultaneous auditory input. Importantly, we also found that this effect was due to sound and not mediated by sound-induced movements, an independent modulator of visual responses. These results clarify the codes underlying this tripartite interaction in this visual region. In Chapter 3, we probed the cortical circuits that support the audiovisual integration in the primary visual cortex. We found that the auditory cortex sends excitatory projections to the visual cortex, and stimulation of these fibers enhances visual response magnitude. However, suppression of this pathway failed to impair audiovisual integration in the primary visual cortex, suggesting the presence of parallel or compensatory mechanisms in this region. In Chapter 4, we explored the subcortical visual circuits that project to the inferior colliculus. We found that the superior colliculus synapses with neurons in the external shell of the inferior colliculus, and stimulation of these projections evokes activity in this auditory midbrain. However, neurons in the inferior colliculus failed to exhibit consistent responses to looming or static audiovisual stimuli, suggesting specificity in the visual tuning of these auditory neurons. Together, these results improve our knowledge of the coding and circuitry principles underlying audiovisual integration in both cortical and subcortical regions, and expand our understanding of how the brain integrates sensory information to generate our smooth perceptual experience.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Neuroscience

First Advisor
Maria N. Geffen

Keywords
Audiovisual integration, Electrophysiology, Optogenetics, Sensory neuroscience, Systems neuroscience

Subject Categories
Neuroscience and Neurobiology

NEURONAL CODES AND CIRCUITS UNDERLYING

AUDIOVISUAL INTEGRATION

Aaron M. Williams

A DISSERTATION

in

Neuroscience

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2021

Supervisor of Dissertation

Maria N. Geffen, PhD

Associate Professor of Otorhinolaryngology

Graduate Group Chairperson

Joshua I. Gold, PhD

Professor of Neuroscience

Dissertation Committee
Dr. Jessica Morgan, PhD, Assistant Professor of Ophthalmology
Dr. Steven Eliades, MD, PhD, Assistant Professor of Otorhinolaryngology
Dr. Jay Gottfried, MD, PhD, Arthur H. Rubenstein University Professor of Neurology
Dr. Alexander Proekt, MD, PhD, Associate Professor of Anesthesiology and Critical Care
Dr. Lindsey Glickfeld, PhD, Associate Professor of Neurobiology, Duke University

ACKNOWLEDGEMENTS

There are so many people that I want to thank for supporting me through graduate school and making this thesis possible. I first want to thank my mentor, Maria. Your scientific ambition, social consciousness, and relaxed personality have provided a great example of a successful career in science. Your mentorship has allowed me to grow as a scientist during these years, and I am grateful for my experience in the lab. I want to thank my lab mates in the Geffen lab, who have built a friendly, collaborative environment that has made my graduate years both productive and fun. I would especially like to thank Dr. Kath Wood and Chris Angeloni, who both provided invaluable scientific mentorship that helped develop my own scientific thinking, skepticism, and rigor. Thank you to Drs. Jessica Morgan, Alex Proekt, Jay Gottfried, Steve Eliades, and Lindsey Glickfeld for all serving on my Thesis Committee and providing scientific and professional guidance. I also want to thank the students and administrators of both NGG and the MSTP for creating a nurturing environment in which I have grown as a physician scientist and learned more about myself, both necessary to prepare for life ahead. Finally, I would like to thank my friends and family. To my friends from childhood, college, and my time here at Penn, you all have made me who I am and for that I am grateful. The games, the parties, the vacations, the solidarity during busy times – our experiences together have supported me more than you know, and I value our friendship and am excited for it to continue to grow. And to Mom, Dad, Ashley, and Bryan, you all have witnessed the highs and lows of my graduate

years and my life more broadly. Your unwavering love throughout these times has driven my academic pursuits, supported my emotional growth, and contributed to my mental wellbeing. I am proud to have you all in my life, and I hope that this scholarly achievement is a celebration of the love and strength we have as a family.

ABSTRACT

NEURONAL CODES AND CIRCUITS UNDERLYING AUDIOVISUAL

INTEGRATION

Aaron M. Williams

Maria N. Geffen

We rely on interactions between our sensory systems to help us communicate with each other and navigate our world. This multisensory integration can improve the accuracy of the sensory systems involved. However, many questions remain on how multisensory integration, specifically audiovisual integration, is mediated within the brain. In Chapter 2, we tested whether sound improves visual processing in the primary visual cortex. We found that both individual and populations of neurons encoded visual stimuli better with simultaneous auditory input. Importantly, we also found that this effect was due to sound and not mediated by sound-induced movements, an independent modulator of visual responses. These results clarify the codes underlying this tripartite interaction in this visual region. In Chapter 3, we probed the cortical circuits that support the audiovisual integration in the primary visual cortex. We found that the auditory cortex sends excitatory projections to the visual cortex, and stimulation of these fibers enhances visual response magnitude. However, suppression of this pathway failed to impair audiovisual integration in the primary visual cortex, suggesting the presence of parallel or compensatory mechanisms in this region. In Chapter 4, we explored the subcortical visual circuits that project to the inferior colliculus. We found that the superior colliculus synapses with neurons in the external shell of the inferior colliculus, and stimulation of these projections evokes activity in this auditory midbrain. However, neurons in the inferior colliculus failed to exhibit

consistent responses to looming or static audiovisual stimuli, suggesting specificity in the visual tuning of these auditory neurons. Together, these results improve our knowledge of the coding and circuitry principles underlying audiovisual integration in both cortical and subcortical regions, and expand our understanding of how the brain integrates sensory information to generate our smooth perceptual experience.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

When we listen to our conversation partner in a crowded room, it often helps to track their lip movements to hear and understand them better. This everyday situation is an example of audiovisual integration, where one sensory modality affects the perception of the other. The neuronal circuits that underlie this sensory processing are not isolated but rather interact with each other. Information is relayed via neuronal connections from peripheral sensory organs to subcortical sensory regions, and then proceeds to the primary sensory cortical areas and upward to higher association areas in the brain. The multitude of brain regions along these sensory pathways provide many opportunities for the sensory systems to exchange information, and this multisensory integration is often beneficial for the perception and accuracy of one or both sensory modalities. In this dissertation, we address questions pertaining to the neuronal mechanisms supporting this multisensory processing, focusing on audiovisual integration.

The goal of our work is to elucidate the neuronal codes that mediate audiovisual integration and reveal the circuitry that underlies this process at both the cortical and subcortical level. In this series of studies, we address the following questions: (1) *How does sound affect visual processing in the primary visual cortex (V1)?* The primary sensory cortical areas are not exclusively responsive to a single sensory modality, but rather are sensitive to the information from other sensory systems and modulate their responses accordingly. Investigating whether and how one form of sensory information, auditory

sound, affects processing in another sensory region, V1, expands our understanding of how multisensory perceptual changes are mediated at a neuronal level. Additionally, consideration of sound-induced movement as a factor in audiovisual integration clarifies the role of this potential confound in audiovisual studies in the awake brain. (2) *What are the neuronal circuits that underlie audiovisual integration in V1?* Our understanding of the audiovisual integration observed in V1 is strengthened by an understanding of the neuronal populations and pathways that mediate this process. We use viral tracing and optogenetic techniques to attempt to identify a causal role of the AC in providing auditory information to V1, and suggest a neuronal circuit that is compatible with the field's current understanding. (3) *How is audiovisual integration mediated in the auditory brainstem?* The inferior colliculus (IC) is the first region in the ascending auditory pathway that receives visual and multisensory information. Visual information in the IC is known to partially originate from the neighboring superior colliculus (SC), which itself participates in multisensory processing. We use electrophysiology and tracing techniques to investigate audiovisual integration at a subcortical level and compare it to its cortical counterpart.

Investigation into the neuronal codes and circuits underlying audiovisual integration, and multisensory processing more broadly, is essential for developing an understanding of how sensory information supports brain function more generally. This information is used by organisms to generate sensory perceptions, thoughts, and behaviors to find food, avoid predators, and reproduce. Therefore, it is useful to begin by reviewing the current understanding of how audiovisual and multisensory integration affect these sensory perceptions and behaviors.

Audiovisual and multisensory integration at a perceptual and behavioral level

Multisensory integration is an important aspect of sensory processing in both humans and animals. The McGurk effect is a common example of audiovisual integration in which the auditory phoneme that is perceived depends on the visual viseme, or syllabic lip movement, that it is paired with (McGurk and MacDonald, 1976). The flash-beep illusion is another audiovisual example in which the number of visual flashes perceived depends on the number of auditory beeps that accompany it (Shams et al., 2002). The McGurk and flash-beep effects demonstrate multisensory perceptual integration in an operationalized manner. However, we rely on multisensory integration in everyday situations such as our attention to body language and facial cues when maintaining a conversation in a crowded environment, the use of both tactile and visual input to help when playing certain instruments, or the interplay between olfactory and gustatory senses when enjoying a meal. The various realms in which multisensory processing is used demonstrates how common and useful it is at a perceptual level.

Such perceptual experiences provide qualitative examples of audiovisual and multisensory integration. Perception is difficult to quantify in an experimental setting – behavioral tasks are instead used in neuroscience and psychology to parameterize, quantify, and compare sensory and multisensory processing. A recent study demonstrated that in a virtual reality driving task, humans performed better with additional auditory and tactile input (e.g. car engine noise and steering wheel vibration, respectively) than with visual input alone (Marucci et al., 2021). Notably, this additional sensory information was more beneficial under "high workload" conditions, in which vision was obscured by mist and rainy conditions, than "low workload" conditions with clear weather. Another study in

school-age children found that reaction time was lower to an audiovisual stimulus than to an isolated auditory or visual stimulus, and this improvement in reaction time was predictive of the same subjects' recognition and working memory performance (Denervaud et al., 2020).

In other animals, behavioral tasks often come in the form of training subjects to detect or discriminate between sensory stimuli, and multisensory studies compare detection and discrimination thresholds to their unisensory counterparts. A 2012 study found that rats were better able to detect and correctly identify the source direction of a lateralized audiovisual stimulus than unisensory auditory or visual stimuli (Gleiss and Kayser, 2012). A similar study in mice found increases in response rate and decreases in reaction time when detecting audiovisual stimuli compared to auditory and visual stimuli (Meijer et al., 2018). These multisensory improvements in behavioral performance extend beyond just audiovisual integration. In a task in which mice were trained to lick in response to tactile stimulation of the forepaw, additional auditory white noise increased lick response rate and decreased lick latency (Godenzini et al., 2021). Additionally, pup odor and vocalizations act synergistically to stimulate pup search and retrieval in maternal dam mice (Cohen et al., 2011; Okabe et al., 2013). Therefore, it becomes clear that the brain often relies on a combination of sensory modalities not only in generating one's perceptual experience but also in forming appropriate behavioral responses.

The range of studies on audiovisual and multisensory integration in a variety of settings in both humans and model organisms has clarified three main factors that are important to understanding when and how multisensory integration occurs. The first of these is that the additional sensory input is most beneficial when the unisensory modality

is at detection threshold. In the 2021 virtual reality study cited above, this came in the form of slightly impaired vision due to inclement weather (Marucci et al., 2021). And in the cited rodent studies, multisensory conditions were most beneficial when compared to unisensory low visual contrast or low auditory intensity (Gleiss and Kayser, 2012; Meijer et al., 2018). This follows a somewhat intuitive understanding that detection or discrimination of sensory stimuli can be most improved when the stimulus is barely at perceptual threshold and performance is therefore low, whereas unisensory stimuli that are at full intensity easily evoke maximum performance accuracy with quick reaction times.

The second important factor in multisensory integration is the spatiotemporal association between the incoming stimuli. Coincidence between the temporal onset of an auditory and visual stimulus as well as their source direction suggests that these inputs share a common source in the external environment. In this case, it would be beneficial to integrate these sensory stimuli and understand them as originating from a single object. However, some multisensory studies do use a temporal offset between the stimuli (Garner and Keller, 2020), demonstrating that the brain is still able to associate a cause-and-effect relationship between sensory cues with appropriate temporal regularity.

The third important factor affecting multisensory integration, particularly audiovisual integration, is temporal congruency between dynamic components of the sensory stimuli. When talking, the changes in one's mouth and tongue movements are closely associated with changes in the intensity and timbre of the spoken syllables. And when an object is quickly approaching, the visual size of the object likely grows larger as the sound it makes grows louder. This temporal congruency has experimentally been shown to be important. In humans, performance on a stimulus discrimination task was

improved with congruent cross-modal stimulation, and actually suffered with incongruent cross-modal stimulation (Laurienti et al., 2004). Chimpanzees are also able to recognize the correspondence between silent movies of faces and their associated auditory vocalization (Izumi and Kojima, 2004). The temporal congruency between incoming stimuli of separate modalities may affect neurons' ability to entrain and phase-lock to the encoded features of the sensory stimuli. There is evidence of this in both the visual and auditory cortical areas (Meijer et al., 2017; Atilgan et al., 2018), hinting at the neuronal codes that mediate multisensory processing.

Despite our awareness of multisensory integration and the psychophysics of it, a detailed understanding of the process at a neuronal level is still developing. Specifically, the relationship between the psychophysics principles outlined above and the underlying neuronal coding principles is unclear. Therefore, it is the goal of the research included in this dissertation to better elucidate the neuronal codes that mediate multisensory integration in the brain. Multisensory improvements in perceptual detection and discrimination of sensory stimuli suggest that similar improvements in sensory processing would be observed at a neuronal level. However, this has yet to be demonstrated, and Chapter 2 focuses on this topic specifically in V1, where many audiovisual studies have been performed. Chapter 3 contributes to this neuronal coding question by identifying which auditory regions are responsible for supporting V1 audiovisual integration. Therefore, we continue by discussing the neuronal codes that underlie audiovisual and multisensory processing and which questions remain unaddressed.

Audiovisual and multisensory integration at a neuronal level

The primary sensory cortical areas are so named because they are the first cortical region in their respective ascending sensory pathways. However, they are capable of responding to input from different sensory modalities. In the study cited above in which sound improved mouse responses to tactile input, the neuronal correlates of this behavioral improvement were localized to the somatosensory cortex (Godenzini et al., 2021). In this region, dendritic and somatic encoding of the tactile input was modulated by sound in layer 2/3 neurons. Another study of multisensory processing in the cortex found that exposing female mice to pup odors reshapes responses to pure tones, natural auditory stimuli, and pup vocalizations in AC neurons (Cohen et al., 2011), a process mediated by the social signaling neuropeptide oxytocin (Marlin et al, 2015). The AC is also sensitive to visual input, as it has been shown that neurons modulate their auditory responses based on the presence of light (Morrill and Hasenstaub, 2018). This finding was particularly prominent in the infragranular layers, a distinction from the prior somatosensory study. Furthermore, in accordance with the above discussion on the temporal congruency of audiovisual input, visual stimuli can be used in the AC to entrain neurons to particular auditory streams that follow a similar temporal pattern (Atilgan et al., 2018). Therefore, while neurons in cortical areas are often tuned for specific features of their primary modality, it is clear that they are often sensitive to the presence of additional sensory input and can adjust their sensory responses accordingly.

Vision is the primary sensory modality in humans, so multisensory processing in the key region for visual processing, V1, has received particular attention in past studies. In V1 of anesthetized mice, neurons sharpen their visual orientation and direction

selectivity when there is coincident auditory noise (Ibrahim et al., 2016). This neuronal finding would suggest a potential mechanism by which sound may improve vision at a perceptual level, however this was in anesthetized mice. In awake mice, a different coding pattern has been observed. The number of neurons responding to a visual stimulus (McClure and Polack, 2019) and visual response magnitude (Meijer et al., 2017) is modulated when sound accompanies the visual input. However, these awake studies have yet to demonstrate that the changes associated with audiovisual stimuli result in improved encoding of the visual stimulus. Chapter 2 directly addresses this question, with additional consideration of how visuo-locomotive integration is involved in this processing. V1 neurons' visual responses are sensitive to movement in awake animals (Neill and Stryker, 2010), and this motor input improves neuronal encoding of the visual stimulus (Dardalat and Stryker, 2017). Therefore, it is important in audiovisual studies to factor in movement of the awake animals to accurately assess how sound affects visual processing.

Multisensory integration has also been observed and characterized at a subcortical level. The SC displays sensitivity to cross-modal inputs such as sound, somatosensation, and eye position. In anesthetized mice, bimodal neurons responding to light and sound, or light and somatosensory input, and even trimodal neurons responding to all three have been observed in deeper layers of the SC (Dräger and Hubel, 1975). The SC also plays a pivotal role in responding to and generating saccadic eye movements (Sparks et al, 2000), the cross-modal intersection between vision and motor planning. The thalamus has received less attention in multisensory studies, perhaps due to its depth in the brain limiting access to experimenters. However, fMRI studies in humans have found audiovisual enhancement of BOLD signals in the visual lateral geniculate body (LGB) and auditory medial

geniculate body (MGB; Noesselt et al., 2010). And improved classification accuracy of emotional audiovisual stimuli was associated with cross-modal BOLD signals with in the thalamus (Kreifelts et al., 2007; Tyll et al., 2011), contributing to our understanding of the relationship between the neuronal and perceptual correlates of multisensory processing. Therefore, it is well known that subcortical sensory regions readily participate in audiovisual and multisensory processing.

In the auditory system, the IC in the midbrain is the recipient of broad cross-modal signals. Neurons in this region are sensitive to motor and somatosensory input. In anesthetized rats, IC neurons' acoustic responses are potentiated by somatosensory input via sciatic nerve stimulation (Syka and Radil-Weiss, 1978). And in anesthetized cats, individual neurons in the IC external nucleus were found to have tactile bodily receptive fields that fully mapped the body across the population (Aitkin et al., 1978; Gruters and Groh, 2012). The barn owl is an animal that relies heavily on its refined sense of hearing for precision while hunting. In the external nucleus of the IC of barn owls, neurons are tuned for spatial location of incoming sounds. And despite these neurons displaying minimal sensitivity to isolated visual stimuli, coincident visual input enhances the response of these neurons to the auditory stimuli (Bergan and Knudsen, 2009). Chapter 4 expands on this foundation by exploring whether IC neurons are sensitive to a broader range of visual stimuli, such as drifting gratings and looming patterns. By mapping the range of visual and audiovisual features that neurons encode, we have a better understanding of which brain regions are involved in the various perceptual improvements associated with audiovisual and multisensory integration.

When considering what and how sensory information is encoded in these brain regions, the relationship between the brain region of interest and the ethological relevance of the sensory features being studied must be factored in. For example, in the barn owl, IC neurons adjust their sound localization tuning with visual input (Bergan and Knudsen, 2009). However, IC neurons are also tuned for auditory frequency, but neuronal responses to this auditory feature in this region are largely invariant to the presence of additional visual stimuli. If one were to only study the integration of visual input and auditory frequency selectivity in IC neurons, it would be incorrect to conclude that the IC is insensitive to visual input and does not participate in audiovisual integration. Therefore, it is easier to demonstrate the presence of cross-modal input than the absence of it in neuronal studies.

Building on the understanding of what information is encoded in specific brain regions, sensory neuroscience is also interested in how these brain regions communicate with each other. Such neuronal communication often entails the filtering and/or integration of the sensory information represented in the afferent brain region. Chapter 3 explores this by focusing on the how the auditory cortex (AC) provides auditory information to support audiovisual integration in V1, and Chapter 4 additionally explores subcortical SC and IC communication. We therefore continue by briefly reviewing the experimental techniques often used to probe the role these circuits play in multisensory integration

Neuronal circuitry underlying audiovisual and multisensory integration

Mapping the pathways and circuits that connect neurons both within and between brain regions complements our functional understanding of neurons. The connectome, as

this comprehensive map is known, could provide clues about where in the brain to physically intervene when medically necessary, such as in cases of epilepsy, movement disorders, and other neurological conditions. The field of sensory and systems neuroscience is particularly well-suited to contribute to this venture by experimentally relating neuronal sensory responses to the physical networks through which information is transmitted and stored. By mapping these neuronal connections at various scales, we build a better understanding of the hierarchical functional networks the brain uses to operate.

In the field of multisensory integration, determining from which brain region cross-sensory information originates can help elucidate the function of that neuronal pathway. To this end, many cortical and subcortical connections have been established (Figure 1.1). The AC has been shown to contribute to audiovisual integration in V1. Optogenetically stimulating AC axons in V1 sharpened orientation tuning in the region (Ibrahim et al., 2017), and pharmacologically inhibiting AC neurons suppressed auditory modulation of visual responses in V1 (Deneux et al., 2019). This suggests AC activity has a causal role in providing auditory information to V1. This AC-V1 connection may also be critical for audiovisual cue association in V1 (Garner and Keller, 2020), implicating this cortico-cortical circuit in the role of sensory learning.

The AC may be connected to motor and somatosensory regions as well. Optogenetic inhibition of AC axons suppresses auditory modulation of tactile sensory responses (Godenzini et al., 2021), demonstrating another causal role the AC has in multisensory integration. However, the AC also receives cortico-cortical input as well. A combination of techniques was used to demonstrate that the secondary motor cortex provides feedforward inhibition to the AC to suppress the salience of self-generated
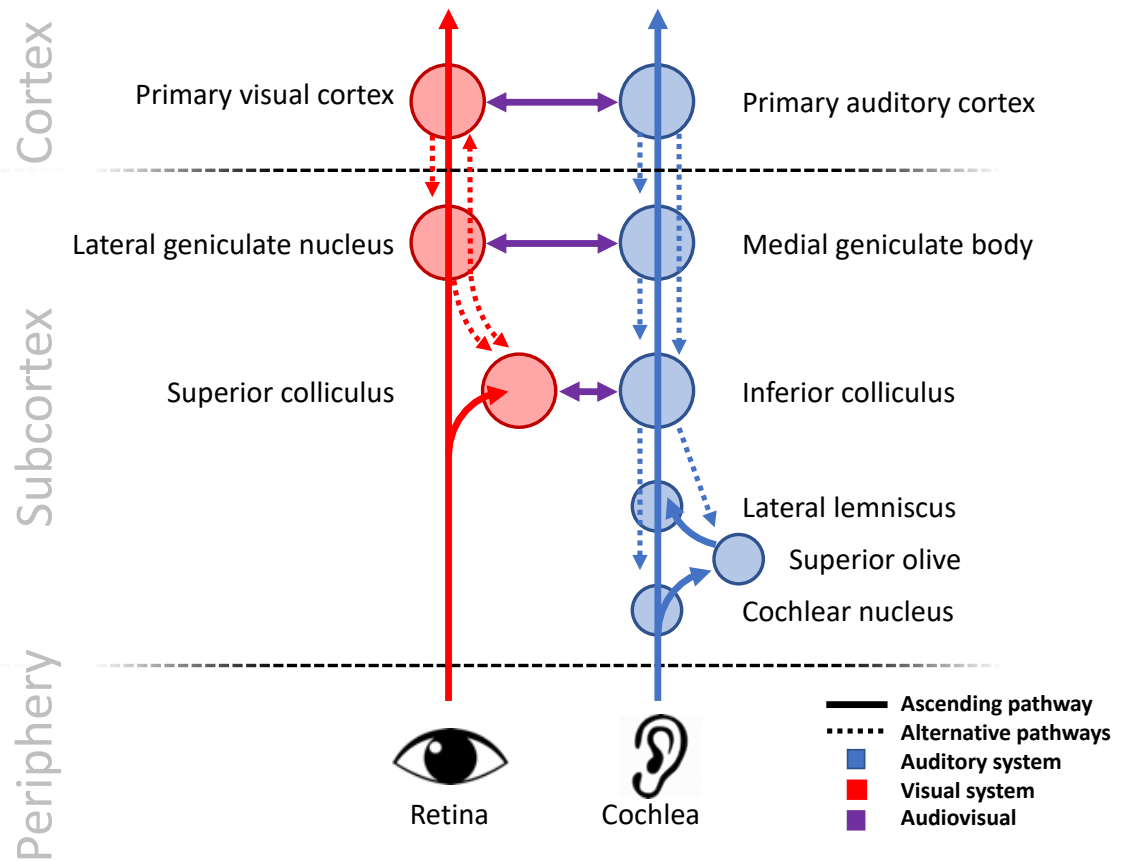
Figure 1.1 | Ascending and non-canonical pathways of sensory information transmission

movement and sounds (Nelson et al., 2013), however some motor-auditory processing may also occur locally in both the AC and the thalamus (Schneider and Mooney, 2018). Given this mutual connection between visual, auditory, and motor regions, it is reasonable to speculate whether the AC plays a pivotal role in the integration of these three signals in V1. As Chapter 2 characterizes this tripartite interaction in V1, Chapter 3 builds on this by investigating how neuronal populations and activity in the AC are involved in this V1 signal integration, a broader, more nuanced multisensory context than prior studies of AC activity.

Work has also been done to demonstrate subcortical cross-sensory connections, especially the visual SC and auditory IC. Returning to the barn owl, spatial tuning in both the SC and the IC arises from an intercollicular network of axonal projections, particularly between the IC external nucleus and the superficial and deep layers of the SC (Brainard and Knudsen, 1993, 1995; Feldman and Knudsen, 1997; King et al., 1998). Cross-sensory corticofugal pathways have been identified as well. The IC receives projections from the somatosensory cortex, again largely terminating in the external nucleus (Lesicko et al., 2016). Optogenetically activating V1 also enhances auditory BOLD signal responses in the IC (Leong et al., 2018), providing a potential functional connection to the anatomical one (Cooper and Young, 1976). And cross-sensory subcortical-cortical pathways have also been hinted at, such as via indirect connections from the mesencephalic locomotor region to V1 (Shik et al., 1966; Lee et al., 2014).

There are various techniques that are used to reveal neuronal connections between brain regions. Viral tracing using AAVs is convenient to fluorescently label axons and cell bodies, and with specific AAVs being engineered for both intracellular and transsynaptic

13

retrograde labeling. A transsynaptic anterograde labeling technique has been more elusive. AAVs can also be used to express photosensitive optogenetic opsins in neurons, a technique allowing the activation or suppression of specific neurons to probe their functional role of their anatomic connections. Pharmacology, while able to be used in coordination with viral and genetic techniques for specific labeling, can also be used to manipulate the activity of broader populations of neurons. By exogenously changing neuronal activity, optogenetic and pharmacologic techniques have the ability to demonstrate causality in a neuronal population's role in providing information to another region. These are useful innovations that build on older techniques from the field that have limited neuronal specificity.

Chapters 3 and 4 both use these techniques to characterize and probe the cortical and subcortical circuits underlying audiovisual integration, providing a deeper understanding of the neuronal codes and circuits that support multisensory perceptual and behavioral changes. As explained earlier, these perceptual and behavioral changes often entail improved sensory processing, as indicated by various multisensory psychophysics studies. Therefore, understanding how improved neuronal encoding supports these behavioral improvements requires an understanding of how neurons represent sensory information. We therefore finish by discussing the techniques to assess, quantify, and compare this sensory information encoding.

The representation and decoding of sensory and neuronal information

In the sensory nervous system, action potentials represent information about the external sensory environment. When incoming signals to a neuron reach a critical threshold

and an action potential is fired, this relays information to other nearby and distant downstream neurons about the stimulus to which the initial neuron was responding. This communication is variable, though, depending on both the internal state of the brain as well as external experimental factors. Therefore, it is useful to understand the neuronal activity as having a signal component, i.e. meaningful information being represented, and a noise component, i.e. uninformative variability in the communication stream.

The concepts of signal and noise form a convenient basis for understanding how well a neuron's activity represents the information that is being communicated, and the signal-to-noise ratio (SNR) quantifies this relationship. The discriminability index metric *d'*, derived from signal detection theory, is another approach to quantifying encoded information with applications in neuroscience and psychology (Stanislaw and Todorov, 1999). The *d'* index measures the separability between two distributions, which may represent a neuron's baseline and stimulus-evoked activity, or a neuron's response to two unique stimuli. The discriminability of these stimuli would be aided by greater separation between the response distribution means (signal) or reduced distribution widths (noise). Information theory is another theoretical framework useful in quantifying neuronal encoding efficiency (Borst and Theunissen, 1999). In particular, the mutual information (MI) metric rests on the principles of bits and entropy to measure how informative one variable, e.g. neuronal activity, is about another variable, e.g. sensory stimulus identity. If a neuron adopts a certain firing pattern following only a specific sensory stimulus (signal), that neuron's activity is informative about the stimulus identity. In contrast, if a neuron's firing pattern may correspond to multiple sensory stimuli (noise), the neuron's activity does not reduce uncertainty about the stimulus identity and is therefore less informative.

Ultimately, signal detection and information theories offer approaches to quantify the information represented in neuronal activity, with higher d' and MI values corresponding to improved sensory encoding. Chapter 2 uses both of these techniques to quantify how well V1 neurons encode the visual stimulus, comparing the unisensory visual to the audiovisual condition.

An additional approach to understanding the information encoded in sensory neuronal activity is to attempt to directly decode the stimulus from the neuronal activity. The decoding technique usually relies on principles derived from statistics or machine learning, where a classifier is trained on a certain set of stimulus and neuronal response data and then later tested on neuronal responses and the corresponding stimulus identities excluded from the training set. One common decoding algorithm is based on maximum-likelihood estimation, which uses Bayesian inference to identify the most likely stimulus identity from the neuronal activity. This approach has been applied to the visual system to measure the effects of noise on sensory representations (Montijn et al., 2014, Stringer et al., 2021). Another decoding algorithm uses a support vector machine to separate and classify response data using a multidimensional linear plane, a technique that has been applied to both the visual and auditory systems (Pagan et al., 2013; Wood et al., 2020). The accuracy of these decoders is a proxy for how well the information is represented by the neurons and can be parsed out and used by downstream brain regions. Chapter 2 again uses these approaches to demonstrate that both individual and populations of V1 neurons more accurately represent the visual stimulus with sound than without sound, providing more evidence for the neuronal mechanisms supporting psychometric improvements.

Sensory neuroscience relies on these tools of information quantification and stimulus decoding to understand, measure, and compare how neuronal systems represent the sensory environment. These techniques are useful because they can be applied to both individual and populations of neurons. Furthermore, these tools are agnostic to the specific units of the input data. In other words, whether the neurons are encoding stimuli through firing rate magnitude, spike timing, or a combination of the two, the concept of distinguishing an informative signal from uninformative variability still applies. However, low stimulus discriminability or decoding accuracy does not necessarily mean the neuron is insensitive to the stimulus feature of interest. It is possible that a neuron uses a different dimension of its neuronal activity than the one directly measured to encode the sensory information. Therefore, these techniques are more useful in demonstrating the presence of information as opposed to the absence. Ultimately, these approaches have been essential to building an understanding of how the brain represents information in a unisensory versus multisensory environment, and we will use them to compare how audiovisual and multisensory integration affect sensory neuronal encoding.

CHAPTER 2: AUDIOVISUAL INTEGRATION AND NEURONAL ENCODING IN MOUSE PRIMARY VISUAL CORTEX

ABSTRACT

In everyday life, we integrate visual and auditory information in routine tasks such as navigation and communication. While it is known that concurrent sound can improve visual perception, the neuronal correlates of this audiovisual integration are not fully understood. Specifically, it remains unknown whether improvement due to sound of detection and discriminability of visual stimuli is reflected in the neuronal firing patterns in the primary visual cortex (V1). Furthermore, presentation of the sound can induce movement in the subject, but little is understood about whether and how sound-induced movement contributes to V1 neuronal activity. Here, we investigated how sound and movement interact to modulate V1 visual responses in awake, head-fixed mice and whether this interaction improves neuronal encoding of the visual stimulus. We presented visual drifting gratings with and without simultaneous auditory white noise to awake mice while recording mouse movement and V1 neuronal activity. Sound modulated the light-evoked

activity of 80% of light-responsive neurons, with 95% of neurons exhibiting increased activity when the auditory stimulus was present. Sound consistently induced movement. However, a generalized linear model revealed that sound and movement had distinct and complementary effects of the neuronal visual responses. Furthermore, decoding of the visual stimulus from the neuronal activity was improved with sound, an effect that persisted even when controlling for movement. These results demonstrate that sound and movement modulate visual responses in complementary ways, resulting in improved neuronal representation of the visual stimulus. This study clarifies the role of movement as a potential confound in neuronal audiovisual responses and expands our knowledge of how multimodal processing is mediated at a neuronal level in the awake brain.

SIGNIFICANCE STATEMENT

Sound and movement are both known to modulate visual responses in the primary visual cortex, however sound-induced movement has remained unaccounted for as a potential confound in audiovisual studies in awake animals. Here, authors found that sound and movement both modulate visual responses in an important visual brain area, the primary visual cortex, in distinct, yet complementary ways. Furthermore, sound improved encoding of the visual stimulus even when accounting for movement. This study reconciles contrasting theories on the mechanism underlying audiovisual integration and asserts the primary visual cortex as a key brain region participating in tripartite sensory interactions.

INTRODUCTION

Our brains use incoming sensory information to generate a continuous perceptual experience. The neuronal systems underlying sensory perceptions of different modalities interact in a way that often improves perception of the complementary modality (Gingras et al., 2009; Gleiss and Kayser, 2012; Bigelow and Poremba, 2016; Hammond-Kenny et al., 2017; Meijer et al., 2018; Stein et al., 2020). In the audiovisual realm, it is often easiest to understand what someone is saying in a crowded room by additionally relying on visual cues such as lip movement and facial expression (Maddox et al., 2015; Tye-Murray et al., 2016). The McGurk effect and flash-beep illusion are other common perceptual phenomena that demonstrate mutual interactions between the auditory and visual systems (McGurk and MacDonald, 1976; Shams et al. 2002). Despite this current awareness of audiovisual integration at a perceptual level, a detailed understanding of the neuronal codes that mediate this improvement has proved elusive.

Previous studies of neuronal correlates of audiovisual integration found that the primary sensory cortical areas participate in this process (Wang et al., 2008; Ibrahim et al., 2016; Meijer et al., 2019; Deneux et al., 2019). The primary visual cortex (V1) contains neurons whose light-evoked firing rates are modulated by sound, as well as neurons that are responsive to sound alone (Knöpfel et al., 2019). Orientation and directional tuning of individual neurons are also affected by sound. In anesthetized mice, layer 2/3 neurons in V1 exhibited sharpened tuning in the presence of sound (Ibrahim et al., 2016). But another study in awake mice found no average differences in visual tuning curve bandwidth with and without sound (Meijer et al., 2017). These contrasting findings raise the question of whether the multisensory perceptual improvements described above are reflected in

20

individual V1 neurons in the awake brain. Furthermore, awake animals are subject to brain-wide changes in neuronal activity due to stimulus-aligned, uninstructed movements (Musall et al., 2019), a factor yet unaccounted for in most audiovisual studies.

Sound-induced movement represents a potential confound for audiovisual studies in awake animals because whisking and locomotion modulate neuronal activity in the sensory cortical areas. In V1, movement enhances neuronal visual responses and improves neuronal encoding of the visual scene (Niell and Stryker, 2010; Dardalat and Stryker, 2017). Conversely, in the auditory cortex (AC), locomotion generally suppresses neuronal spontaneous and auditory responses (Nelson et al., 2013; Schneider and Mooney, 2018; Bigelow et al., 2019). Therefore, movement is an important factor in neuronal sensory responses that often correlates with stimulus features.

Thus, audiovisual integration in V1 may not simply represent afferent information from auditory brain regions, as supported by studies demonstrating that V1 neurons are sensitive to the optogenetic stimulation (Ibrahim et al., 2016) and pharmacologic suppression (Deneux et al., 2019) of AC neurons. Indeed, the modulation of V1 activity may instead by a byproduct of uninstructed sound-induced movements which themselves modulate visual responses (Bimbard et al., 2021). However, because previous studies were either performed in anesthetized subjects (Ibrahim et al., 2016), or trials during which the mouse moved were excluded from analysis (Deneux et al., 2019) or pooled together (Iurilli et al., 2012; Meijer et al., 2017), these alternative explanations have not been quantified. We tested to what extent locomotion contributed to audiovisual integration in V1 by performing extracellular recordings of neuronal activity in V1 concurrent with monitoring movement in awake mice presented with audiovisual stimuli. We found that the majority

of neurons in V1 were responsive to visual and auditory stimuli. We found that sound and movement exerted distinct yet complementary effects on shaping the visual responses. Importantly, sound improved discriminability of the visual stimuli both in individual neurons and at a population level, an effect that persisted when accounting for movement.

RESULTS

Sound enhances the light-evoked firing rate of a subset of V1 neurons

Previous work identified that sound modulates visual responses in V1 (Ibrahim et al., 2016; Meijer et al., 2018; McClure and Polack, 2019), yet how that interaction affects stimulus encoding in individual neurons and as a population remains unclear. Furthermore, whether that interaction can be exclusively attributed to sound or rather to sound-induced motion is controversial (Bimbard et al., 2021). To elucidate the principles underlying audiovisual integration, we presented audiovisual stimuli to awake mice while performing extracellular recordings in V1 (Figure 2.1A). The visual stimulus consisted of drifting gratings in 12 directions presented at 5 visual contrast levels (Figure 2.1B). On half of the trials, we paired the visual stimulus with a 70-dB burst of white noise from a speaker positioned next to the screen (Figure 2.1C), affording 10 trials of each unique audiovisual stimulus condition (Figure 2.1C). Twelve recording sessions across six mice were spike sorted, and the responses of these sorted neurons were organized by trial type to compare across audiovisual stimulus conditions. Figure 2.1D-G demonstrates an example unit tuned for gratings aligned to the 30°-210° axis whose baseline and light-evoked firing rate are increased by the sound.

Sound modulated the activity of the majority of V1 neurons. We used a generalized linear model (GLM) to classify neurons as light-responsive and/or sound-responsive based on their firing rate at the onset (0-300 ms) of each trial. Using this classification method, we found that 86.2% (703/816) of units were responsive to increasing visual stimulus contrast levels, and of these visually responsive units, 80.1% (563/703 neurons, 12 recording sessions in 6 mice) were significantly modulated by the presence of sound (Figure 2.2A). We constructed an average PSTH from the response profiles of sound-modulated light-responsive neurons, which revealed that the largest change in light-evoked firing rate occurs at the onset of the stimulus (Figure 2.2B). Averaged across neurons, we found a robust increase in the magnitude of the visually evoked response across visual contrast levels (Figure 2.2C; p(vis)=1.2e-100, p(aud)=1.6e-88, p(interact)=5.7e-4, paired 2-way ANOVA; $p_{c=0}$=2.1e-51, $p_{c=0.25}$=2.6e-62, $p_{c=0.5}$=5.7e-75, $p_{c=0.75}$=1.1e-81, $p_{c=1}$=2.0e-81, post hoc Bonferroni-corrected paired t-test, Table 1). This difference was driven by the majority of neurons (95%) that increased their firing rate in the presence of sound. However, some neurons exhibited lower light-evoked and sound-evoked firing rates relative to baseline.

This change in firing rate can be described as supra-linear or sub-linear based on whether the audiovisual response is greater or less than, respectively, the sum of the unimodal light-evoked and sound-evoked firing rates. At medium to high visual contrast levels, integration of the audiovisual stimulus was predominantly supra-linear (Figure 2.2D-E; p=1.6e-12, Kruskal-Wallis test; $p_{c=0.25}$=0.053, $p_{c=0.5}$=0.004, $p_{c=0.75}$=4.6e-8, $p_{c=1}$=2.1e-5, post hoc Bonferroni-corrected Wilcoxon signed rank test, Table 1). In
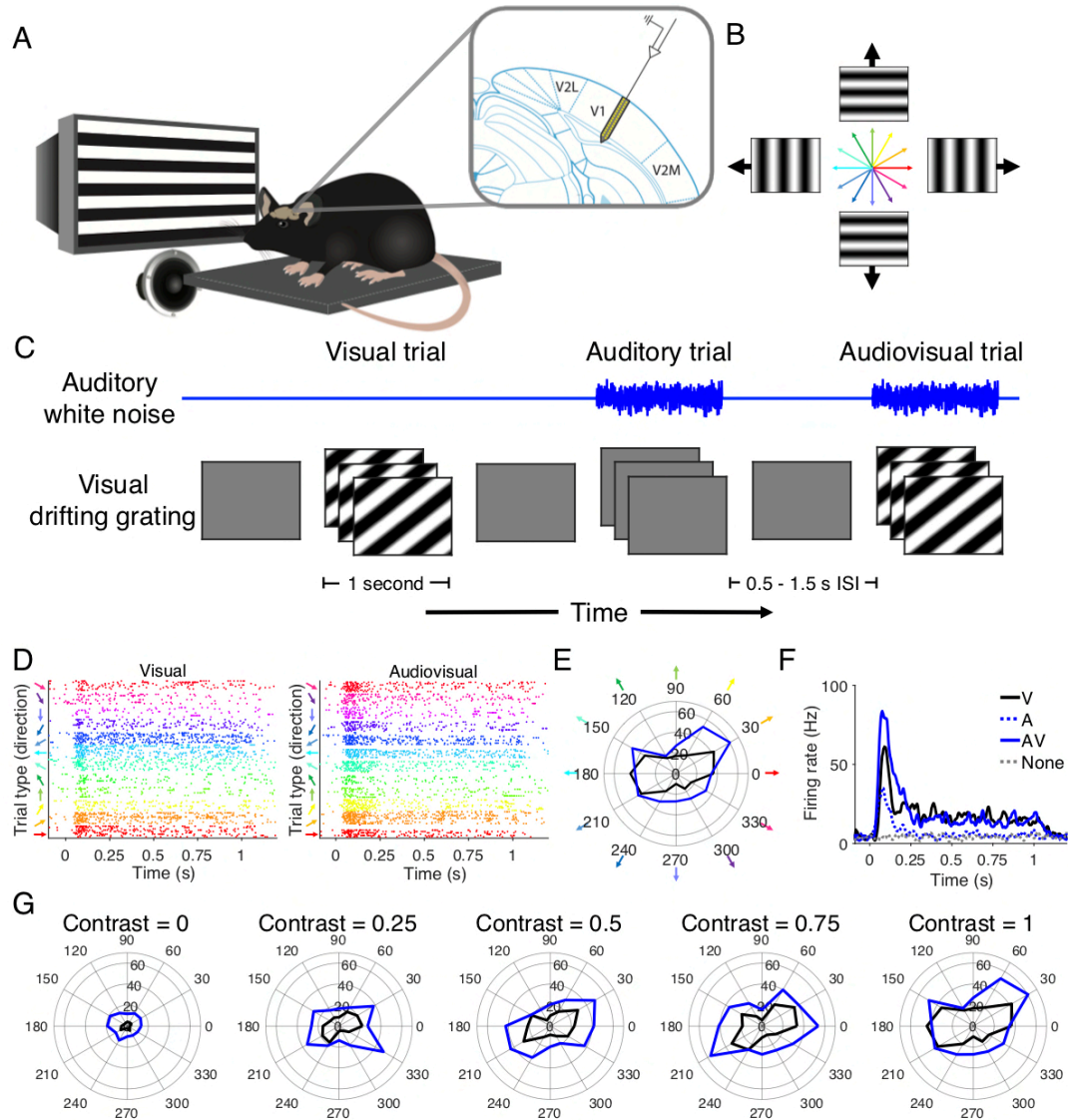
Figure 2.1 | Audiovisual stimulus presentation (A) Diagram (left) demonstrating that mice were head-fixed and presented with audiovisual stimuli from the right spatial field while electrophysiological recordings were performed in V1 (right). (B) Visual stimuli consisted of drifting gratings of 12 directions. (C) Auditory, visual, and audiovisual trials were randomly ordered and spaced with variable inter-stimulus intervals. (D) Raster plots of visual (left) and audiovisual (right) trials of an example neuron. (E) Polar plot demonstrating the orientation tuning and magnitude of response (Hz) of the same example neuron in *E*. (F) PSTH of the same neuron in *E* demonstrating enhanced firing in response to audiovisual stimuli compared to unimodal stimuli. (G) Example neuron in *E* displays enhanced firing rate with sound across visual contrast levels.

summary, these results show that sound supra-linearly increases the magnitude of the light-evoked response in the majority of V1 neurons.

Sound reduces the orientation- and direction-selectivity of tuned neurons

Having observed sound-induced changes in the magnitude of the visual response, we next assessed whether these changes in magnitude affected neuronal tuning. V1 neurons have receptive fields tuned to a specific visual stimulus orientation and, to a lesser extent, stimulus direction (Métin et al, 1988; Rochefort et al., 2011; Fahey et al., 2019). We first tested whether sound altered tuning preferences of V1 neurons. In light-responsive neurons, we calculated the orientation and direction-selective indices (OSI and DSI) as well as pseudo indices based on random permutations of the trials (see Methods), and classified neurons in which the true indices were >95% of the pseudo indices as "orientation-" or "direction-selective." Using this stringent selection criterion, we found that 13.9% (78/563) of neurons were orientation-selective, whereas 2.1% (12/563) were direction-selective. In these neurons, we observed shifts in the preferred direction from the visual to audiovisual condition (Fig 2.2 Sup 1A). This shift in visual tuning preference may be due to auditory input, or it may reflect noise in the neuronal responses. To test this, we performed an additional permutation test by repeatedly sampling the visual responses. We found that the resulting distribution of preferred direction shifts resembled the observed distribution under the audiovisual condition (Fig 2.2 Sup 1B), and the observed mean shift in degrees was within the limits of the sampled distribution (Fig 2.2 Sup 1C). Therefore, we cannot conclude that the shift in directional tuning preferences is associated with the presence of sound.

25
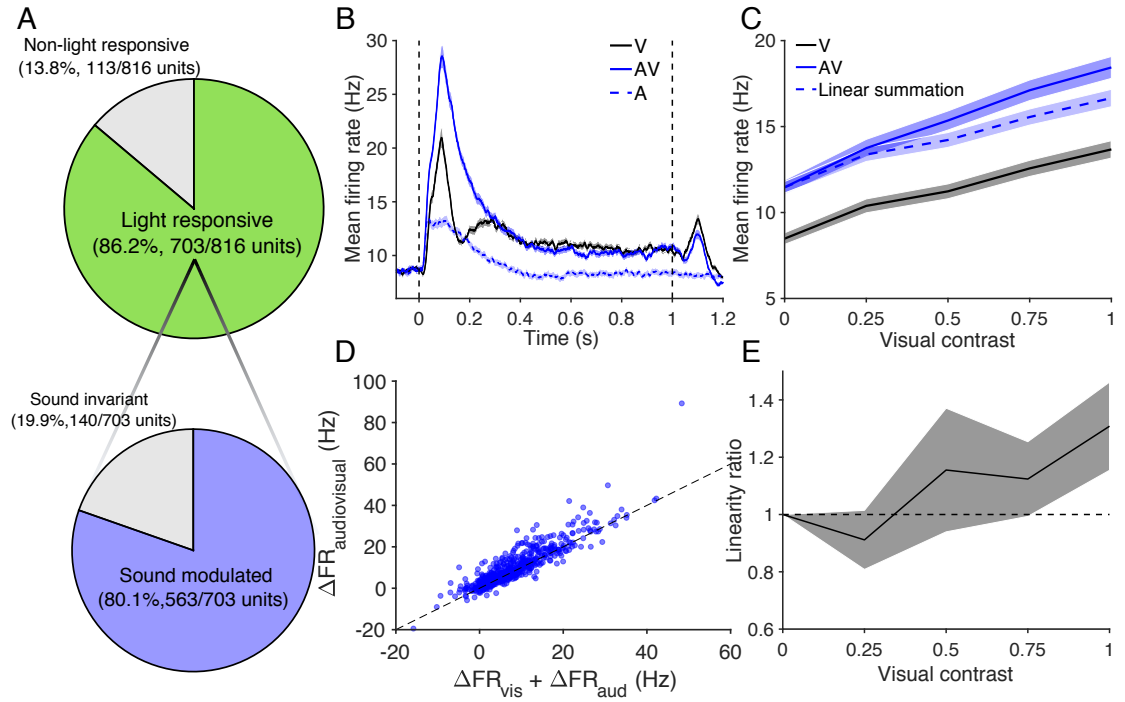
Figure 2.2 | Sound enhances visual responses in a supra-linear manner (A) Sound modulates visually evoked activity in 80.1% of light-responsive neurons in V1. (B) Comparison of visual, auditory, and audiovisual PSTHs averaged across all light-responsive sound-modulated neurons. Visual and audiovisual PSTHs correspond to the highest visual contrast level. (C) The magnitude of audiovisual onset responses (0-300ms) is greater than that of the visual response in light-responsive sound-modulated neurons (n=563, p(vis)=1.2e-100, p(aud)=1.6e-88, p(interact)=5.7e-4, 2-way repeated measures ANOVA; post hoc Bonferroni-corrected paired t-test). The expected linear sum of the unimodal auditory and visual responses is included. (D) At full visual contrast, the observed audiovisual response in the majority of neurons is greater than the linear sum of the unimodal auditory and visual responses. (E) A linearity ratio above 1 demonstrates audiovisual responses in V1 represent supra-linear integration of the unimodal signals (n=563, p=1.6e-12, Kruskal-Wallis test, post hoc Bonferroni-corrected Wilcoxon signed rank test).

In addition to testing a shift in preferred direction, we investigated whether sound altered the neurons' tuning selectivity. Tuning selectivity captures how strongly an individual neuron responds to stimuli of a certain condition, e.g. grating orientation and drift direction, as compared to others. We found a small reduction in the OSI from the visual to audiovisual conditions (Fig 2.2 Sup 1D-E; p=0.0018, paired Student's t-test), which may reflect disproportionate changes in firing rate at the preferred versus orthogonal directions. We also found a reduction in the DSI in the presence of sound (Fig 2.2 Sup 1F-G; p=0.021, paired Student's t-test). Combined, these results suggest that sound's enhancement of the magnitude of light-evoked responses has minimal or potentially diminishing effects on the tuning selectivity of neurons.

Sound reduces the latency, increases onset duration, and decreases variability of visual responses in neurons

Behaviorally, certain cross-modal stimuli elicit shorter reaction times than their unimodal counterparts (Diederich and Colonius, 2004; Colonius and Diederich, 2017; Meijer et al., 2018). Therefore, we hypothesized that sound reduces the latency of the light-evoked response at a neuronal level as well. For each neuron, we calculated the response latency as the first time bin after stimulus onset at which the firing rate exceeded 1 standard deviation above baseline (Fig 2.2 Sup 2A), and found that sound reduced the response latency across contrast levels (Fig 2.2 Sup 2B; p(vis)=6.9e-4, p(aud)=6.8e-15, p(interact)=0.045, paired 2-way ANOVA; $p_{c=0.25}$=2.3e-4, $p_{c=0.5}$=7.1e-12, $p_{c=0.75}$=4.6e-5, $p_{c=1}$=9.9e-4, post hoc Bonferroni-corrected paired t-test, Table 1). We additionally calculated the slope of the onset response of light-responsive sound-modulated neurons,

27

measured from trial onset until the time at which each neuron achieved its peak firing rate (Fig 2.2 Sup 2C). We found that sound increased the slope of the onset response (Fig 2.2 Sup 2D; p(vis)=3.5e-121, p(aud)=2.7e-15, p(interact)=0.038, paired 2-way ANOVA; $p_{c=0.25}$=1.4e-4, $p_{c=0.5}$=8.9e-13, $p_{c=0.75}$=3.6e-12, $p_{c=1}$=5.5e-8, post hoc Bonferroni-corrected paired t-test, Table 1), both indicating that the response latency was reduced in the audiovisual condition compared to the visual condition. Additionally, the duration of the light-evoked response, defined as the full width at half maximum of the peak onset firing rate, increased in the presence of sound (Fig 2.2 Sup 2E,F; p(vis)=1.3e-10, p(aud)=8.7e-98, p(interact)=0.23, paired 2-way ANOVA). Both of these timing effects were relatively constant across contrast levels. Therefore, the latency and onset duration of light-evoked responses in V1 neurons is enhanced by sound.

Having observed changes in response magnitude and timing, we next investigated the effect of sound on the variability of light-evoked responses. If individual neurons encode the visual stimulus using changes in their firing rate, a more consistent response would entail less spread in the response magnitude relative to the mean response across trials of a single stimulus type. We quantified this relationship using the coefficient of variation (CV) defined as the ratio of the standard deviation to the response mean (Gur et al., 1997). We hypothesized that sound reduces the CV of light-evoked responses, corresponding to reduced response variability and higher SNR. Fig 2.2 Sup 2G depicts the relationship between response magnitude and CV in an example sound-modulated light-responsive neuron, demonstrating that increased response magnitude correlates with reduced CV. Consistent with sound increasing the visual response magnitude in the majority of sound-modulated light-responsive neurons (Figure 2.2), we observed a

28

reduction of CV in the audiovisual condition relative to the visual condition when averaged across these neurons (Fig 2.2 Sup 2H; p(vis)=0.28, p(aud)=4.2e-103, p(interact)=0.38, paired 2-way ANOVA). Taken together, these results indicate that sound not only modulates the magnitude of the visual response (Figure 2.2), but also improves the timing and consistency of individual neurons' responses (Fig 2.2 Sup 2).

Sound-induced movement does not account for sound's effect on visual responses

It is known that whisking and locomotive behaviors modulate neuronal activity in mouse visual cortex (Niell and Stryker, 2010) and auditory cortex (Nelson et al., 2013; Schneider and Mooney, 2018; Bigelow et al., 2019). Therefore, having established that sound robustly modulates visual responses (Figure 2.2), we tested whether these observed changes were more accurately attributable to sound-induced movement. In an additional cohort of mice, we performed V1 extracellular recordings with the same audiovisual stimuli described above while recording movement activity of the mice throughout stimulus presentation. We found that sound did evoke whisking and locomotive behavior in mice, leading to increased movement on audiovisual trials compared to visual trials (Figure 2.3A; p=9.1e-5, paired t-test). However, there were many visual trials in which substantial movement occurred, as well as audiovisual trials in which little movement was detected (Figure 2.3B). Because of this large variability in sound-induced movement, we were able to control for movement when comparing visual and audiovisual activity in the recorded neurons.

Similar to above, we used a GLM to classify each neuron as light-, sound-, and/or motion-responsive based on the neuron's firing rate and mouse's movement activity during

the onset (0-300ms) of the trial. The vast majority of light-responsive neurons, 71.1% (249/350), displayed both sound- and motion-modulated visual responses (Figure 2.3C). 11.1% (39/350) and 5.2% (18/350) of light-responsive neurons were purely sound- or motion-modulated, respectively. An additional 12.6% (44/350) were invariant to sound or motion. We then compared the visually and audiovisually evoked firing rates of neurons when controlling for movement. Among sound- and motion-modulated light-responsive neurons, the firing rate was higher on audiovisual trials than visual trials when movement was held constant (Figure 2.3D), especially when mice showed limited movement.

On trials in which the mice were largely stationary (z-score<-0.5, 43% of visual trials, 32% of audiovisual trials) or displayed moderate levels of movement (-0.5<z-score<1.5, 51% of visual trials, 57% of audiovisual trials), the mean firing rate of neurons was 54-62% higher when sound was presented than when sound was absent. The firing rates under the two stimulus conditions converged on trials in which the mice displayed high movement activity (z-score>1.5, 4.8% of visual trials, 11% of audiovisual trials; Figure 2.3D,E; p(move)=0.010, p(aud)=1.4e=13, p(interact)=1.8e-8, unbalanced 2-way ANOVA; $p_{stationary}$=1.5e-14, $p_{low\ motion}$=7.1e-10, $p_{high\ motion}$=0.6, post hoc Bonferroni-corrected two-sample t-test, Table 1). Notably, increasing movement activity was correlated with increased firing rates on visual trials, but was correlated with decreasing firing rates among audiovisual trials (Figure 2.3E). These results indicate that sound modulated visually evoked neuronal activity even when accounting for sound-induced movement in awake mice.
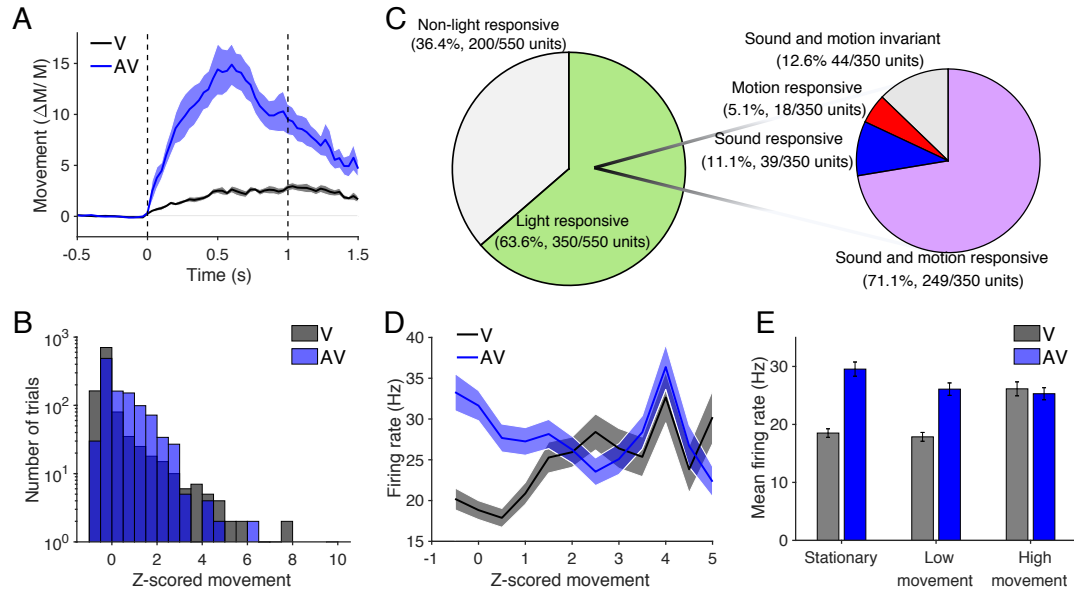
Figure 2.3 | Sound modulates visual activity when controlling for stimulus-induced movement (A) Mice displayed more movement response to audiovisual trials than in visual trials (n=9 recording sessions; p=9.1e-5, paired t-test). (B) Histogram of trials' z-scored movements show a range of levels of movement during both visual and audiovisual trials. (C) Venn diagram demonstrating that 87% of light-responsive neurons exhibited some combination of sound- and movement-responsiveness. (D) Comparison of firing rate of sound- and motion-modulated light-responsive neurons across trials with a range of z-scored movement. (E) Responses to audiovisual stimuli evoke larger magnitude responses than visual stimuli when mice were stationary (z-score<-0.5) or displayed low to moderate movement (-0.5<z-score<1.5), but responses were not significantly different when mice displayed the highest amount of movement (z-score>1.5; p(motion)=0.001, p(aud)=1.4e-13, p(interact)=1.8e-8, 2-way ANOVA, post hoc Bonferroni-corrected two-sample t-test)

Sound and movement have distinct and complementary effects on visual responses

To further parse out the role of sound and movement on audiovisual responses, we used a separate GLM to capture the time course of these parameters' effects on visual activity. For each neuron, we used a GLM with a sliding 10ms window to reconstruct the PSTH based on the visual contrast level, sound presence, and movement during that time window (Figure 2.4A). Figure 2.4B shows an example neuron in which the GLM accurately captures the light-evoked, sound-evoked, and audiovisually evoked PSTHs using the average movement for each trial type. Across neurons, the GLM-estimated PSTHs accurately reconstructed observed PSTHs, with the highest correlation when all parameters were included in the estimate (Figure 2.4C-E). We leveraged the coefficients fit to each neuron (Figure 2.4A) to estimate the unique contribution of each predictor to the firing rates as a function of time (see Materials and Methods). In the absence of movement, sound predominantly enhanced neuronal activity at the onset of the visual response and suppressed activity during the response's sustained period (Figure 2.4F; n=295 fitted neurons, paired t-test at each time window [1391], α=3.6e-5). Conversely, movement had little effect on the onset activity in the absence of sound, but rather enhanced firing rates during the response's sustained period (Figure 2.4G; n=295 fitted neurons, paired t-test at each time window [1391], α=3.6e-5). Together, sound and movement have complementary effects in which both the onset and sustained portions of the visual response are enhanced (Figure 2.4H; n=295 fitted neurons, paired t-test at each time window [1391], α=3.6e-5). Again notably, the peak onset response under the audiovisual condition was lower when movement was included in the estimate (Figure 2.4H). These findings indicate not only that movement is unable to account for the changes in onset response reported above, but

32
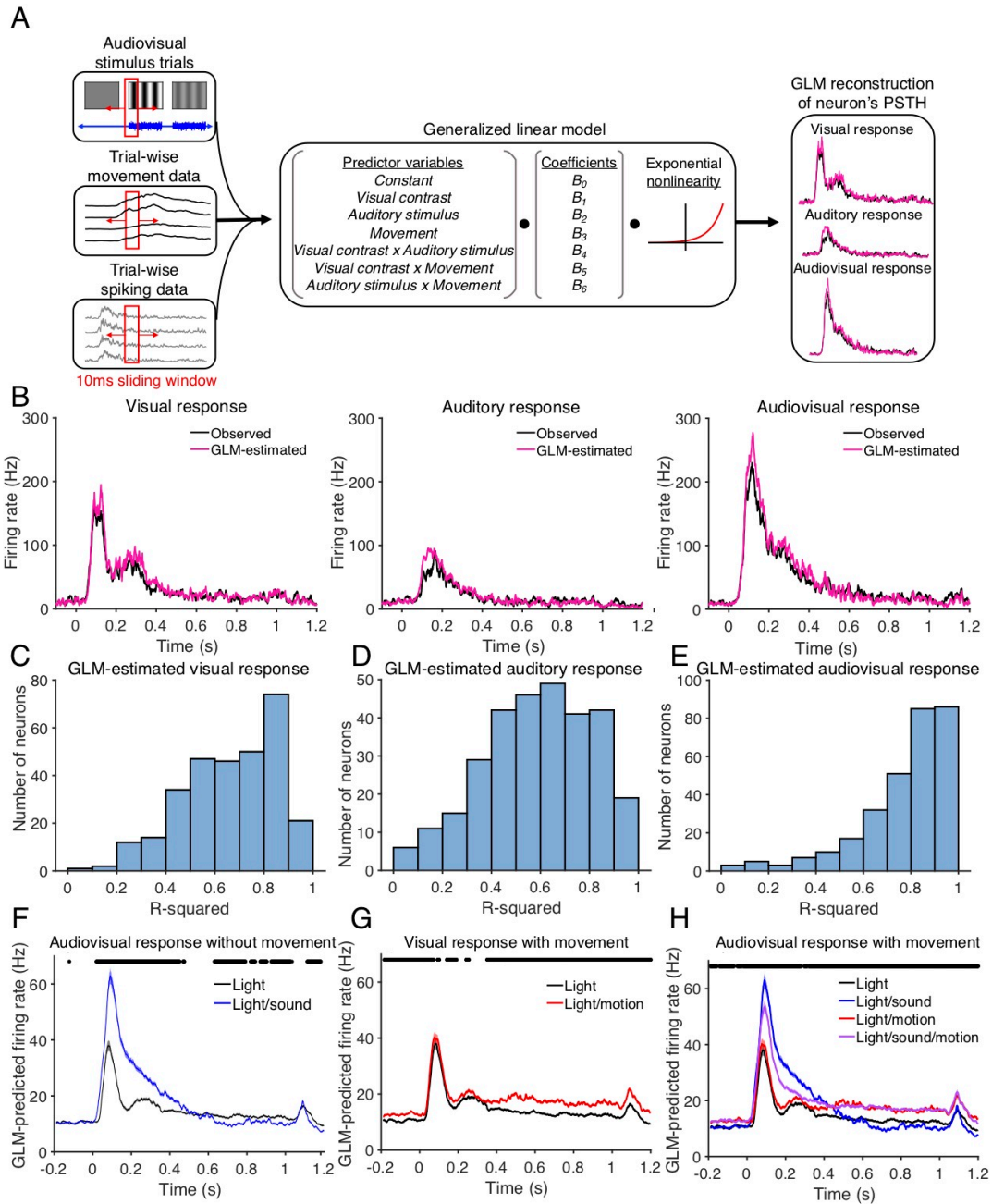
Figure 2.4 | Sound and movement modulate visual responses in distinct but complementary ways (A) Diagram illustrating the use of a GLM to reconstruct individual neurons' PSTHs based on neuronal responses and mouse movement during stimulus presentation. The GLM was then used to predict the time course of neuronal responses audiovisual stimuli with and without movement. (B) Observed trial-averaged PSTHs for visual-only (left), auditory-only (middle), and audiovisual (right) trials overlaid with GLM estimates based on the selected stimulus features. (C-E) Histograms demonstrating R^2 values of the GLM-estimated PSTHs, averaged across sound- and motion-modulated light-responsive neurons. Moderate to high R^2 values across the population indicate a good ability for the GLM to estimate neuronal firing rates. (F-H) GLM-predicted

33

visually evoked PSTHs with and without sound and motion. Asterisks indicate time windows in which there was a significant difference between the *light* prediction and the *light+sound*, *light+motion*, and *light+sound+motion* predictions, respectively. (F) Excluding motion highlights that sound primarily enhances the onset response. Asterisks indicate time windows in which there was a significant difference (n=295 fitted neurons; paired t-test, α=3.6e-5). (G) Excluding sound highlights that motion primarily enhances the sustained portion of the response. Asterisks indicate time windows in which there was a significant difference (n=295 fitted neurons; paired t-test, α=3.6e-5). (H) Sound and motion together enhance both the onset and sustained periods of the visually evoked response. (n=295 fitted neurons; paired t-test, α=3.6e-5).

also that sound and motion have distinct and complementary effects on the time course of visually evoked activity in V1.

Decoding of the visual stimulus from individual neurons is improved with sound

Behaviorally, sound can improve the detection and discriminability of visual responses, however whether that improved visual acuity is reflected in V1 audiovisual responses is unknown. Despite many studies reporting neuronal correlates of audiovisual integration in V1, whether sound improves neuronal encoding of the visual stimulus has yet to be demonstrated. The increase in response magnitude and decrease in CV suggest that sound may improve visual stimulus discriminability in individual V1 neurons. Consistent with these changes in response magnitude and variability, we observed sound-induced improvements in the *d'* sensitivity index between responses to low contrast drifting grating directions among orientation- and direction-selective neurons (Fig 2.5 Sup 1), further indicating improved orientation and directional discriminability in individual neurons. To directly test this hypothesis, we used the neuronal responses of individual neurons to estimate the visual stimulus drifting grating orientation and direction. We trained a maximum likelihood estimate (MLE)-based decoder (Montijn et al., 2014; Meijer et al., 2017) on trials from the preferred and orthogonal orientations in orientation-selective neurons and on trials from the preferred and anti-preferred directions in direction-selective neurons. We used leave-one-out cross-validation and cycled the probe trial through the repeated trials of the stimulus condition calculate the mean decoding performance. The MLE decoder's output was the orientation or direction with the maximum posterior likelihood based on the training data (Figure 2.5A). This decoding technique achieves high

decoding accuracy (Figure 2.5B). When averaged across sound-modulated orientation-selective neurons, decoding performance was improved on audiovisual trials compared to visual trials (Figure 2.5C; p(vis)=4.8e-112, p(aud)=7.8e-4, p(interact)=0.71, paired 2-way ANOVA), with the greatest improvements at low to intermediate contrast levels (Figure 2.5D). We applied this approach to sound-modulated direction-selective units and found similar trends towards improvements at low contrast levels (Figure 2.5E,F; p(vis)=2.1e-4, p(aud)=0.18, p(interact)=0.78, paired 2-way ANOVA), limited by fewer and weaker direction-selective neurons in V1. These results demonstrate that sound-induced changes in response magnitude and consistency interact in order to improve neuronal representation of the visual stimulus in individual neurons.

Population-based decoding of the visual stimulus improves with sound

V1 uses population coding to relay information about the various stimulus dimensions to downstream visual areas (Montijn et al., 2014, Berens et al., 2012), so we next tested whether these improvements in visual stimulus encoding in individual neurons extended to the population level. We began by training a support vector machine (SVM) to perform pairwise classification of visual drifting grating directions based on neuronal population activity. We again used a leave-one-out cross-validation approach when training and testing the SVM (Figure 2.6A). Decoding accuracy improved as more neurons were included in the population (Fig 2.6 Sup 1A), achieving an accuracy of ~90% when averaged across all pairwise orientation comparisons. At full visual contrast, there was little difference between the performance on visual and audiovisual trials. However, at low to intermediate visual contrast levels, classification performance robustly increased on
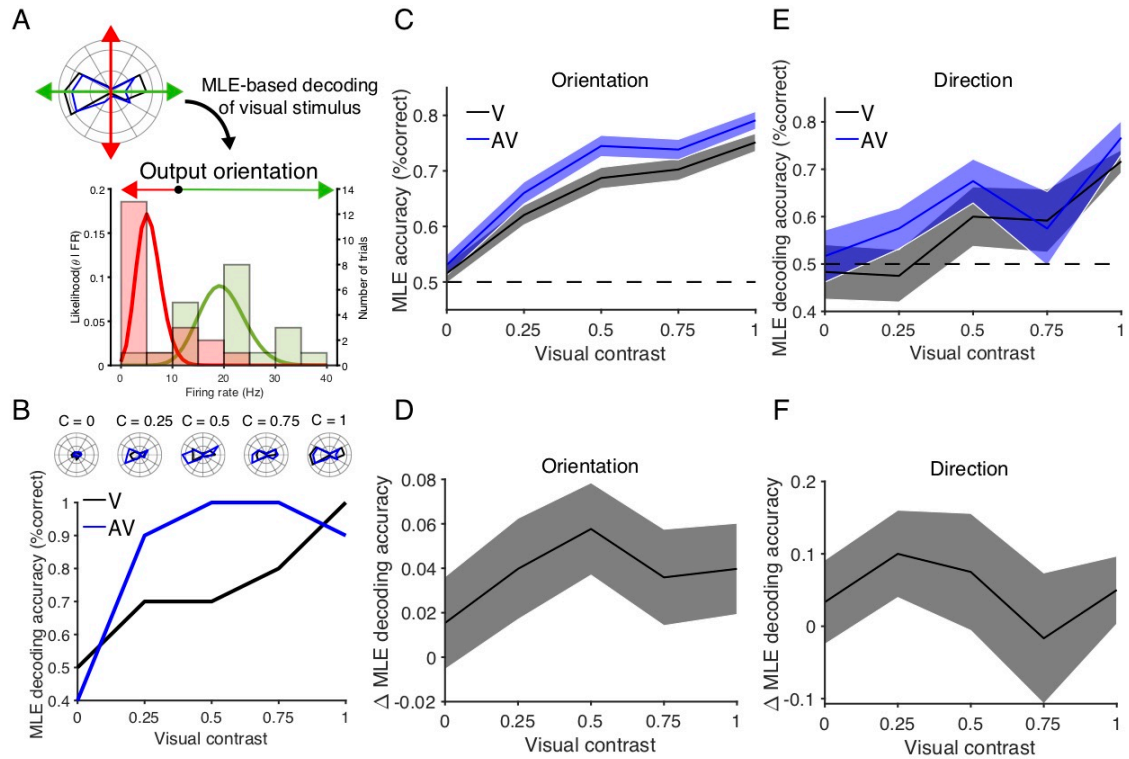
Figure 2.5 | Sound improves decoding of drifting grating direction and orientation in individual neurons (A) Diagram illustrating MLE-based decoding of an individual neuron's preferred versus orthogonal orientations. (B) Performance of the MLE decoder, trained on an example orientation-selective neuron, in decoding the neuron's preferred versus orthogonal orientations. The neuron's polar plots are shows in the above inset. (C-D) Absolute (C) and difference (D) in decoding accuracy of preferred versus orthogonal orientations, averaged across sound-modulated orientation-selective neurons, demonstrating higher performance in the audiovisual condition (n=78, p(vis)=4.8e-112, p(aud)=7.8e-4, p(interact)=0.71, paired 2-way ANOVA). (E-F) Absolute (E) and difference (F) in decoding accuracy of preferred versus anti-preferred directions, averaged across sound-modulated direction-selective neurons. No significant effect of sound on decoding accuracy was observed (n=12, p(vis)=2.1e-4, p(aud)=0.18, p(interact)=0.78, paired 2-way ANOVA).

audiovisual trials as compared to visual trials (Figure 2.6B). This improvement in performance was greatest when comparing orthogonal drifting grating orientations (Figure 2.6C; p(vis)=1.8e-61, p(aud)=1.9e-8, p(interact) = 2.4e-4, 2-way ANOVA; $p_{c=0}$=0.12, $p_{c=0.25}$=0.0016, $p_{c=0.5}$,=0.0014, $p_{c=0.75}$=0.0023; $p_{c=1}$=1, post hoc Bonferroni-corrected paired t-test, Table 1). However, a similar improvement was also observed in decoding opposite drifting grating directions (Figure 2.6D, p(vis)=1.1e-21, p(aud)=9.0e-9, p(interact)=0.0019, 2-way ANOVA; $p_{c=0}$=0.55, $p_{c=0.25}$=5.3e-5, $p_{c=0.5}$=0.0036, $p_{c=0.75}$=0.17, $p_{c=1}$=0.0036, post hoc Bonferroni-corrected paired t-test, Table 1). These results indicate that sound improves neuronal population encoding of grating orientation and drift direction.

Similar performance levels were also observed when decoding drifting grating orientation and direction using an MLE-based population decoder, indicating that the results were not specific to the decoding algorithm. Again, performance improved with increasing population sizes (Fig 2.6 Sup 1B), and accuracy was higher on audiovisual trials than visual trials (Figure 2.6E-G; orientation: p(vis)=2.3e-66, p(aud)=0.61, p(interact)=9.6e-11, 2-way ANOVA; $p_{c=0}$-5.8e-4, $p_{c=0.25}$=1.8e-4, $p_{c=0.5}$=0.3, $p_{c=0.75}$=0.53, $p_{c=1}$=0.15, post hoc Bonferroni-corrected paired t-test, Table 1; direction: p(vis)=4.6e-26, p(aud)=0.51, p(interact)=4.1e-6, 2-way ANOVA; $p_{c=0}$=0.037, $p_{c=0.25}$=6.4e-6, $p_{c=0.5}$=0.036, $p_{c=0.75}$-0.16, $p_{c=1}$=0.14, post hoc Bonferroni-corrected paired t-test, Table 1). Expanding on the SVM approach, the MLE-based decoder allowed us to perform not only pairwise classification, but also classification of 1 out of all 12 drifting grating directions. When trained and tested in this fashion, MLE decoding performance again improved at low to intermediate contrast levels on audiovisual trials (Figure 2.6H-I), before reaching
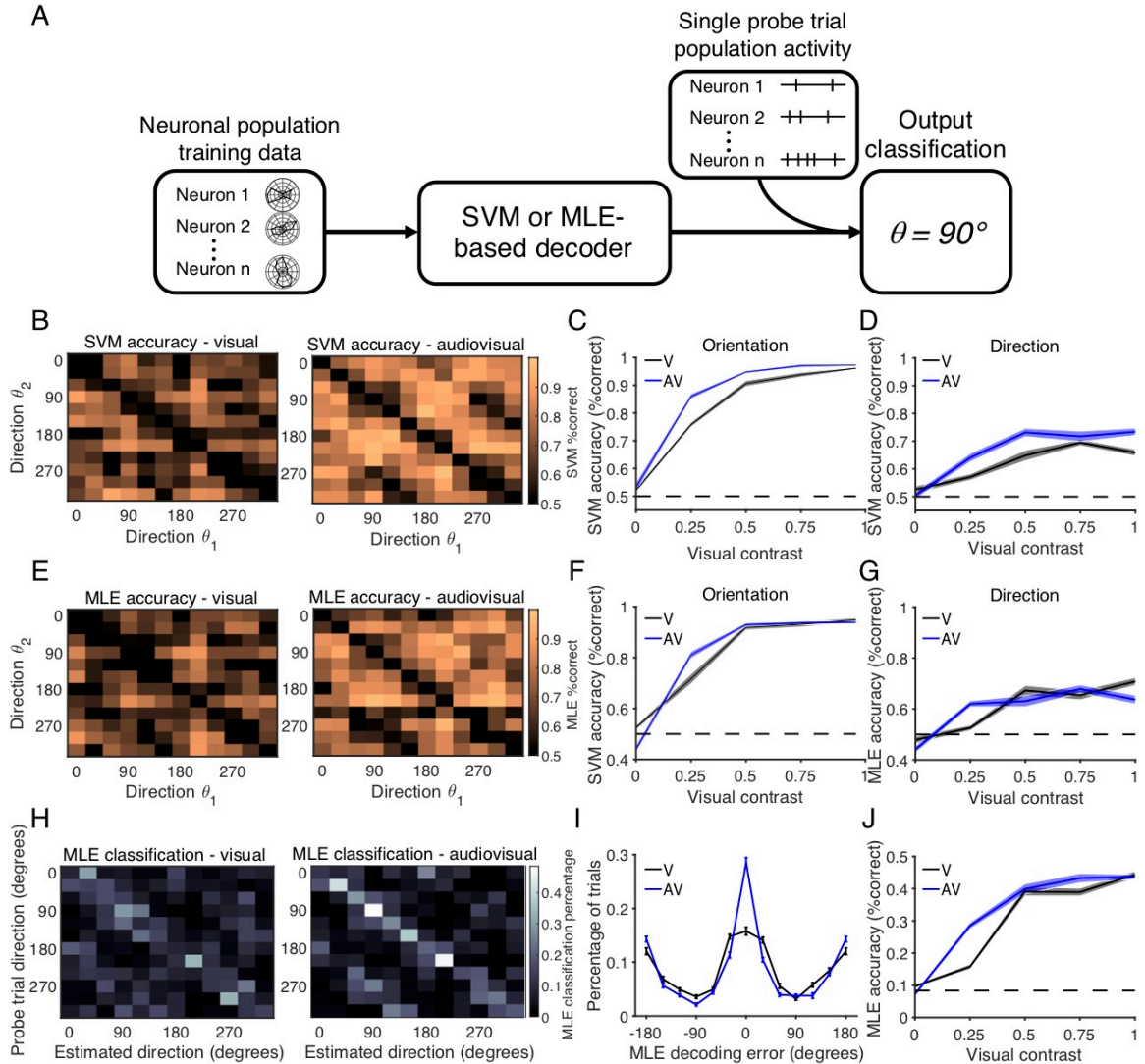
Figure 2.6 | Sound improves accuracy of population-based visual stimulus decoding (A) Schematic illustrating the decoding of the drifting grating direction using either an SVM or MLE decoder trained on neuronal population activity. (B) Accuracy of SVM pairwise classification of drifting grating directions on visual (left) and audiovisual (right) trials, contrast 0.25. (C) SVM decoding accuracy improved with sound when classifying orthogonal drifting grating orientations (n=10 randomizations, p(vis)=1.8e-61, p(aud)=1.9e-8, p(interact)=2.4e-4, 2-way ANOVA, post hoc Bonferroni-corrected paired t-test). (D) SVM decoding accuracy when classifying opposite drifting grating directions, demonstrating improved performance with sound (n=10 randomizations, p(vis)=1.1e-21, p(aud)=9.0e-9, p(interact)=0.0019, 2-way ANOVA, post hoc Bonferroni-corrected paired t-test). (E) Accuracy of MLE pairwise classification of drifting gratings on visual (left) and audiovisual (right) trials, contrast 0.25. (F) MLE decoding accuracy when classifying orthogonal drifting grating orientations improved with sound (n=10 randomizations, p(vis)=2.3e-66, p(aud)=0.61, p(interact)=9.6e-11, 2-way ANOVA, post hoc Bonferroni-corrected paired t-

39

test). (G) MLE decoding accuracy when classifying opposite drifting grating directions, demonstrating less effect of sound on performance (n=10 randomizations, p(vis)=4.6e-26, p(aud)=0.51, p(interact)=4.1e-6, 2-way ANOVA, post hoc Bonferroni-corrected paired t-test). (H) Heat map of actual vs MLE-output directions under visual (left) and audiovisual (right) trials, contrast 0.25. MLE decoder could choose between all 12 drifting grating directions. (I) MLE decoder classification percentage, comparing estimated direction to actual direction. (J) Overall decoding accuracy of MLE decoder when choosing between all 12 drifting grating directions improved with sound (n=20 randomizations, p(vis)=2.2e-92, p(aud)=1.9e-5, p(interact)=2.7e-11, 2-way ANOVA, post hoc Bonferroni-corrected paired t-test).

asymptotic performance of ~45% at full visual contrast (Figure 2.6J; p(vis)=2.2e-92, p(aud)=1.9e-5, p(interact)=2.7e-11, 2-way ANOVA; $p_{c=0}$=0.012, $p_{c=0.25}$=1.4e=10, $p_{c=0.5}$=0.48, $p_{c=0.75}$=0.0013, $p_{c=1}$=0.5, post hoc Bonferroni-corrected paired t-test, Table 1). Taken together, these results indicate that sound improves neuronal encoding of the visual stimulus both in individual neurons and at a population level, especially at intermediate visual contrast levels.

Sound improves stimulus decoding when controlling for sound-induced movements

It is known that locomotion improves visual processing in V1 (Dardalat and Stryker, 2017). We next tested whether the sound-induced improvement in visual stimulus representation (Figure 2.6) was attributable to sound's effect on visual responses or indirectly via sound-induced movement. We observed previously that sound was primarily responsible for enhancing the visual response onset, whereas motion enhanced the sustained portion (Figure 2.4). We therefore hypothesized that the improvement on MLE decoding performance, based on the visual response onset, would be present even when accounting for sound-induced uninstructed movements. We tested this hypothesis by expanding on the GLM-based classification of neurons described in Figure 2.3. Using the same GLM generated for each neuron, we modified the movement variable and its corresponding pairwise predictors to the lowest observed value, and then used the GLM coefficients and the exponential nonlinearity to estimate each neuron's audiovisual response magnitude when regressing out the effect of motion (Figure 2.7A, Materials and Methods). We then input these estimated trial-wise neuronal responses into the same MLE-based decoder described above. Using this approach, we found that in individual

orientation-selective neurons, controlling for the effect of motion on audiovisual trials minimally changed the accuracy of the population decoder across contrast levels (Figure 2.7B-C; p(vis)=7.7e-93, p(aud)=0.055, p(interact)=0.058, paired 2-way ANOVA, Table 1). However, regressing out both sound and motion from the audiovisual responses resulted in decoding accuracy that resembled that on visual trials (Figure 2.7B-C; p(vis)=8.1e-95, p(aud) = 0.55, p(interact)=0.24, paired 2-way ANOVA, Table 1). These results in individual neurons indicate that sound and not movement primarily drives the improvements in decoding accuracy in audiovisual trials. We found similar results when implementing this approach in the MLE-based population decoder. We again found that that decoding performance on audiovisual trials when regressing out motion was still significantly improved compared to that on visual trials (Figure 2.7D-E; p(vis)=1.4e-38, p(aud)=6.0e-8, p(interact)=0.0015, 2-way ANOVA; $p_{c=0}$=0.30, $p_{c=0.25}$=0.0012, $p_{c=0.5}$=0.0022, $p_{c=0.75}$=0.0044, $p_{c=1}$=0.35, Bonferroni-corrected paired t-test). Furthermore, regression of both sound and movement from audiovisual trials resulted in population decoding performance similar to that on visual trials (Figure 2.7D-E; p(vis)=2.5e-39, p(aud)=0.48, p(interact)=0.99, 2-way ANOVA). These results demonstrate that at both an individual neuron and population level, sound improves visual stimulus decoding on audiovisual trials even when controlling for sound-induced motion.

DISCUSSION

Audiovisual integration is an essential aspect of sensory processing (Stein et al., 2020). In humans, audiovisual integration is used in everyday behaviors such as speech perception and object recognition (Fujisaki et al., 2014). In animal models, audiovisual
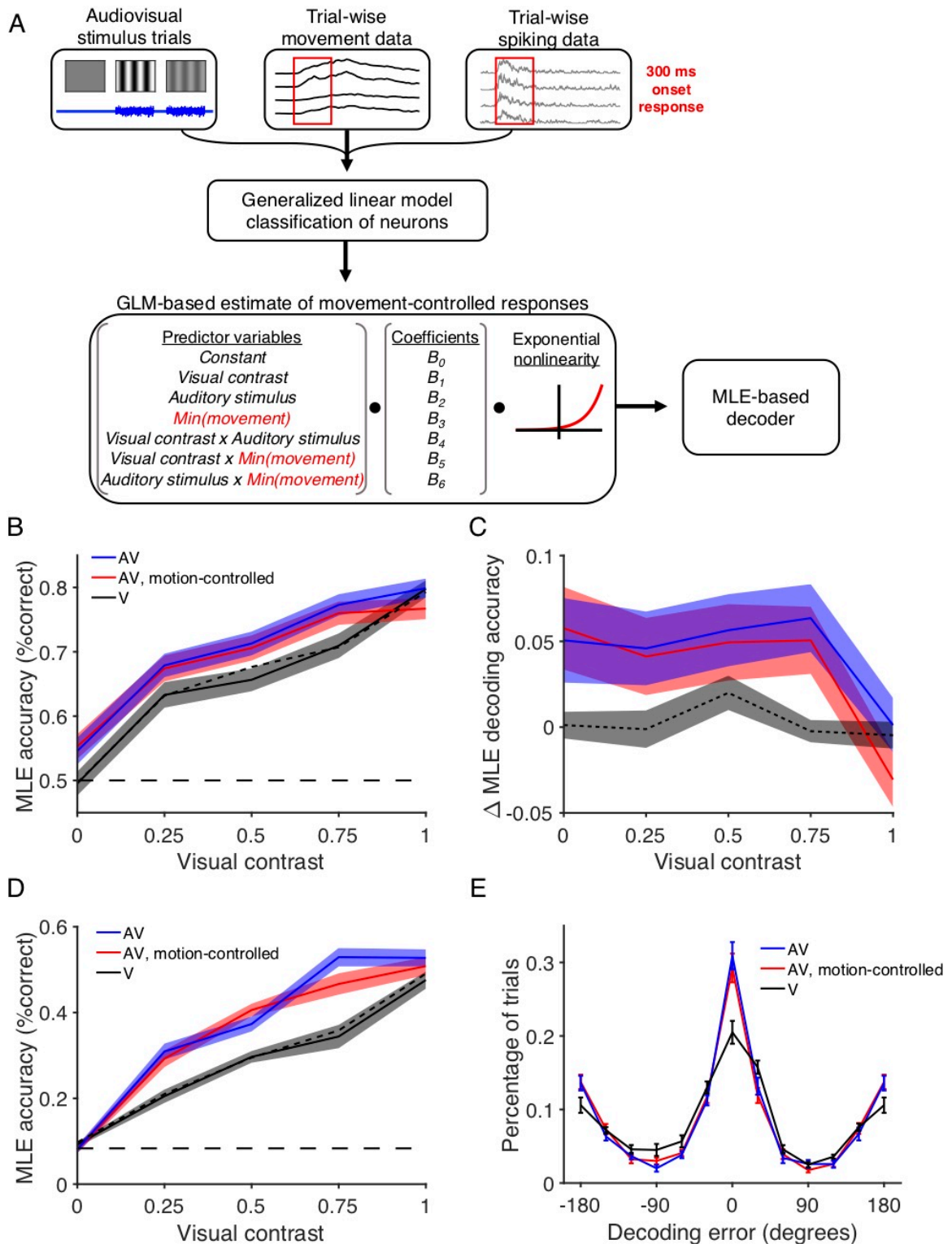
Figure 2.7 | Sound improved decoding performance when controlling for motion. (A) Diagram illustrating the use of a GLM to calculate each predictor variable's coefficient. These are then used when varying the predictor variables to estimate trial-wise neuronal responses, which are then into the MLE-based

decoder. (B) Absolute accuracy of decoding orientation among orientation-selective, sound/motion-modulated light-responsive neurons, comparing visual responses (black, solid) to audiovisual responses (blue) and audiovisual responses when regressing out motion (red). The finely dotted line represents audiovisual responses when controlling for the effects of both motion and sound. (C) Relative decoding accuracy compared to decoding on visual trials. Regressing out motion did not reduce performance compared to audiovisual trials (n=85 neurons, p(vis)=7.7e-93, p(aud)=0.055, p(interact)=0.058, paired 2-way ANOVA), whereas regressing out both motion and sound resulted in comparable performance to visual trials (n=85 neurons, p(vis)=8.1e-95, p(aud)=0.55, p(interact)=0.24, paired 2-way ANOVA). (D) Population decoding accuracy of population-based decoder on audiovisual trials (blue) is preserved even when controlling for motion (red) compared to decoding of visual trials (black; n=10 randomizations, p(vis) = 1.4e-38, p(aud)=6.0e-8, p(interact)=0.0015, 2-way ANOVA; $p_{c=0}$=0.30, $p_{c=0.25}$0.0012, $p_{c=0.5}$=0.0022, $p_{c=0.75}$=0.0044, $p_{c=1}$=0.35, Bonferroni-corrected paired t-test). The finely black dotted line represents decoding accuracy when regressing out both sound and motion. (E) MLE decoder classification percentage, comparing estimated direction to actual direction, contrast 0.25. Little difference is observed between audiovisual trials and audiovisual trials when controlling for motion, whereas both are more accurate than visual trials.

integration improves the detection and discriminability of unisensory auditory and visual stimuli (Gleiss and Kayser, 2012; Meijer et al., 2018). However, the neuronal mechanisms underlying these behavioral improvements are still being revealed. Specifically, it remains unclear how sound-induced changes in neuronal activity affect encoding of the visual stimulus. Furthermore, whether the reported audiovisual integration can more accurately be attributed to sound-induced movement has yet to be studied.

The goal of the present study was to test the hypothesis that sound improves neuronal encoding of visual stimuli in V1 independent of sound-induced movement. We performed extracellular recordings in V1 while presenting combinations of visual drifting gratings and auditory white noise and recording movement of awake mice. The drifting gratings were presented at a range of visual contrast levels to determine the threshold levels at which sound is most effective. As in previous studies, we found neurons in V1 whose spontaneous and visually evoked firing rates are modulated by sound (Figure 2.2). Notably, the effects we observed were stronger and more positive than in previous studies (80.1% of neurons were modulated by sound, with ~95% exhibiting sound-induced increases in firing rate). When accounting for movement in awake animal subjects, we found that the neurons' audiovisual responses actually represented a mixed effect of both sound- and movement-sensitivity (Figure 2.3), an effect in which sound primarily enhances the onset response whereas movement complementarily enhances the sustained response (Figure 2.4). We also found that sound-induced changes in response magnitude and consistency combined to improve the discriminability of drifting grating orientation and direction in individual neurons and at a population level (Figure 2.5,2.6). The improvements in neuronal encoding were most pronounced at low to intermediate visual contrast levels, a

finding that supports the current understanding that audiovisual integration is most beneficial for behavioral performance under ambiguous unisensory conditions (Gleiss and Kayser, 2012; Meijer et al., 2018; Stein et al., 2020). Importantly, the improvement in neuronal encoding was based on firing at the onset of the visual response, indicating that the auditory signal itself is responsible for improvements in visual encoding and not attributable to uninstructed movements. This was directly demonstrated by the persistence of sound-induced improvements in stimulus decoding, even when controlling for the effect of motion (Figure 2.7).

Auditory and locomotive inputs distinctly shape visual responses

We present the novel finding that sound and movement have distinct and complementary effects on visual response. Specifically, we found that sound primarily enhances the firing rate at the onset of the visual response, whereas motion enhances the firing rate during the sustained period of the visual response (Figure 2.4F-H). Our initial classification of sound-modulated neurons and the subsequent decoding analyses were based on firing rates during the onset period. Therefore, despite robust differences in movement during visual and audiovisual trials, motion was unable to account for the sound-induced changes in neuronal responses that resulted in improved neuronal encoding (Figure 2.7). The distinct effects that sound and locomotion have on visual responses also adds nuance to our understanding of how motion affects visual processing, as other groups have predominantly used responses averaged across the duration of the stimulus presentation in categorizing motion responsive neurons in V1 (Neil and Stryker, 2010; Dardalat and Stryker, 2017). Our findings indicate that the timing of cross-sensory

interactions is an important factor in the classification and quantification of multisensory effects.

We also observed that motion decreases the magnitude of the enhancing effect that sound has on the onset of the visual response (Figure 2.3E, 2.4H). This finding suggests a degree of suppressive effect that motion has on this audiovisual interaction. A potential mechanism for this result may relate to the circuits underlying audiovisual integration in V1. Other groups have shown using retrograde tracing, optogenetics and pharmacology that the AC projects directly to V1 and is responsible for the auditory signal in this region (Falchier et al., 2002; Ibrahim et al., 2016; Deneux et al., 2019). It is currently understood that unlike in V1, in other primary sensory cortical areas including the AC movement suppresses sensory evoked activity (Nelson et al., 2013; Schneider and Mooney, 2018; Bigelow et al., 2019). Therefore, one explanation for this observation is that despite motion enhancing the visual response magnitude in the absence of sound, the suppressive effect that motion has on sound-evoked responses in the AC leads to weaker AC enhancement of visual activity on trials in which the mice move. A detailed experimental approach using optogenetics or pharmacology would be required to test this hypothesis of a tripartite interaction and would also reveal the potential contribution of other auditory regions.

Enhanced response magnitude and consistency combine to improve neuronal encoding

Signal detection theory indicates that improved encoding can be mediated both by enhanced signal magnitude as well as reduced levels of noise (von Trapp et al., 2016). When using purely magnitude-based metrics of discriminability, OSI and DSI, we found a

small reduction from the visual to audiovisual conditions (Fig 2.2 Sup 1). However, we also observed that sound reduced the CV of visual responses (Fig 2.2 Sup 2), a measure of the trial-to-trial variability in response. When we measured the *d'* sensitivity index of neuronal responses, a measure that factors in both the response magnitude and distribution, we found that sound improved the discriminability of drifting grating orientation and direction (Fig 2.4 Sup 1). These findings indicate that the improved discriminability of visual responses in individual neurons was mediated not only by changes in response magnitude but also by the associated improvement in response consistency between trials. Therefore, it is important to consider response variability in addition to magnitude-based metrics when quantifying tuning and discriminability in neurons (Churchland et al., 2011).

Prior studies using calcium imaging found equivocal results when investigating whether sound-induced changes in visual responses led to improved population encoding of the visual stimulus (Meijer et al., 2017). The improved discriminability of grating orientation and direction by individual neurons supports our finding that the presence of sound enhances population encoding of the visual stimulus. One explanation for this difference may be the recording modality and analysis parameters. We performed electrophysiological recordings of spiking activity and limited our quantification to the onset of the stimulus (0-300 ms), the time window in which there was the greatest change in firing rate across neurons. Calcium imaging, on the other hand, may lack the temporal resolution required to detect the trial-by-trial differences in spiking activity associated with improved neuronal discriminability. Additionally, extracellular electrophysiology allowed us to take advantage of large numbers of neurons in awake animals to include in the population analysis, as opposed to patch-clamp approaches with a limited number of

neurons (Ibrahim et al., 2016). Finally, presenting a wide range of visual contrast levels allowed use to demonstrate that sound improves neuronal encoding at low to intermediate contrasts, above which further improvement is difficult to demonstrate due to already reliable encoding in the absence of sound.

Stimulus parameters relevant to audiovisual integration

Sensory neurons are often tuned to specific features of unisensory auditory and visual stimuli, and these features are relevant to cross-sensory integration of the signals. In the current study we paired the visual drifting gratings with a static burst of auditory white noise as a basic well-controlled stimulus. Previous studies found that temporally congruent audiovisual stimuli, e.g. amplitude-modulated sounds accompanying visual drifting gratings, evoke larger changes in response than temporally incongruent stimuli in the mouse visual cortex (Meijer et al., 2017), and therefore using such stimuli would potentially result in even stronger effects than we observed. Auditory pure tones can also induce changes in V1 visual responses (McClure and Polack, 2019). However, in other brain regions such as the inferior colliculus, audiovisual integration is highly dependent on spatial congruency between the unimodal inputs (Bergan and Knudsen, 2009). Our results show that spatially congruent, static white noise is sufficient to improve the neuronal response magnitude and latency to light-evoked response. However, additional studies are needed to explore the full range of auditory stimulus parameters relevant to visual responses in V1. Additionally, visual drifting gratings are often used to evoke robust responses in V1, but it would be valuable to determine whether sound is also capable of modulating responses to looming stimuli and more complex visual patterns as well.

Neuronal correlates of multisensory behavior

Our findings of multisensory improvements in neuronal performance are supported by numerous published behavioral studies in humans and various model organisms (Gleiss and Kayser, 2012; Meijer et al., 2018; Stein et al., 2020). Training mice to detect or discriminate audiovisual stimuli allows the generation of psychometric performance curves in the presence and absence of sound. We would hypothesize that the intermediate visual contrast levels in which we see improvements in neural encoding would align with behavioral detection threshold levels. One could also correlate the trial-by-trial neural decoding of the visual stimulus with the behavioral response on a stimulus discriminability task, an analysis that could provide information about the proximity of the V1 responses to the behavioral perception and decision. Additionally, a behavioral task could allow the comparison of neural responses between passive and active observing, helping to reveal the role of attention on how informative or distracting one stimulus is about the other.

Multisensory integration in other systems

It is useful to contextualize audiovisual integration by considering multisensory integration that occurs in other primary sensory cortical areas. The auditory cortex contains visually responsive neurons and is capable of binding temporally congruent auditory and visual stimulus features in order to improve deviance detection within the auditory stimulus (Atilgan et al., 2018; Morrill and Hasenstaub, 2018). Additionally, in female mice, pup odors reshape AC neuronal responses to various auditory stimuli and drive pup retrieval behavior (Cohen et al., 2011; Marlin et al., 2015), demonstrating integration of auditory and olfactory signals. However, whether these forms of multisensory integration rest on

similar coding principles of improved SNR observed in the current V1 study is unknown. Investigation into this relationship between the sensory cortical areas will help clarify the neuronal codes that support multisensory integration, and the similarities and differences across sensory domains.

METHODS

Mice

All experimental procedures were in accordance with NIH guidelines and approved by the IACUC at the University of Pennsylvania. Mice were acquired from Jackson Laboratories (5 male, 6 female, aged 10-18 weeks at time of recording; B6.Cast-$Cdh23^{Ahl+}$ mice [Stock No: 018399]) and were housed at 28°C in a room with a reversed light cycle and food provided ad libitum. Experiments were carried out during the dark period. Mice were housed individually after headplate implantation. Euthanasia was performed using CO_2, consistent with the recommendations of the American Veterinary Medical Association (AVMA) Guidelines on Euthanasia. All procedures were approved by the University of Pennsylvania IACUC and followed the AALAC Guide on Animal Research. We made every attempt to minimize the number of animals used and to reduce pain or discomfort.

Surgical procedures

Mice were implanted with skull-attached headplates to allow head stabilization during recording, and skull-penetrating ground pins for electrical grounding during recording. The mice were anesthetized with 2.5% isoflurane. A ~1mm craniotomy was

performed over the right frontal cortex, where we inserted a ground pin. A custom-made stainless steel headplate (eMachine Shop) was then placed on the skull at midline, and both the ground pin and headplate were fixed in place using C&B Metabond dental cement (Parkell). Mice were allowed to recover for 3 days post-surgery before any additional procedures took place.

Electrophysiological recordings

All recordings were carried out inside a custom-built acoustic isolation booth. 1-2 weeks following the headplate and ground pin attachment surgery, we habituated the mice to the recording booth for increasing durations (5, 15, 30 minutes) over the course of 3 days. On the day of recording, mice were placed in the recording booth and anesthetized with 2.5% isoflurane. We then performed a small craniotomy above the left primary visual cortex (V1, 2.5mm lateral of midline, 0-0.5 mm posterior of the lambdoid suture). Mice were then allowed adequate time to recover from anesthesia. Activity of neurons were recorded using a 32-channel silicon probe (NeuroNexus A1x32-Poly2-5mm-50s-177). The electrode was lowered into the primary visual cortex via a stereotactic instrument to a depth of 775-1000μm. Following the audiovisual stimulus presentation, electrophysiological data from all 32 channels were filtered between 600 and 6000 Hz, and spikes belonging to single neurons and multi-units were identified in a semi-automated manner using KiloSort2 (Pachitariu et al., 2016).

Audiovisual stimuli

The audiovisual stimuli were generated using MATLAB (MathWorks, USA), and presented to mice on a 12" LCD monitor (Eyoyo) and through a magnetic speaker (Tucker-Davis Technologies) placed to the right of the mouse. The visual stimulus was generated using the PsychToolBox package for MATLAB and consisted of square wave drifting gratings 1 s in duration, 4-Hz temporal frequency, and 0.1 cycles/°. The gratings moved in 12 directions, evenly spaced 0°-360°, and were scaled to a range of 5 different visual contrast levels (0, 0.25, 0.5, 0.75, 1), totaling 60 unique visual stimuli. The auditory stimulus was sampled at 400 kHz and consisted of a 1 s burst of 70 dB white noise. The visual grating was accompanied by the auditory noise on half of trials (120 unique trial types, 10 repeats each), with simultaneous onset and offset. The auditory-only condition corresponded to the trials with a visual contrast of 0. The trial order was randomized and was different for each recording.

Data analysis and statistical procedures

Spiking data from each recorded unit was organized by trial type and aligned to the trial onset. The number of spikes during each trial's first 0-300ms was input into a generalized linear model (GLM; predictor variables: visual contrast [continuous variable 0, 0.25, 0.5, 0.75 ,1], sound [0 or 1]; response variable: number of spikes during 0-300ms; Poisson distribution, log link function), allowing the classification of each neuron's responses as having a main effect ($p < 0.05$) of light, sound, and/or a light-sound interaction. Neurons that were responsive to both light and sound or had a significant light-sound interaction term were classified as "light-responsive sound-modulated." To quantify the

supra- or sub-linear integration of the auditory and visual responses, we calculated the linearity ratio of neurons' audiovisual responses. This ratio was defined as $FR_{AV} / (FR_V + FR_A)$, and the sound-only response FR_A was calculated using the trials with a visual contrast of 0.

We quantified changes in response timing by calculating response latency, onset slope, and onset response duration. First, mean peristimulus time histograms (PSTH) were constructed for each trial type using a 10 ms sliding window. The latency was calculated as the first time bin after stimulus onset in which the mean firing rate at full contrast exceeded 1 standard deviation above baseline. The slope Hz/ms slope was calculated from the trial onset to the time of the peak absolute value firing rate. The response duration was calculated using the full width at half maximum of the peak firing rate at stimulus onset (limited to 0-300 ms).

Orientation selectivity and direction selectivity were determined for all light-responsive neurons. The preferred direction of each direction-selective neuron was defined as the drifting grating direction that evoked the largest mean firing rate at the highest contrast level (FR_{pref}). We calculated orientation and direction-selective indices (Zhao et al., 2013) for each neuron according to:

$$\text{OSI} = \frac{FR_{pref} - FR_{ortho}}{FR_{pref} + FR_{ortho}} \qquad \text{DSI} = \frac{FR_{pref} - FR_{antipref}}{FR_{pref} + FR_{antipref}}$$

where FR_{ortho} and $FR_{antipref}$ are the mean firing rates in the orthogonal (90°) and anti-preferred (180°) directions, respectively. One-tailed permutation testing was performed by comparing these OSI and DSI values to pseudo OSI and DSI values obtained by 200 random shuffles of the firing rates from the pooled preferred and orthogonal or anti-

preferred trials. If a neuron's actual OSI or DSI value was >95% of shuffled OSI or DSI values, the neuron was classified as "orientation-" or "direction-selective," respectively. To determine whether there were statistically significant changes in the preferred direction from the visual to audiovisual conditions, we applied a bootstrapping procedure, subsampling the visual trials for each neuron 1000 times and creating a confidence interval of the mean shift in preferred direction (degrees) for each population randomization.

We assessed and controlled for sound-induced movement as a potential confound for the audiovisual effects observed. During a subset of V1 recordings (9 recordings, 5 mice), mouse movement was tracked throughout stimulus presentation. Video recording was performed using a Raspberry Pi 4 Model B computer system with an 8MP infrared Raspberry Pi NoIR Camera V2 attachment. The video was converted to MP4 format, and motion was quantified by calculating the frame-by-frame difference, an approach that captured both whisking and locomotive behavior. This movement value for each recording was then aligned to the trials of the audiovisual stimulus from the recording trials for further analysis.

Similar to above, a GLM (predictor variables: visual contrast level, sound presence, average motion during each trial; response variable: trial spikes during 0-300ms; Poisson distribution, log link function) classified each neuron as having a main effect ($p<0.05$) of light, sound, or motion, as well as the pairwise interactions of these parameters. Light-responsive sound-modulated neurons, according to the above definition, that additionally displayed either a main effect of motion or significant light-motion or sound-motion interaction terms were classified as "motion-modulated" and were included for further analysis.

To reconstruct peristimulus time histograms of light-responsive, sound-modulated, motion-modulated neurons, we used a separate GLM. Using a 10ms sliding window across all trials, we input the visual contrast level, sound presence, and motion during that window (discretized into five bins) as predictor variables, and the number of spikes during that window as response variables, into the GLM (Poisson distribution, log link function) to calculate coefficients for light, sound, motion, and their pairwise interactions. This approach allowed us to reconstruct the mean PSTH of individual neurons observed during each trial type by calculating:

$$\text{Spikes}_t = \exp\left(\sum_i p_{t,i} \cdot c_{t,i}\right)$$

where the spikes in time window t are determined by the values p and coefficients c of predictor variable i. From there, we used this same equation to estimate the shape of the PSTHs when varying sound and motion in order to determine differential effects these parameters had on the temporal trajectory of neurons' visual responses.

The d' sensitivity index (Stanislaw and Todorov, 1999; von Trapp et al., 2016) was used to calculate the directional discriminability of direction-selective neurons. The d' sensitivity index between two directions θ_1 and θ_2 is calculated as:

$$d' = \frac{\mu_{\theta_1} - \mu_{\theta_2}}{\sqrt{\frac{1}{2}(\sigma_{\theta_1}^2 + \sigma_{\theta_2}^2)}}$$

where μ_θ and σ_θ are the response mean and standard deviation, respectively, for direction θ. For each neuron, the sensitivity index was calculated in a pairwise manner for preferred direction versus all other directions and then aligned relative to the preferred direction to test sensitivity index as a function of angular distance from preferred direction.

We used a maximum likelihood estimate approach (Montijn et al., 2014; Meijer et al., 2017) to decode the visual stimulus direction from the neuronal responses based on Bayes rule:

$$P(\theta|A_{trial}) = \frac{P(A_{trial}|\theta)P(\theta)}{P(A_{trial})}$$

For decoding using individual neurons, the likelihood $P(A_{trial}|\theta)$ for each orientation or direction was computed based on the Poisson response distribution across all trials of that orientation or direction, with a leave-one-out cross-validation technique in which the probe trial (A_{trial}) was excluded from the training data. The prior $P(\theta)$ was uniform, and the normalization term $P(A_{trial})$ was similarly applied to all directions. Therefore, the posterior probability $P(\theta|A_{trial})$ was proportional to and based on evaluating the likelihood function at the value of the probe trial. For orientation-selective neurons, decoding was performed between the preferred and orthogonal orientations, and for direction-selective neurons, decoding was performed between the preferred and anti-preferred directions. For decoding using populations of neurons, neurons were pooled across recording sessions. A similar approach was used; however, here, the posterior probability $P(\theta|A_{pop})$ was proportional to the joint likelihood $P(A_{pop}|\theta)$ of the single-trial activity across all N neurons in the population (A_{pop}):

$$P(A_{pop}|\theta) = \prod_{neuron\ i}^{N} P(A_{trial}|\theta)_i$$

With this population-based analysis, pairwise decoding was performed between every orientation and its orthogonal orientation (1 of 2 options), as well as decoding one direction from all possible directions (1 of 12 options).

Additionally, we used a support vector machine (SVM) to corroborate the findings of the MLE-based decoder. The SVM was implemented using MATLAB's fitcsvm function with a linear kernel to predict the drifting grating direction based on single-trial population responses. Similarly, a leave-one-out cross-validation technique was used, and pairwise decoding was performed between every combination of two stimulus directions.

Statistics

Figure data are displayed as means with standard error of the mean (SEM), unless otherwise noted. Shapiro-Wilk tests were used to assess normality, and the statistical tests performed are indicated in the text, figures, and Table 1. For multi-group and multivariate analysis (e.g., ANOVA and Kruskal-Wallis tests) in which a significant ($p<0.05$) interaction was detected, we subsequently performed a post hoc Bonferroni-corrected test. P-values reported as 0 are too small to be accurately calculated by Matlab ($p<2.2e-301$), due to characteristically large data sets. See Table 1 for a detailed summary of statistical results and post hoc comparisons.

SUPPLEMENTARY FIGURES
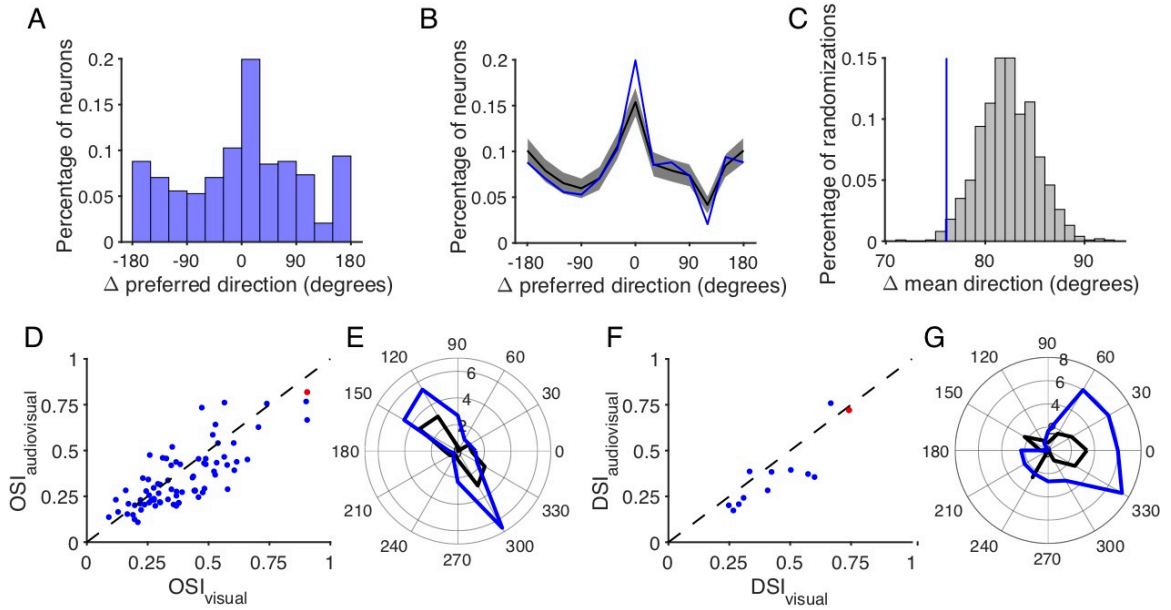
Figure 2.2 Supplementary 1 | Sound minimally reduces tuning selectivity in individual neurons (A) Histogram depiction of changes in preferred drifting grating directions with sound in orientation-selective neuron. (B) Observed changes in preferred direction (blue) compared to shuffled permutations (black) using the mean and standard deviation of observed responses. (C) The observed mean change in preferred direction (blue) is within the expected distribution (gray) based on visual response variability. (D,E) A slight reduction in the orientation selectivity index was observed in orientation-selective neurons (n=78, p=0.0018, paired t-test). The visual tuning of the red data point in *D* is displayed in *E*. (F,G) A slight reduction in the direction selectivity index was also observed in direction-selective neurons (n=12, p=0.021, paired t-test), with the tuning of the red data point in *F* displayed in *G*.
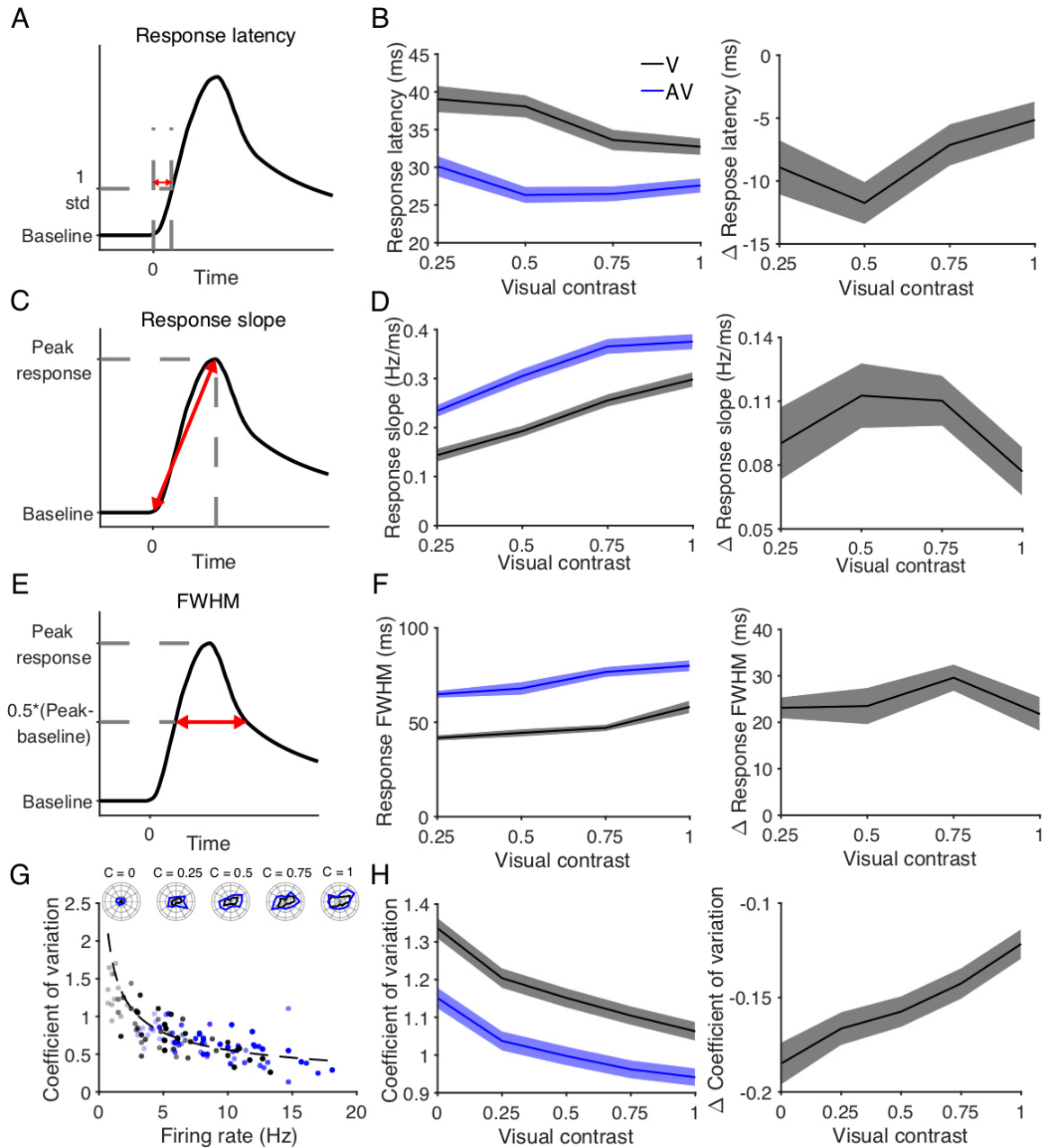
Figure 2.2 Supplementary 2 | Sound reduces the latency, increases duration, and reduces variability of light-evoked responses in individual neurons (A) Diagram of the calculation of response latency, the first time bin in which the FR exceeds 1 std above baseline. (B) Response latency is reduced by sound (left: absolute, right: difference; p(vis)=6.9e-4, p(aud)=6.8e-15, p(interact)=0.045, paired 2-way ANOVA, post hoc Bonferroni-corrected paired t-test, Table 1). (C) Diagram of the calculation of response onset slope, the peak change in FR over the latency to peak response. (D) Sound increases the slope of the onset response (left: absolute, right: difference; n=563, p(vis)=3.5e-121, p(aud)=2.7e-15, p(interact)=0.038, paired 2-way ANOVA, post hoc Bonferroni-corrected paired t-test). (E) Diagram of the calculation of FWHM, the width of the onset response at half maximum FR. (F) Sound increases the FWHM duration of the onset response

60

(left: absolute, right: difference; n=367, p(vis)=1.3e-10, p(aud)=8.7e-98, p(interact)=0.23 paired 2-way ANOVA). (G) An example neuron demonstrating that increased response magnitude corresponds to lower CV according to an inverse square root relationship. The black and blue dots represent visual and audiovisual responses, respectively, and the dot transparency corresponds to visual contrast level. The dotted lines are fitted y=c/sqrt(x) curves, where c is a constant. The above inset is the polar plots corresponding to the example neuron. (H) Lower coefficient of variation indicates reduced response variability in audiovisual compared to visual responses (left: absolute, right: difference; n=563, p(vis)=0.28, p(aud)=4.2e-103, p(interact)=0.38, paired 2-way ANOVA).
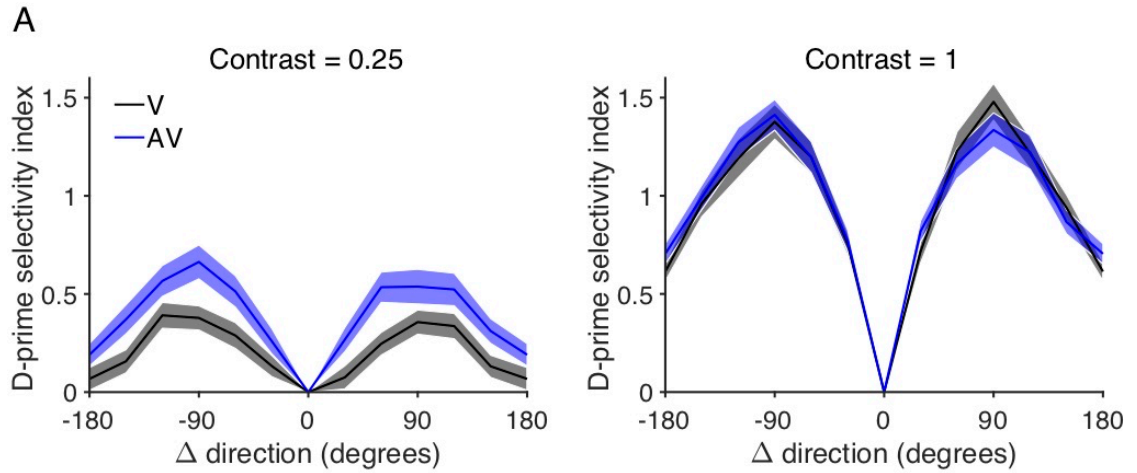
Figure 2.5 Supplementary 1 | Sound enhances the *d'* sensitivity index at low contrast levels (A) The *d'* sensitivity index between neuronal responses to drifting grating directions, averaged across orientation- and direction-selective neurons. Enhancements are observed at low visual contrast (left), whereas minimal changes are present at full contrast (right).
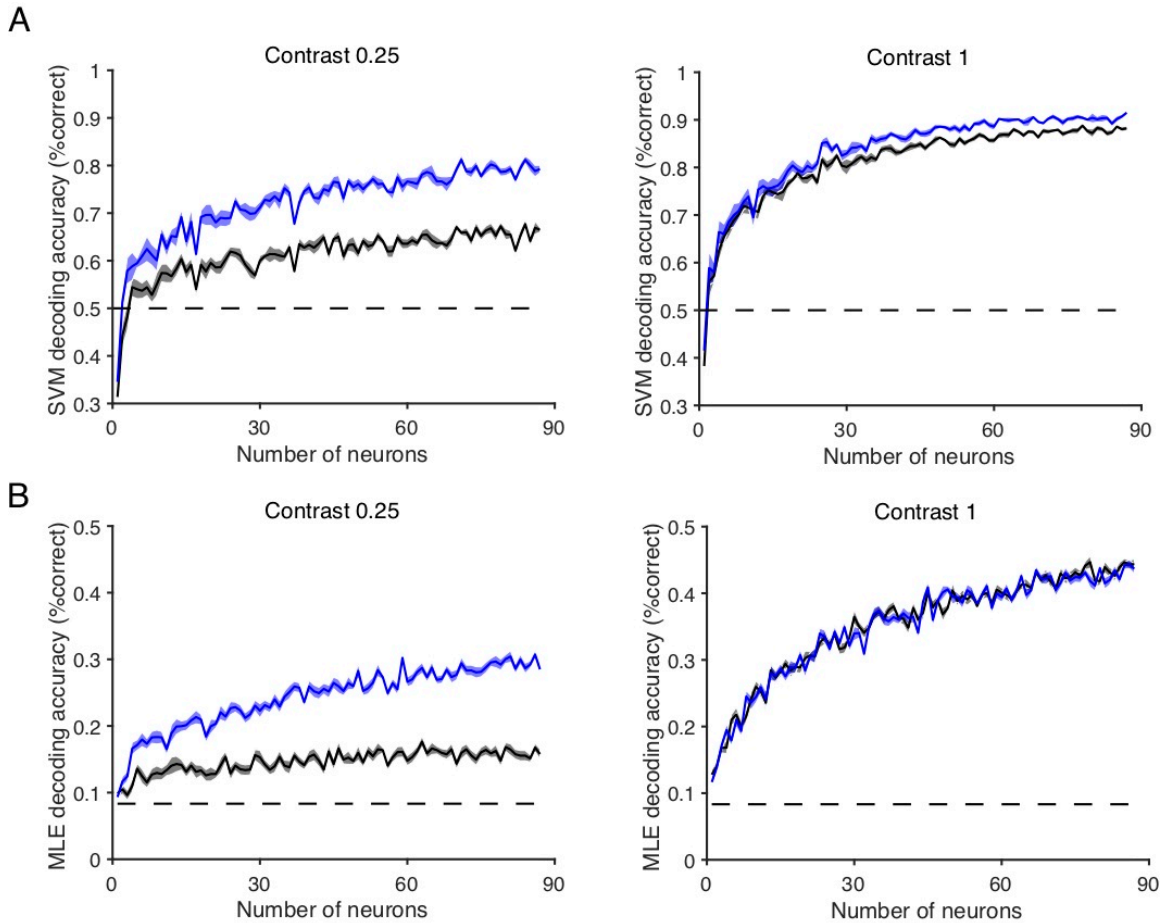
Figure 2.6 Supplementary 1 | Decoding accuracy increases with population size (A) Accuracy of SVM pairwise classification, average across all direction pairs, as the neuronal population size included in the decoder increases. Visual contrast 0.25 is on the left, and full visual contrast is on the right. (B) Accuracy of MLE decoding 1 of 12 drifting grating options, as the neuronal population size increases. Again, visual contrast 0.25 is on the left, and full visual contrast is on the right.

Table 1: Statistical comparisons

Comparison	Fig	Test	Test statistic	N	df	p-value	Post hoc test	Post hoc α	Post hoc comparison	Post hoc p-value
Mean firing rate, V vs AV	2.2C	Paired 2-way ANOVA	F(vis)=340 F(aud)=506 F(interact)=75	565 neurons	vis=4 aud=1 interact = 4	p(vis) = 1.2e-100 p(aud) = 1.6e-88 p(interact) = 5.7e-4	Bonferroni-corrected paired t-test	0.01	Contrast 0, V vs AV	2.1e-50
									Contrast 0.25, V vs AV	2.6e-62
									Contrast 0.5, V vs AV	5.7e-75
									Contrast 0.75, V vs AV	1.1e-81
									Contrast 1, V vs AV	2.0e-81
Linearity ratio, V vs AV	2.2E	Kruskal-Wallis test	Chi-sq = 61	555 neurons	4	p = 1.6e-12	Bonferroni-corrected Wilcoxon signed rank test	0.013	Contrast 0 vs 0.25	0.053
									Contrast 0 vs 0.5	0.0040
									Contrast 0 vs 0.75	4.6e-8
									Contrast 0 vs 1	2.1e-5
Sound induced movement	2.3A	Paired t-test	t-stat = -7.2	9 recording sessions	8	p = 9.1e-5				
Firing rate across movement range, V vs AV	2.3E	Unbal-anced 2-way ANOVA	F(motion)=6.9 F(sound)=55 F(interact)=18	Variable trial count	mot=2 aud=1 Interact=2	p(motion) = 0.001 p(sound) = 1.4e-13 p(interact) = 1.8e-8	Bonferroni corrected two-sample t-test	0.016	Stationary, V vs AV	1.5e-14
									Low motion, V vs AV	7.1e-10
									High motion, V vs AV	0.60
PSTH, light vs light/sound	2.4F	Paired t-test	1391 unique t-stats	295 neurons	294	1391 unique p-values, α= 0.05/1391= 3.6e-5				
PSTH, light vs light/motion	2.4G	Paired t-test	1391 unique t-stats	295 neurons	294	1391 unique p-values, α= 0.05/1391= 3.6e-5				
PSTH, light/sound vs light/sound/motion	2.4H	Paired t-test	1391 unique t-stats	295 neurons	294	1391 unique p-values, α= 0.05/1391= 3.6e-5				
Orientation selectivity index, V vs AV	Fig 2.2 Sup1D	Paired t-test	t-stat = 3.2	78 neurons	77	p = 0.0018				
Direction selectivity index, V vs AV	Fig 2.2 Sup 1F	Paired t-test	t-stat = 2.7	12 neurons	11	p = 0.0206				
Onset response latency, V vs AV	Fig 2.2 Sup 2B	Paired 2-way ANOVA	F(vis)=5.7 F(aud)=64 F(interact)=2.7	517 neurons	vis=3 aud=1 interact=3	p(vis)=6.9e-4 p(aud)=6.8e-18 p(interact)=0.045	Bonferroni-corrected paired t-test	0.01	Contrast 0.25, V vs AV	2.3e-4
									Contrast 0.5, V vs AV	7.1e-12
									Contrast 0.75, V vs AV	4.6e-5
									Contrast 1, V vs AV	9.9e-4

Table 1 (cont.)

Comparison	Fig	Test	Test statistic	N	df	p-value	Post hoc test	Post hoc α	Post hoc comparison	Post hoc p-value
Onset response slope, V vs AV	Fig 2.2 Sup 2D	Paired 2-way ANOVA	F(vis)=70 F(aud)=66 F(interact)=2.8	563 neurons	vis=3 aud=1 interact=3	p(vis)=3.5e-121 p(aud) = 2.7e-15 p(interact) = 0.038	Bonferroni-corrected paired t-test	0.01	Contrast 0.25, V vs AV	1.4e-4
									Contrast 0.5, V vs AV	8.9e-13
									Contrast 0.75, V vs AV	3.6e-12
									Contrast 1, V vs AV	5.5e-8
Onset response duration, V vs AV	Fig 2.2 Sup2F	Paired 2-way ANOVA	F(vis)=17 F(aud)=129 F(interact)=1.4	367 neurons	vis=3 aud=1 Interact=3	p(vis)=1.3e-10 p(aud) = 8.7e-98 p(interact) = 0.23				
Response coefficient of variation, V vs AV	Fig 2.2 Sup 2H	Paired 2-way ANOVA	F(vis)=1.3 F(aud)=834 F(interact)=1.0	564 neurons	vis=4 aud=1 Interact=4	p(vis) = 0.28 p(aud) = 4.2e-103 p(interact) = 0.38				
Orientation decoding accuracy, individual neurons, V vs AV	Fig 2.5C	Paired 2-way ANOVA	F(vis)=67 F(aud)=12 F(interact)=0.54	78 neurons	vis=4 aud=1 interact=4	p(vis)=4.8e-112 p(aud)=7.8e-4 p(interact) = 0.71				
Direction decoding accuracy, individual neurons, V vs AV	2.5E	Paired 2-way ANOVA	F(vis)=6.9 F(aud)=2.0 F(interact)=0.43	12 neurons	vis=4 aud=1 interact=4	p(vis)=2.1e-4 p(aud)=0.18 p(interact)=0.78				
Orientation decoding accuracy, SVM, population, V vs AV	2.6C	2-way ANOVA	F(vis)=526 F(aud)=38 F(interact)=6	10 repeats	vis=4 aud=1 interact=4	p(vis) = 1.8e-61 p(aud) = 1.9e-8 p(interact) = 2.4e-4	Bonferroni-corrected paired t-test	0.01	Contrast 0, V vs AV	0.12
									Contrast 0.25, V vs AV	0.0016
									Contrast 0.5, V vs AV	0.0014
									Contrast 0.75, V vs AV	0.0023
									Contrast 1, V vs AV	1
Direction decoding accuracy, SVM, population, V vs AV	2.6D	2-way ANOVA	F(vis)=48 F(aud)=40 F(interact)=4.6	10 repeats	vis=4 aud=1 interact=4	p(vis) = 1.1e-21 p(aud) = 9.0e-9 p(interact) = 0.0019	Bonferroni-corrected paired t-test	0.01	Contrast 0, V vs AV	0.55
									Contrast 0.25, V vs AV	5.3e-5
									Contrast 0.5, V vs AV	0.0036
									Contrast 0.75, V vs AV	0.17
									Contrast 1, V vs AV	0.0036
Orientation decoding accuracy, MLE, population, V vs AV	2.6F	2-way ANOVA	F(vis)=682 F(aud)=0.27 F(interact)=18	10 repeats	vis=4 aud=1 interact=4	p(vis)=2.3e-66 p(aud)=0.61 p(interact)=9.6e-11	Bonferroni-corrected paired t-test	0.01	Contrast 0, V vs AV	5.8e-4
									Contrast 0.25, V vs AV	1.8e-4
									Contrast 0.5, V vs AV	0.30
									Contrast 0.75, V vs AV	0.53
									Contrast 1, V vs AV	0.15
Direction decoding accuracy, MLE, population, V vs AV	2.6G	2-way ANOVA	F(vis)=67 F(aud)=0.43 F(interact)=8.9	10 repeats	vis=4 aud=1 interact=4	p(vis)=4.6e-26 p(aud)=0.51 p(interact)=4.1e-6	Bonferroni-corrected paired t-test	0.01	Contrast 0, V vs AV	0.037
									Contrast 0.25, V vs AV	6.4e-6
									Contrast 0.5, V vs AV	0.036
									Contrast 0.75, V vs AV	0.16
									Contrast 1, V vs AV	0.014

Table 1 (cont.)

Comparison	Fig	Test	Test statistic	N	df	p-value	Post hoc test	Post hoc α	Post hoc comparison	Post hoc p-value
Overall decoding accuracy, MLE, population, V vs AV	2.6J	2-way ANOVA	F(vis)=411 F(aud)=19 F(interact)=16	20 repeats	vis=4 aud=1 interact=4	p(vis)=2.2e-92 p(aud)=1.9e-5 p(interact)=2.7e-11	Bonferroni - corrected paired t-test	0.01	Contrast 0, V vs AV	0.012
									Contrast 0.25, V vs AV	1.4e-10
									Contrast 0.5, V vs AV	0.48
									Contrast 0.75, V vs AV	0.0013
									Contrast 1, V vs AV	0.50
Orientation decoding accuracy, individual neurons, V vs AV	2.7B	Paired 2-way ANOVA	F(vis) = 74 F(aud) = 19 F(interact) = 1.5	85 neurons	vis=4 aud=1 interact=4	p(vis) =0 p(aud)=3.5e-5 p(interact)=0.21				
Orientation decoding accuracy, individual neurons, V vs motion-corrected AV	2.7B	Paired 2-way ANOVA	F(vis) = 64 F(aud) = 13 F(interact) = 3	85 neurons	vis=4 aud=1 interact=4	p(vis) =0 p(aud)=5.9e-4 p(interact)=0.019	Bonferroni-corrected paired t-test	0.01	Contrast 0, V vs AV	0.019
									Contrast 0.25, V vs AV	0.071
									Contrast 0.5, V vs AV	0.029
									Contrast 0.75, V vs AV	0.011
									Contrast 1, V vs AV	0.0602
Orientation decoding accuracy, individual neurons, AV vs motion-corrected AV	2.7B	Paired 2-way ANOVA	F(vis) = 34 F(aud) = 3.8 F(interact) = 2.4	85 neurons	vis=4 aud=1 interact=4	p(vis) = 7.7e-93 p(aud) = 0.055 p(interact) = 0.058				
Orientation decoding accuracy, individual neurons, V vs motion/sound-corrected AV	2.7B	Paired 2-way ANOVA	F(vis) = 56 F(aud) = 0.36 F(interact) = 1.4	85 neurons	vis=4 aud=1 interact=4	p(vis)=8.1e-95 p(aud)=0.55 p(interact)=0.24				
Population decoding accuracy, V vs AV	2.7D	2-way ANOVA	F(vis) = 166 F(aud) = 52 F(interact) = 8.2	10 repeats	vis=4 aud=1 interact=4	p(vis)=1.1e-40 p(aud)=1.6e-10 p(interact)=1.1e-5	Bonferroni-corrected paired t-test	0.01	Contrast 0, V vs AV	0.34
									Contrast 0.25, V vs AV	2.2e-5
									Contrast 0.5, V vs AV	0.0019
									Contrast 0.75, V vs AV	8.7e-6
									Contrast 1, V vs AV	0.013
Population decoding accuracy, V vs motion-corrected AV	2.7D	2-way ANOVA	F(vis) = 147 F(aud) = 35 F(interact) = 4.8	10 repeats	vis=4 aud=1 interact=4	p(vis)=1.4e-38 p(aud)=6.0e-8 p(interact)=0.0015	Bonferroni-corrected paired t-test	0.01	Contrast 0, V vs AV	0.30
									Contrast 0.25, V vs AV	0.0012
									Contrast 0.5, V vs AV	0.0022
									Contrast 0.75, V vs AV	0.0044
									Contrast 1, V vs AV	0.35
Population decoding accuracy, V vs motion/sound-corrected AV	2.7D	2-way ANOVA	F(vis) = 154 F(aud) = 0.50 F(interact) = 0.088	10 repeats	vis=4 aud=1 interact=4	p(vis)=2.5e-39 p(aud) = 0.48 p(interact) = 0.99				

CHAPTER 3: CORTICAL CIRCUITRY UNDERLYING AUDIOVISUAL INTEGRATION IN V1

ABSTRACT

Humans commonly rely on interactions between sounds and visual signals for behaviors such as spatial navigation and communication. While it is known that sound can affect visual processing on a neuronal level, the neuronal circuits that mediate this process are not well understood. We previously found that sound enhanced encoding of the visual stimulus in the primary visual cortex (V1), in an effect that was more excitatory than previous studies reported. Therefore, detailing the neuronal circuits that support this audiovisual integration may elucidate the unique nature of this interaction. We used retrograde tracing to identify the auditory cortex (AC) as the primary auditory region that projects directly to V1. We observed little colocalization between these infragranular AC projection neurons and the inhibitory neuronal marker glutamate decarboxylase, indicating that they are largely excitatory neurons. Optogenetic stimulation of these AC neurons enhanced baseline and visual-evoked firing rates in V1. However, optogenetic suppression of these same neurons failed to suppress audiovisual integration in V1. Therefore, it is unlikely that the AC neurons are solely responsible for the observed V1 audiovisual integration, and alternative subcortical and disinhibitory are likely involved.

INTRODUCTION

Audiovisual integration relies on the transmission of information between auditory and visual regions, therefore the neuronal pathways and circuits that mediate this communication are of interest to the field of sensory neuroscience. Understanding which brain regions are involved in the integration of this cross-sensory information will clarify the function of these neuronal pathways in supporting multisensory behaviors more broadly. Additionally, detailing the circuits underlying audiovisual integration can improve our understanding of sensory disorders and behavioral deficits in neurological injuries.

Audiovisual integration has been observed and described in the primary visual cortex (V1), and studies have been conducted that have identified the auditory cortex (AC) as an important contributor to this process. A subset of AC neurons project to superficial layers of V1, primarily synapsing with inhibitory neurons, and stimulation of these AC axon terminals sharpened visual orientation tuning in V1 neurons in anesthetized mice (Ibrahim et al., 2016). However, other studies have observed a more excitatory effect of sound on visual responses of individual neurons in awake animals (Meijer et al., 2017). Despite the range of effects, pharmacologic and optogenetic suppression of AC activity may ablate audiovisual integration in V1 (Deneux et al., 2019). Therefore, the AC has a role on transmitting auditory information to be integrated with the visual stream in V1.

Our prior work, detailed in Chapter 2, investigated how sound affects V1 visual processing when taking self-generated movement into account. The study found a largely excitatory effect that sound had on visual responses, which ultimately improved neuronal encoding of the visual stimulus (Williams et al., 2021). We also found distinct effects that

68

sound and motion had on the time course of the visual response. These findings differed from prior studies which were either performed in anesthetized subjects (Ibrahim et al., 2016) or actively excluded trials in which the mice displayed movement (Deneux et al., 2019). Therefore, the role of the AC in modulating V1 visual responses in the context of self-generated movement has remained unexplored. It is unclear whether given this movement, the AC is still responsible for the excitatory signal associated with sound and enhancement of the visual response.

In the current study, we investigate the role of the AC in mediating V1 audiovisual integration. We hypothesized that despite the presence of sound-induced movement, suppressing AC activity would ablate sound's effect on the visual response onset, the portion of the visual response attributable to sound and not movement. We identified infragranular AC neurons that project directly to V1 using retrograde tracing, and demonstrate that stimulation of these excitatory neurons modulates visual responses in V1. However, optogenetic suppression of these AC neurons failed to ablate audiovisual integration in V1, and we therefore propose potential explanations and circuits that would account for these findings.

RESULTS

Auditory cortex projects directly to the primary visual cortex

We investigated the role of the AC in providing auditory information to V1 to mediate audiovisual integration in that region. We began by using viral retrograde tracing to label brain regions that send direct projections to V1. After injecting retroAAV-eGFP into V1 (Figure 3.1A), we sliced and visualized the entire brain to identify the brain regions

that were fluorescently labeled. We observed cell body labeling in the subcortical lateral geniculate nucleus (LGN), validating the use of our retroAAV construct. The only auditory region that consistently contained labeled cell bodies was the primary auditory cortex (A1) and the surrounding belt region (Figure 3.1B). The neurons in these AC regions were predominantly localized in the infragranular layers 5/6, with sparser labeling in supragranular layers (Figure 3.1C). We also consistently observed axonal labeling in various regions including the medial geniculate body (MGB) and inferior colliculus (IC), suggesting that the axons labeled AC neurons co-terminate in both V1 and these subcortical auditory regions. Additional cortical cell body labeling was found in secondary visual regions and sparsely in the frontal cortex and secondary motor regions. In a separate control mouse, we injected AAV5-eGFP into V1 and found no cell body labeling outside of the injection site. This confirms that the fluorescent labeling observed in the experimental animals was due to retrograde activity of the retroAAV, and not diffusion of the virus through the brain tissue. Together, these viral tracing studies confirm that a range of visual and non-visual regions project to V1, of which the AC is the principal auditory component.

AC neurons that project to V1 are primarily excitatory

We next wanted to characterize these labeled AC neurons that project to V1 by determining whether they were excitatory or inhibitory. Understanding this aspect of their activity would help determine how this intercortical circuit modulates firing rates in V1. We approached this question by performing immunohistochemical (IHC) GAD65/67 staining on the retrogradely labeled tissue from above (Figure 3.1). Glutamate decarboxylase (GAD) is an enzyme critical in the formation of the inhibitory

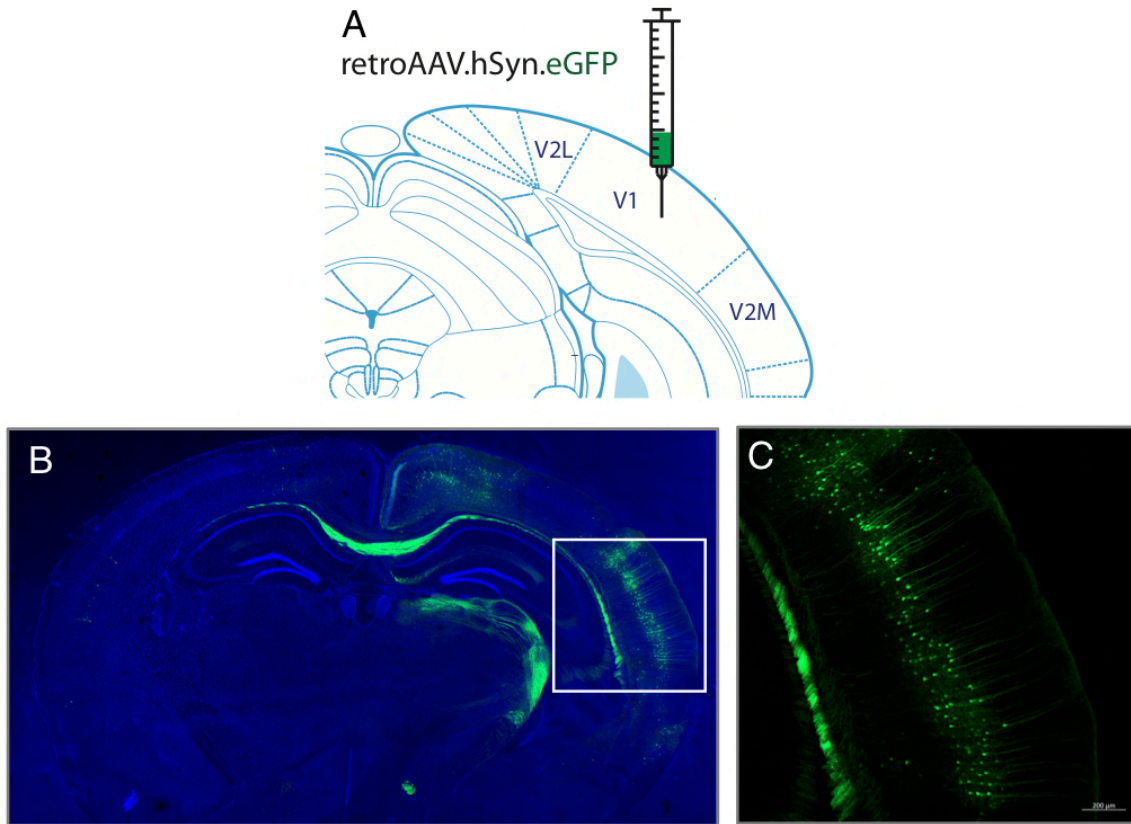**Figure 3.1 | The auditory cortex projects directly to the primary visual cortex** (A) We injected a retroAAV encoding the fluorescent marker GFP into V1 to label brain regions that directly synapse within this region. (B) The AC was the only consistently labeled auditory brain region, indicating direct connections with V1. (C) The labeled AC neurons were primarily located in the infragranular cortical layers.

neurotransmitter GABA, and is therefore a marker of inhibitory neurons. Given that infragranular cortical neurons that project between brain regions are largely excitatory, we hypothesized that IHC-labeled inhibitory neurons and retrogradely labeled AC neurons would represent two distinct neuronal populations. After pairing the GAD65/67 primary antibody with a red secondary antibody, we appropriately saw GAD+ neurons labeled throughout the cortex (Figure 3.2A). However, we observed little colocalization between the GFP+ AC neurons and the GAD+ inhibitory neurons (Figure 3.2B). This was true in all AC layers in which GFP+ neurons were labeled. Quantification of this colocalization across mouse subjects revealed that 95-99% of GFP+ AC neurons were GAD- (Figure 3.2C). These results indicate that the AC neurons that project directly to V1 are almost exclusively excitatory.

Optogenetic stimulation of AC neurons evokes activity in V1

Having observed anatomical evidence that excitatory AC neurons in infragranular layers project directly to V1, we next wanted to determine whether and how activity of these AC neurons translated to activity in V1. This would provide a functional component to the regions' anatomical connection, and could indicate whether AC activity is sufficient to modulate V1 visual responses. We began by injecting a retroAAV-CAG-hChR2-tdTomato construct into V1 in order to express the excitatory opsin channelrhodopsin (ChR2) in neurons that directly project to V1 (Figure 3.3A). During the same surgery, we installed an optic cannula into the auditory cortex to allow delivery of 473nm blue laser to excite neurons expressing ChR2. We then performed electrophysiological recordings in V1 in awake head-fixed mice while presenting pulses of blue laser into the AC. We found that

72

A

B

GFP

GAD

Merge

C

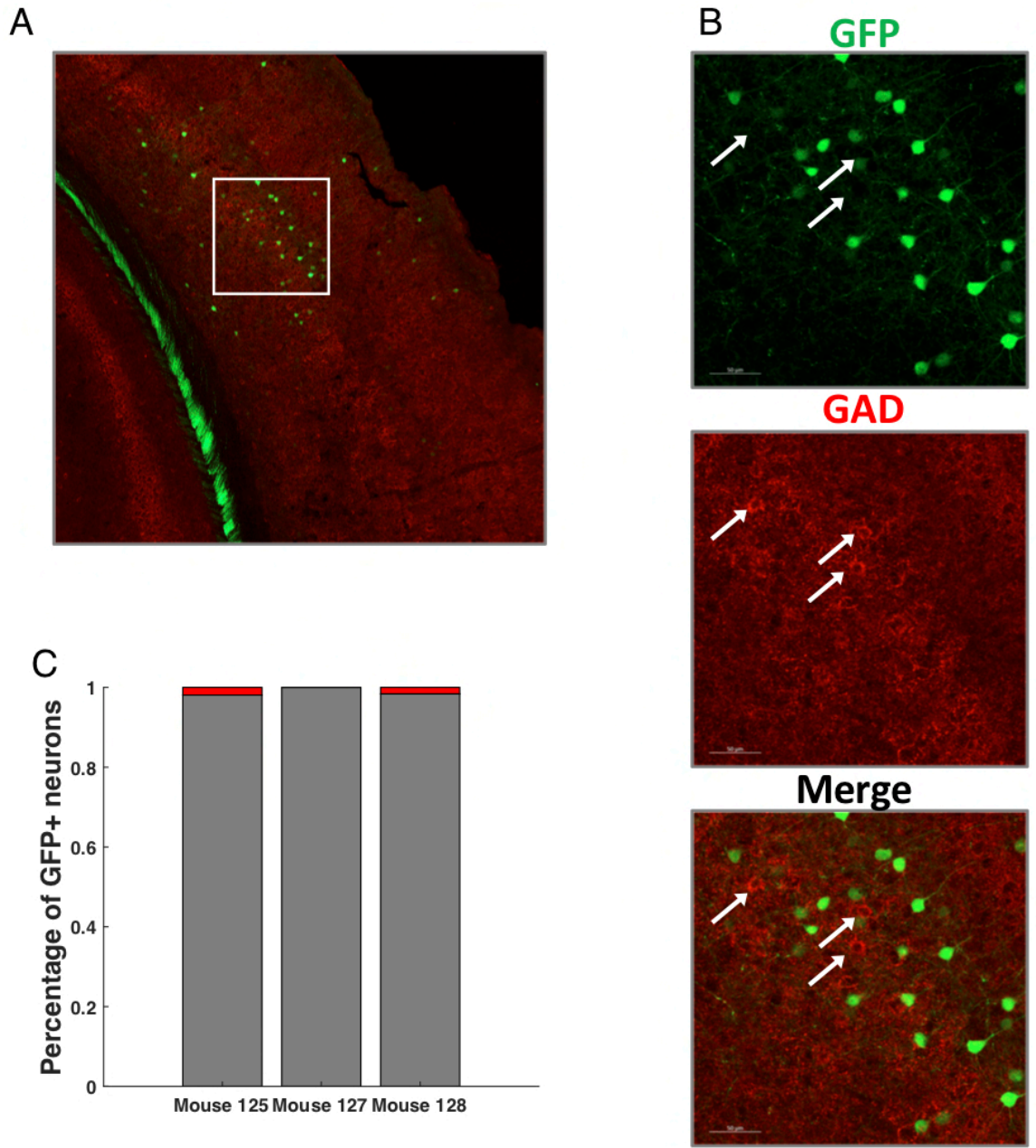**Figure 3.2 | AC$_{V1}$ neurons are predominantly excitatory** (A) A zoomed out image of the retrogradely labeled AC co-stained for GAD with a Alexa Fluor- 568 conjugated secondary antibody. (B) Zoomed in images of the GAD stain, demonstrating little colocalization between GFP+ AC$_{V1}$ neurons and GAD+ inhibitory neurons. (C) Quantification of the GFP and GAD colocalization from three mouse subjects all show minimal colocalization.

54.9% (147/268 units) of V1 neurons were responsive to AC laser stimulation. Responsivity ranged from time-locked spikes (Figure 3.3B, left) to more sustained trains of intermittent spiking activity (Figure 3.3B, right). We then quantified the response latency relative to laser onset. We found a bimodal distribution of latencies, with one group of neurons clustered around 5-8 ms, another group clustered around 12-14 ms, and a small tail of neurons with longer response latencies (Figure 3.3C). These findings indicate that optogenetic stimulation of AC neurons evokes activity in downstream V1 neurons.

We next wanted to determine whether stimulation of this intercortical AC neurons during visual stimulus presentation was sufficient to modulate visual responses in V1. In the same awake and head-fixed mice described above, we delivered optogenetic stimulation to the AC while presenting visual and audiovisual stimuli. We used the same audiovisual stimulus that previously had been determined to evoke robust changes in neuronal activity (outlined in Chapter 2): the visual stimulus consisted of 12 drifting gratings ordered randomly at intermediate contrast, and the auditory stimulus consisted of 70 dB white noise. Consistent with the above findings, we found that AC optogenetic stimulation alone was sufficient to evoke activity in V1 neurons. When we paired the optogenetic stimulation with visual stimulus presentation, we observed further increases in visual response magnitude across the neuronal population (Figure 3.3D; p=3.0e-6, Student's t-test). Furthermore, this laser-enhanced visual response magnitude scaled with the visual response magnitude when sound was present, with no significant difference between the laser-enhanced visual response magnitude and the audiovisual response magnitude (Figure 3.3E; p=0.33, Student's t-test). Together, these results indicate not only

74
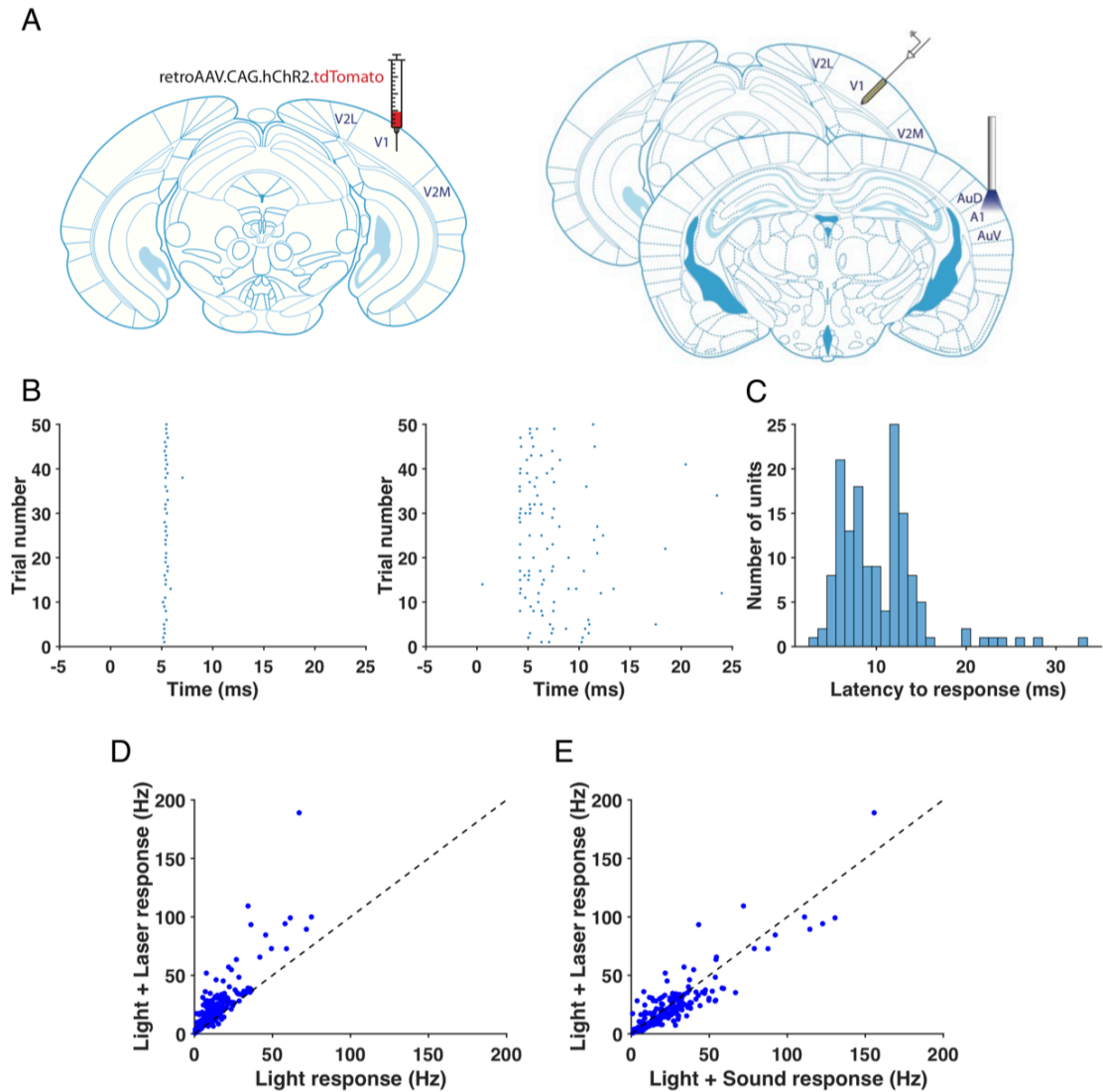
Figure 3.3 | AC stimulation enhances visual responses in V1 (A) Diagram detailing the injection of retroAAV encoding the excitatory opsin hChR2 into V1, followed by simultaneous V1 recording and AC optogenetic stimulation. (B) Example raster plots from units exhibiting time-locked (left) and sustained (right) laser responses. (C) Histogram of the laser response latencies, demonstrating a bimodal distribution. (D) Laser stimulation enhanced the visual response of neurons (p=3.0e-6, Student's t-test). (E) Laser-evoked visual responses were not significantly different from the audiovisual response (p=0.33, Student's t-test).

that optogenetic AC stimulation evokes activity in V1, but also that AC activity is sufficient to enhance visual response magnitude to a similar degree that sound does.

Suppression of AC neurons fails to reduce audiovisual integration in V1

Having observed direct anatomical and functional connections between a subpopulation of AC neurons and V1, we next wanted to directly test whether audiovisual integration in V1 was directly attributable to activity in these AC neurons. Given our own findings and those of other groups, we hypothesized that suppression of activity in these AC neurons would reduce the degree to which sound modulated visual responses in V1. We approached this by injecting retroAAV-hSyn-Cre-GFP into V1, and injecting AAV5-Flex-ArchT-tdTomato into the AC (Figure 3.4A). This combination of viral vectors leads to exclusive Cre-dependent expression of the inhibitory opsin ArchT in AC neurons that project to V1. We first tested whether we were able to successfully suppress AC activity using this approach. We performed electrophysiological recordings in the AC while presenting auditory white noise to awake mice and delivering pulses of 532nm laser to the AC. We found that the laser was able to suppress baseline activity in a subset of neurons (Figure 3.4B-C; 28.6%, 38/133 units; p=1.6e-8, paired t-test). In these same neurons, laser delivery also suppressed sound-evoked activity (Figure 3.4B,D; p=3.8e-8, paired t-test), although few neurons exhibited complete ablation of sound-evoked responses. We also consistently observed rebound spiking activity in these laser-suppressed neurons following laser offset, a common artifact reported with this opsin. These results indicate our ability to optogenetically suppress sound-evoked activity in the AC.

76
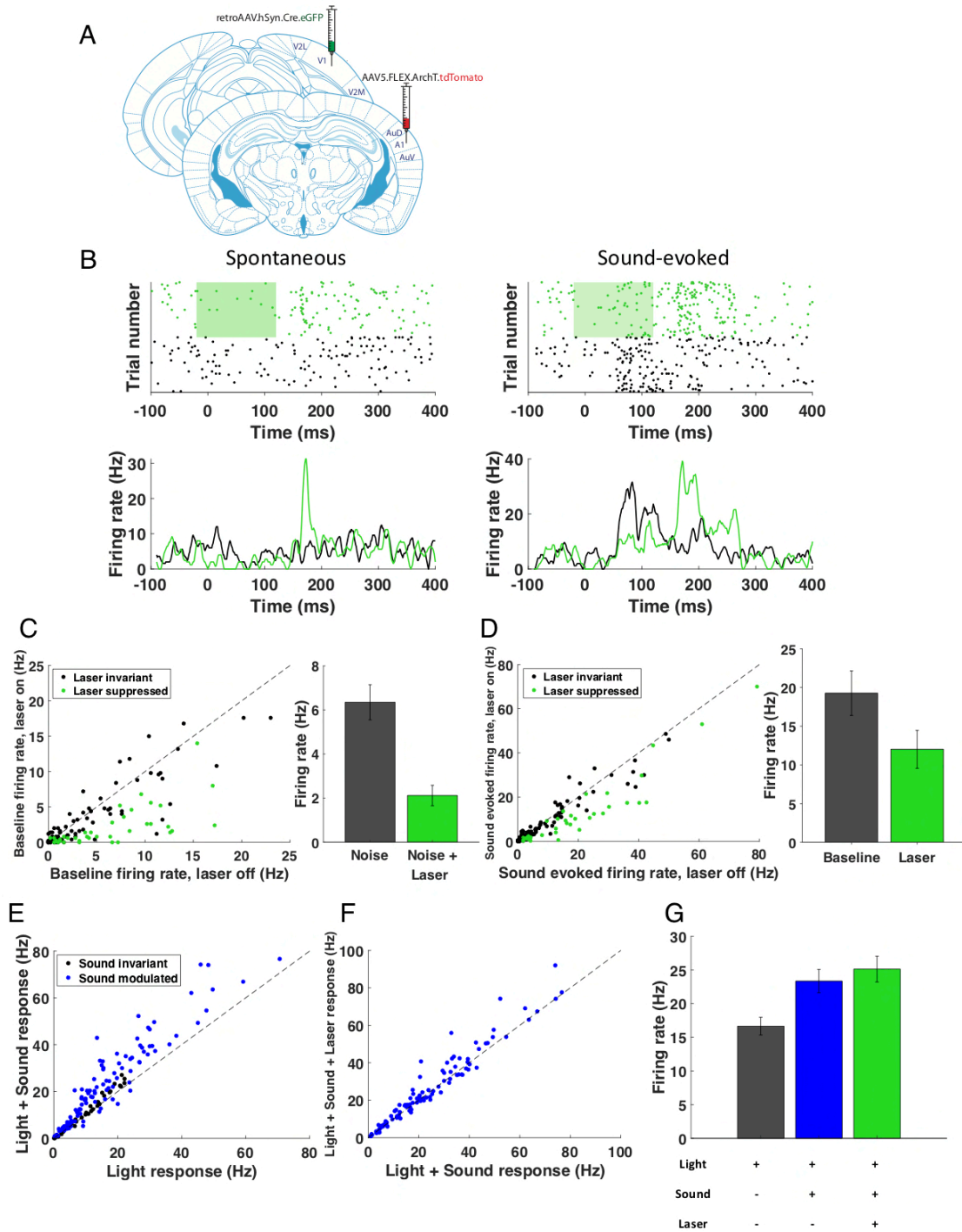
Figure 3.4 | Optogenetic suppression of AC neurons failed to inhibit V1 audiovisual integration (A)
Diagram illustrating the injection of retroAAV encoding Cre into V1, with co-injection of AAV5 encoding
the inhibitory opsin ArchT into the AC. (B) Raster plots and PSTHs from a single example unit demonstrating
laser-suppressed spontaneous and sound-evoked activity. (C) Scatter plot (left) demonstrating baseline and

laser-suppressed activity across the population, with the bar plot quantification from the laser-suppressed neurons on the right (p=1.6e-8, paired t-test). (D) Scatter plot and bar plot quantification of sound-evoked activity across the population (3.8e-8, paired t-test). (E) Sound increased the visual response magnitude in 72.6% of neurons, confirming the previously observed audiovisual effect. (F) 21.7% of sound-modulated light-responsive neurons significantly increased their firing rate with AC projection neurons. (G) Quantification of the visual, audiovisual, and audiovisual response with AC laser, showing paradoxically increased V1 audiovisual responses with AC suppression.

We then tested whether suppressing activity in these AC neurons affected audiovisual integration in V1. We used the same viral and optogenetic approach described above, but instead recorded in V1 while presenting audiovisual stimuli to awake head-fixed mice. We confirmed our previous results by observing that sound increased the magnitude of visual responses in the majority of recorded neurons (Figure 3.4E; 72.6%, 106/146 units). However, surprisingly we did not find that the laser reduced the magnitude of sound-evoked changes in V1. Instead, the majority of neurons exhibited no significant changes in their audiovisual response magnitude when the laser was present (Figure 3.4F), and a minority of neurons paradoxically exhibited higher audiovisual firing rates with laser presentation than without it (Figure 3.4G; 21.7%, 23/106 units; p=1.4e-4, paired t-test). Therefore, optogenetic suppression of sound-evoked activity in projection AC neurons had minimal and a perhaps disinhibitory effect on audiovisual responses in V1.

DISCUSSION

Our previous work identified robust audiovisual integration in V1, entailing changes in response magnitude and timing that resulted in improved neuronal encoding of the visual stimulus (Chapter 2; Williams et al., 2021). We therefore were interested in understanding the circuitry that underlies this cross-sensory phenomenon, specifically determining which auditory area was responsible for providing auditory information to V1 to be integrated with the auditory stream. We began by using retrograde viral tracing to broadly label brain regions that project to V1, and identified the AC as the primary auditory region among that cohort (Figure 3.1). We then characterized these projection neurons as predominantly excitatory (Figure 3.2), and showed that stimulation of these neurons

enhanced visual responses in V1 (Figure 3.3). However, we were surprised to find that suppression of these AC output neurons did not suppress audiovisual integration in V1, and in some neurons paradoxically increased activity (Figure 3.4).

We propose several theories for why suppression of these AC neurons did not reduce audiovisual integration in V1, differing from other studies that have observed such effects (Deneux et al., 2019). First, our optogenetic approach was only able to achieve partial reduction in the sound-evoked activity of AC neurons (Figure 3.3). It is possible that this partially reduced activity was still sufficient to modulate visual responses to the full extent, suggesting some degree of redundancy that this AC activity has with itself and affording it the ability to compensate even when diminished. Secondly, it is known that V1 neurons are sensitive not only to auditory input, but also locomotion as well (Neill and Stryker, 2010; Dardalat and Stryker, 2017). Previously reported studies probing the effect of AC activity on V1 responses used either anesthetized mice (Ibrahim et al., 2016) or excluded trials in which mice displayed noticeable changes in arousal or movement (Deneux et al., 2019). In our reported studies, we included all trials in our analysis regardless of arousal or movement data. It is possible that suppression of AC activity actually did reduce sound's effect on visual responses, but sound-induced movement still enhanced visual responses enough to mask the effect of the laser. However, we have shown previously that sound and movement have evoke distinct changes on the time course of the visual response, with sound primarily enhancing the onset response (Chapter 2; Williams et al., 2021). Therefore it is unclear whether sound-induced movement would be able to effectively mask the effect of the laser on the onset response magnitude. Ultimately, more

research is needed to fully characterize and parse out the tripartite interaction between V1, AC, and motor region activity.

Additionally, other groups have demonstrated using tracing and whole cell recording that AC neurons primarily synapse onto inhibitory neurons in layer 1 and 2/3 of V1 (Ibrahim et al., 2016). Therefore, it is consistent with this known circuitry that we observed neurons that were disinhibited by AC suppression. However, this explanation would not be able to fully account for how this normally feedforward inhibitory circuit would mediate the largely excitatory effect that sound has on visual responses. It is reasonable to speculate, therefore, that either a feedforward excitatory circuit between these cortical regions exists in parallel, or V1 is receiving excitatory auditory input from alternative regions. The AC was the only auditory region consistently labeled by our retrograde tracing. However, it is possible that the retroAAV virus is only uptaken by specific subtypes of neurons, with bias towards cortical circuits. Indeed, we did not observe consistent labeling of the superior colliculus (SC), a known origin of afferent input to V1 (Ahmadlou et al., 2018). And in a single mouse, we observed robust retrograde labeling of neurons in the external and lateral nuclei of the inferior colliculus (IC). These subcortical tracing results lead us to speculate whether activity from the IC is being relayed to V1 either directly or via its dense connections with the SC. This alternative pathway could potentially compensate for our laser suppression of AC, or it could be the principal driver of audiovisual integration in V1. Further studies using tracing, optogenetics, and pharmacological techniques would be necessary to detail the various candidates for cortical and subcortical pathways converging on V1.

METHODS

Mice

All experimental procedures were in accordance with NIH guidelines and approved by the IACUC at the University of Pennsylvania. Mice were acquired from Jackson Laboratories (7 male, 8 female, aged 10-18 weeks at time of experimentation; B6.Cast-*Cdh23*^*Ahl+* mice [Stock No: 018399]) and were housed at 28°C in a room with a reversed light cycle and food provided ad libitum. Experiments were carried out during the dark period. Mice were housed individually after viral injection or headplate and optic cannula implantation. Euthanasia was performed using CO_2, consistent with the recommendations of the American Veterinary Medical Association (AVMA) Guidelines on Euthanasia. All procedures were approved by the University of Pennsylvania IACUC and followed the AALAC Guide on Animal Research. We made every attempt to minimize the number of animals used and to reduce pain or discomfort.

Surgical procedures

Mice were implanted with skull-attached headplates to allow head stabilization during recording, and skull-penetrating ground pins for electrical grounding during recording. The mice were anesthetized with 2.5% isoflurane. A ~1mm craniotomy was performed over the right frontal cortex, where we inserted a ground pin. Additionally, a craniotomy was performed over the left auditory cortex (2.6mm posterior of Bregma, 4.2mm lateral of midline), where an optic cannula was inserted. A custom-made stainless steel headplate (eMachine Shop) was then placed on the skull at midline, and the ground pin, the optic cannula, and headplate were fixed in place using C&B Metabond dental

cement (Parkell). Mice were allowed to recover for 3 days post-surgery before any additional procedures took place.

Viral injections

During surgery, viral vectors were injected to express fluorophores or opsins in neurons to enable labeling and optogenetic manipulations, respectively. Viral particles were injected (500 nL) unilaterally via glass syringe (30-50μm) using a syringe pump (Pump 11 Elite, Harvard Apparatus) targeted to AC (~2.6mm posterior from Bregma, ~4.2mm lateral of midline) or V1 (0-0.5mm posterior of lambdoid suture, 2.0-2.5mm lateral of midline). The viruses used were retroAAV-hSyn-GFP (Addgene #50465), retroAAV-CAG-hChR2-tdtomato (Addgene #8017), retroAAV-hSyn-Cre-eGFP (Addgene #105540), and AAV5-Flex-ArchT-tdTomato (Addgene #28305). For optogenetic experiments, fiber-optic cannulas (ThorLabs, ⌀200μm Core, 0.22 NA) were implanted in the craniotomy over the auditory cortex and secured using silicon and C&B Metabond dental cement (Parkell). Craniotomies without hardware implanted were filled using bone wax.

Fluorescent tracing and immunohistochemistry

For tracing experiments, 2-3 weeks following viral injection, mice were anesthetized using a ketamine/dexmedetomidine combination, and perfused with saline. Brains were removed from the skull and placed in 4% PFA overnight, followed by 48 hours submersion in 30% sucrose. Fixed brains were then sliced into 40μm sections and either directly mounted on slides for visualization, or placed in cryoprotective media for long-

term storage. Immunohistochemical staining was performed to test for colocalization with GAD. Sections were first pretreated in a citrate buffer at 75°C for 30 minutes then washed 3x with 0.3% Triton X-100 in PBS at room temperature to enhance membrane permeability. Sections were then transferred to blocking solution of 0.5% Triton X-100 and 10% goat serum in PBS for 1 hours. The primary antibody solution consisted of the same blocking solution, with 1:200 primary mouse anti-GAD67 (Millipore #MAB5406), applied overnight. Sections were then washed 3xwith 0.3% Triton X-100 in PBS, and secondary goat anti-mouse antibody, Alexa Fluor-568-conjugated was applied 1:500 in blocking solution for 1 hour. Sections were then washed in PBS, and mounted on slides for visualization.

Electrophysiological recordings

All recordings were carried out inside a custom-built acoustic isolation booth. 1-2 weeks following the headplate and ground pin attachment surgery, we habituated the mice to the recording booth for increasing durations (5, 15, 30 minutes) over the course of 3 days. On the day of recording, mice were placed in the recording booth and anesthetized with 2.5% isoflurane. We then performed a small craniotomy above either the left auditory cortex (2.6mm posterior of Bregma, 4.2mm lateral of midline) or the left primary visual cortex (2.5mm lateral of midline, 0-0.5 mm posterior of the lambdoid suture). Mice were then allowed adequate time to recover from anesthesia. Activity of neurons were recorded using a 32-channel silicon probe (NeuroNexus A1x32-Poly2-5mm-50s-177). The electrode was lowered into the cortex via a stereotactic instrument to a depth of 775-1000μm. Following the audiovisual stimulus presentation, electrophysiological data from

all 32 channels were filtered between 600 and 6000 Hz, and spikes belonging to single neurons and multi-units were identified in a semi-automated manner using KiloSort2 (Pachitariu et al., 2016).

Audiovisual stimuli

The audiovisual stimuli were generated using MATLAB (MathWorks, USA), and presented to mice on a 12" LCD monitor (Eyoyo) and through a magnetic speaker (Tucker-Davis Technologies) placed to the right of the mouse. The visual stimulus was generated using the PsychToolBox package for MATLAB and consisted of square wave drifting gratings 1 s in duration, 4-Hz temporal frequency, and 0.1 cycles/°. The gratings moved in 12 directions, evenly spaced 0°-360°, and were scaled to a range of 5 different visual contrast levels (0, 0.25, 0.5, 0.75, 1), totaling 60 unique visual stimuli. The auditory stimulus was sampled at 400 kHz and consisted of a 1 s burst of 70 dB white noise. The visual grating was accompanied by the auditory noise on half of trials (120 unique trial types, 10 repeats each), with simultaneous onset and offset. The auditory-only condition corresponded to the trials with a visual contrast of 0. The trial order was randomized and was different for each recording.

Optogenetic stimulation and suppression

Optogenetic stimulation of ChR2 was performed using a blue 473nm laser (BL473T3-150). The laser was administered through an optic cannula, and was pulsed at 40 Hz with 50% duty cycle. Optogenetic suppression of ArchT was performed using a green 532 nm DPSS laser (GL532T3-300, Slocs 155 lasers, 3 mW power at cannula tip or

OptoEngine, MGL-III-532, 15 mW power at cannula tip). This laser was similarly administered through an optic cannula, however it was not pulsed.

Data analysis and statistical procedures

Spiking data from each recorded unit was organized by trial type and aligned to the trial onset. The number of spikes during each trial's first 0-300ms was input into a generalized linear model (GLM; predictor variables: visual contrast [continuous variable 0, 0.5], sound [0 or 1]; response variable: number of spikes during 0-300ms; Poisson distribution, log link function), allowing the classification of each neuron's responses as having a main effect ($p<0.05$) of light, sound, and/or a light-sound interaction. Neurons that were responsive to both light and sound or had a significant light-sound interaction term were classified as "light-responsive sound-modulated."

Statistics

Figure data are displayed as means with standard error of the mean (SEM), unless otherwise noted. Shapiro-Wilk tests were used to assess normality, and the statistical tests performed are indicated in the text, figures, and Table 1. For multi-group and multivariate analysis (e.g., ANOVA and Kruskal-Wallis tests) in which a significant ($p<0.05$) interaction was detected, we subsequently performed a post hoc Bonferroni-corrected test.

CHAPTER 4: THE SUPERIOR COLLICULUS AND AUDIOVISUAL INTEGRATION IN THE AUDITORY MIDBRAIN

ABSTRACT

As an essential component of sensory processing, auditory and visual signals are integrated to improve acuity and processing of the sensory signals. Despite anatomic connections between sensory pathways occurring in many regions throughout the brain, cortical audiovisual integration has received more attention, whereas subcortical integration has been less well studied. We therefore investigated the circuits and codes used in audiovisual integration in the inferior colliculus (IC), the first region in the ascending auditory pathway to receive cross-modal input. We began by performing retrograde tracing, and identifying the neighboring superior colliculus (SC) as the primary visual region that projected to the IC. Transsynaptic anterograde tracing revealed that the SC synapses with neurons in the external shell of the IC. Stimulation of these SC neurons resulted in increased baseline and sound-evoked activity in the inferior colliculus. Despite the connections from the SC, we failed to observe visual or audiovisual responsiveness in the IC using static, drifting gratings, and looming visual stimuli. The presence of robust anatomic input from the SC, the lack of observed audiovisual integration in the IC suggests that neurons in this auditory region have narrow specificity and tuning for visual stimuli.

INTRODUCTION

Everyday perception relies on integration of incoming visual and auditory signals. The cross-communication between sensory streams is important for communicating, spatial navigation, and movement coordination. This importance is highlighted within the auditory system of patients with hearing aids who increasingly rely on visual cues to improve perception of auditory features degraded by auditory assistive devices (Arnold and Köpsel, 1996), as well as those with neurological injuries resulting in balance and movement impairment when closing their eyes (Forbes and Cronovich, 2020). Despite the importance of audiovisual and multisensory integration, the neuronal mechanisms that control this process are still being uncovered.

The inferior colliculus (IC) is the first region in the ascending auditory pathway to receive cross-modal sensory inputs. Non-lemniscal regions of the IC receive inputs from cortical auditory areas, somatosensory regions, as well as visual brain regions such as the superior colliculus (SC; Gruters and Groh, 2012). Studies in the barn owl have demonstrated a rich interconnected network of projections between auditory and visual subdivisions of both the superior and inferior colliculi, important for development of visual spatial tuning in the SC (Knudsen, 1985; Brainard and Knudsen, 1993). Despite the dense bidirectional projections between these neighboring regions, how activity in the visual SC directly modulates activity in the auditory IC has not been directly studied. Understanding how this subcortical circuit operates can provide insight into the function of audiovisual integration in this early auditory processing center.

The audiovisual responses of neurons within the IC have also largely been described in the barn owl. Studies showed that IC neurons sharpen their spatial tuning with

coincident visual input from congruent spatial locations (Bergan and Knudsen, 2009). However, whether IC neurons are also sensitive to a broader range of visual stimuli is unclear. Furthermore, the mouse IC receives auditory and other cross-modal projections (Gruters and Groh, 2012; Lesicko et al., 2020), but whether audiovisual responses are also observed in the mouse IC and how the visual input is integrated with the auditory information in these neurons remains unclear. Studying the parameters of audiovisual processing in the mouse IC, particularly in the context of the subcortical circuitry present in the region, will expand our understanding of audiovisual integration in a model organism in which experimental interventions are readily available.

In the present study, we seek to understand the role of subcortical projections from the SC in mediating audiovisual integration in the IC. We use tracing and optogenetic techniques to demonstrate that activity in the SC is capable of modulating spontaneous and sound-evoked responses in the IC. We also study whether visual drifting gratings and looming stimuli are capable of evoking changes in neuronal activity in the IC, with the goal of determining the role of the SC in this visual tuning. However, we find the negative result that the IC is insensitive to this set of visual stimuli, leaving open the role of this circuit in audiovisual integration in mice.

RESULTS

Retrograde tracing reveals afferent visual projections

We began investigating audiovisual integration in the IC by characterizing the anatomical inputs to the brain region. We used retrograde tracing to identify the visual regions that project directly to the IC. Prior studies in the barn owl indicate that the SC is

89

the primary visual region that synapses with the IC (Brainard and Knudsen, 1993), and we wanted to confirm that this was also true in our mouse model. We injected retroAAV-Cre into mice with global expression of Cre-dependent flex-tdTomato, allowing labeling and identification of afferent neurons through activated expression of tdTomato. Using this technique, we observed labeling in various neighboring and distant brain regions (Figure 4.1). Fluorescent cell bodies in the infragranular layers of the auditory cortex (AC) confirmed retrograde action of the retroAAV virus. The SC was the primary visual region with observable tdTomato+ cell bodies, located in both superficial and deep layers. Additionally, we identified sparse cell body labeling in the primary visual cortex (V1), Ultimately, this approach confirmed the SC as the primary visual region projecting to the IC, although additional projections from V1 are present as well.

Superior colliculus projects to the IC external shell

We were next interested in characterizing the IC neurons that received these projections from the SC, hereto referred as IC_{SC} neurons. The external shell of the IC, consisting of the external and lateral nuclei, are traditionally considered the non-lemniscal regions of the IC and receive cross-modal non-auditory input. Therefore, we hypothesized that the IC_{SC} neurons would primarily be located in this IC subregion. We tested this by performing a novel anterograde tracing technique using the transsynaptic anterograde properties of AAV1 (Zingg et al., 2017). A small percentage of AAV1 is capable of jumping to the downstream synapse, and can be used in combination with the Cre-Lox system to trigger expression exclusively in the postsynaptic neurons. We injected AAV1-Cre into the superior colliculus of wildtype mice and also injected AAV9-flex-tdTomato
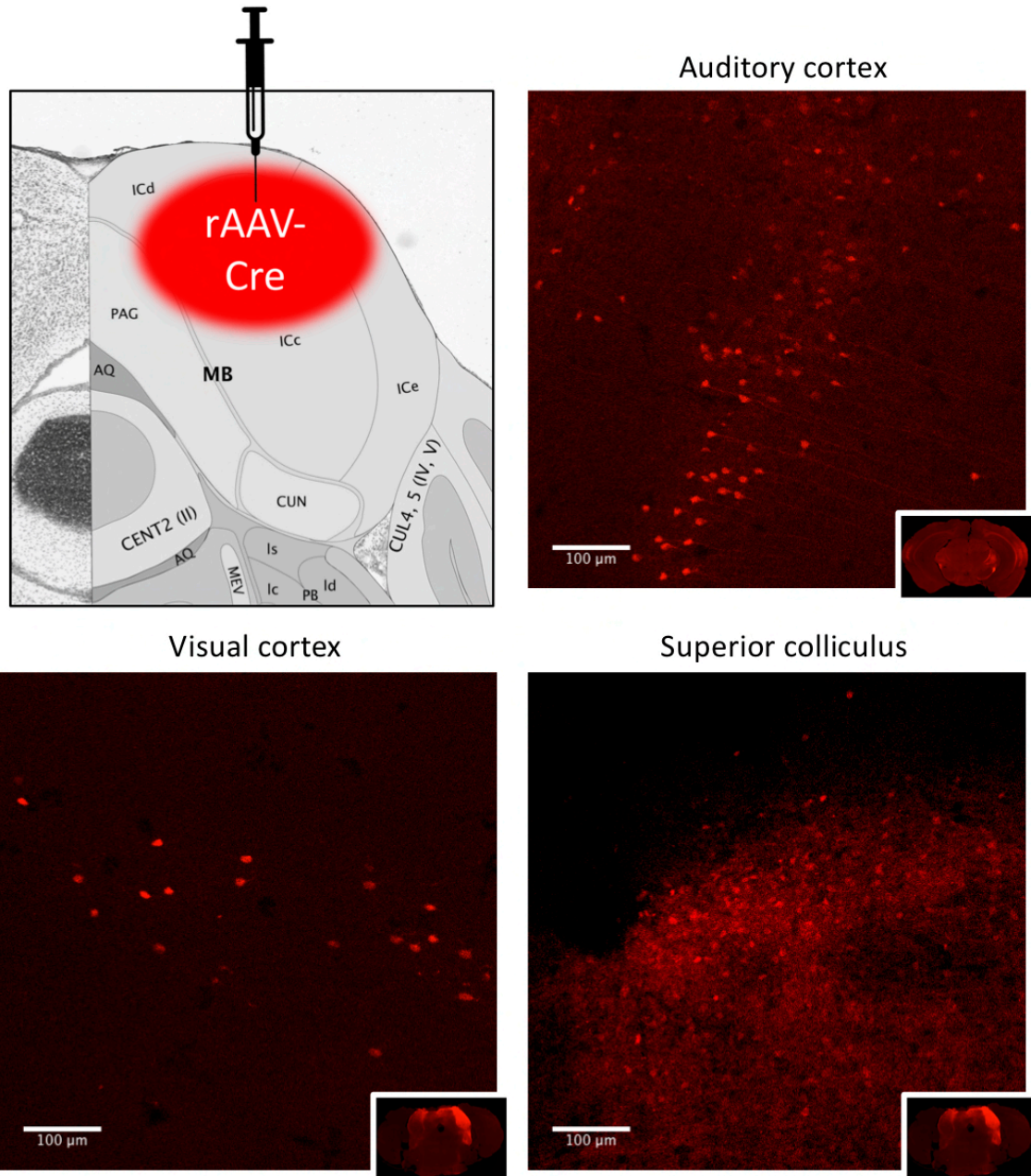
90

Figure 4.1 | The superior colliculus is the primary visual input to the inferior colliculus (A) Retrograde tracing using retroAAV-encoding Cre, injected into a mouse globally expressing flex-tdTomato, labeled neurons in various regions including the auditory cortex, visual cortex, and superior colliculus.

into the IC (Figure 4.2A). This technique would label IC_{SC} neurons as tdTomato+. Using this approach, we successfully labeled IC_{SC} neurons, consistently observing tdTomato+ neurons in mouse subjects (Figure 4.2B). We registered the brain sections with the Allen Brain Atlas (Shamash et al., 2018) and determined that the majority of IC_{SC} neurons were in fact located in the IC external and dorsal nuclei, with very few labeled neurons in the central nucleus (Figure 4.2D). Additionally, sole injection of AAV9-flex-tdTomato into the IC of a control mouse without injecting AAV1-Cre led to no fluorescent neurons in the IC, confirming that the observed labeling in the experimental group was due to activity of the AAV1 virus and not aberrant expression of the downstream virus.

Given the bidirectional connections between the SC and the IC, we wanted to confirm that this AAV1-mediated labeling was in fact due to transsynaptic anterograde labeling, and not due to retrograde activity of the virus. We used this same viral genetic approach in a known unidirectional pathway, the top-down AC-IC connection. We injected AAV1-Cre into the AC and AAV9-flex-tdTomato into the IC. Anterograde activity of the virus would lead to fluorescent neurons in the IC, whereas retrograde activity would lead to no expression due to a lack of IC neurons that project to the AC. We did observe tdTomato+ IC_{AC} neurons in these mice, with the majority of neurons being located in the dorsal nucleus (Figure 4.2C-D). The use of this AC-IC control pathway confirms anterograde activity of the AAV1 virus, and confirms that we successfully identified IC_{SC} neurons in the external and dorsal nuclei of the IC.
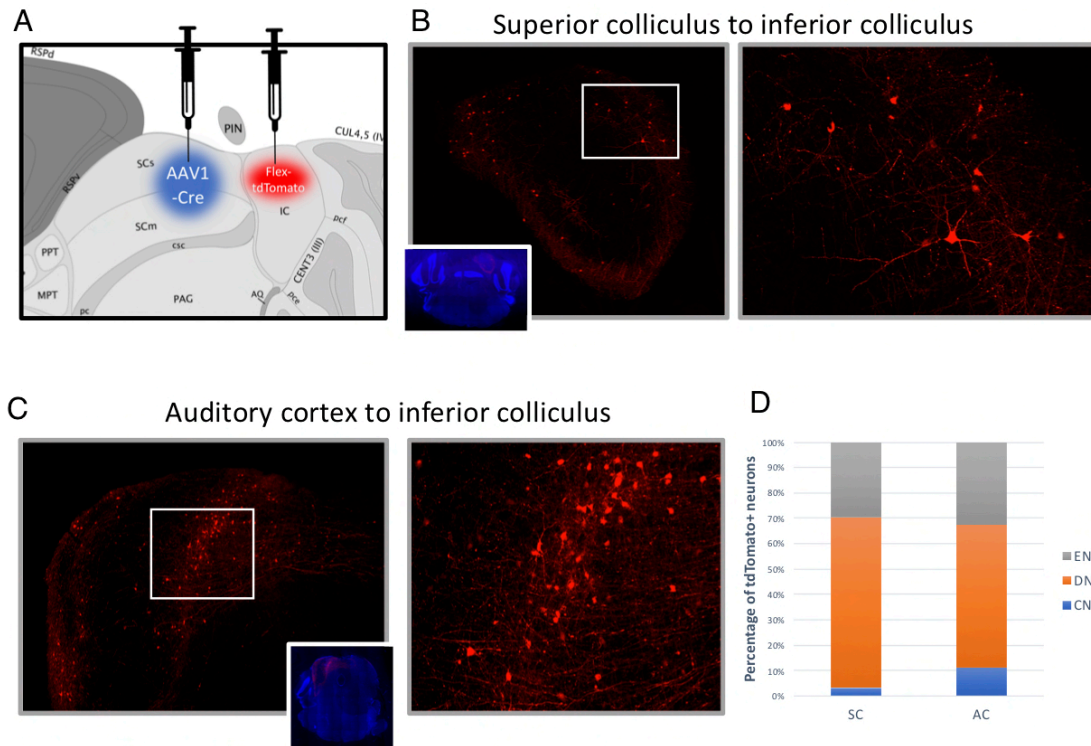
Figure 4.2 | The superior colliculus projects to the external shell of the inferior colliculus (A) Diagram illustrating the injection of AAV1-Cre into the SC and AAV9-flex-tdTomato for anterograde tracing. (B) Example section from brain with anterograde tracing labeled neurons in the IC. (C) Example section from brain with anterograde tracing between the AC and IC also labeled neurons in the IC. (D) Quantification of brain sections from these tracing experiments demonstrates that the AC and SC both predominantly project to the external shell of the IC.

Optogenetic stimulation of the SC modulates baseline and sound-evoked IC activity

Having identified the SC as the primary visual region connecting to the IC and described the distribution pattern of these IC$_{SC}$ neurons within the IC, we next wanted to complement this anatomic understanding with a functional one. Specifically, we wanted to determine how SC activity affects IC baseline and sound-evoked activity. We began by injecting AAV5-hSyn-hChR2-mCherry into the SC and installing an optic cannula in the region, and then performing electrophysiological recordings in the IC while delivering 473 nm light to the SC to stimulate those neurons (Figure 4.3A). We found that in the absence of sound, SC stimulation evoked activity in the IC in a subset of neurons (Figure 4.3B). These IC neurons responded with low and consistent latency (3-7ms), suggesting a monosynaptic connection from the SC. These findings indicate that SC activity leads to an increase in the baseline activity of IC neurons, predominantly in a low-latency time-locked manner.

We then optogenetically stimulated the SC while presenting 70 dB white noise to the awake head-fixed mice in order to determine whether SC activity modulates sound-evoked activity in this downstream region. We again found that SC stimulation enhanced the firing rate of IC neurons (Figure 4.3C). We observed an even distribution between neurons whose onset response magnitude was enhanced and other neurons whose sustained portion of the sound response was enhanced by SC activity. On average, the percentage increase in firing rate was relatively small (Figure 4.3D; p=9.9e-11, paired t-test). Nonetheless, these results together indicate that exogenous stimulation of SC activity translates into increased activity in the downstream IC.
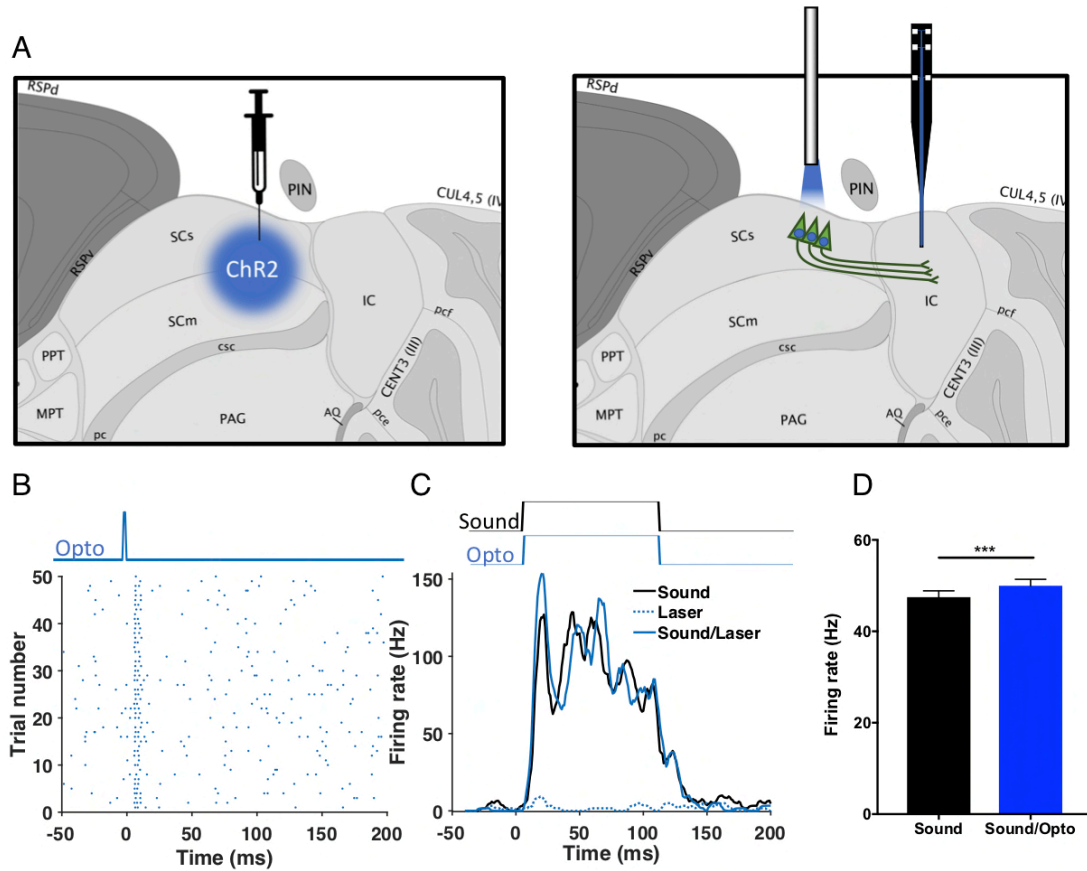
Figure 4.3 | Optogenetic stimulation of the superior colliculus enhances sound responses in the inferior colliculus (A) Diagram illustrating injection of AAV5-ChR2 into the SC, followed by IC recording and optogenetic stimulation of the SC. (B) Raster plot of example IC unit demonstrating a time-locked response to SC stimulation. (C) PSTH of example unit demonstrating laser-enhanced sound-responses. (D) Quantification across the entire population, demonstrating SC laser stimulation enhances IC sound-responses (p=9.9e-11, paired t-test).

Visual stimuli evoke little change in IC auditory response

Finally, having characterized the anatomic and functional connection between the IC and its primary visual input, the SC, we wanted to determine how visual input affected baseline and sound responses in the IC. IC visual responses have primarily been described in the barn owl animal model, so we wanted to determine whether there was similar visual responsiveness in the IC of our mouse model. We presented audiovisual stimuli to awake head-fixed mice while performing electrophysiological recordings in the IC. For one audiovisual stimulus, the visual component of the stimulus consisted of drifting gratings of random directions, and the auditory component was white noise at a range of sound intensities (0-80 dB). For another audiovisual stimulus, the visual component was looming and receding black circle that was growing or shrinking in size, respectively, while the paired auditory stimulus was white noise that grew or lessened in intensity (Figure 4.4A). For both sets of audiovisual stimuli, we used a 2-way ANOVA to determine the effects of the auditory and visual components on the spiking activity of each neuron. We found that the majority of recorded neurons in the IC were sound-responsive to the static white noise or the amplitude-modulated noise. However, very few (~5%) of neurons were responsive to the visual component of either audiovisual stimulus (Figure 4.4B), and none of these neurons were significantly responsive to visual stimulation following multiple comparisons corrections. This suggests that the IC was unresponsive to this set of visual stimuli, either through their baseline or sound-evoked activity.

We also studied the audiovisual responses of IC_{SC} neurons to determine whether their properties were distinct from the rest of the recorded population. We approached this by again using the AAV1-mediated transsynaptic labeling technique, however in this
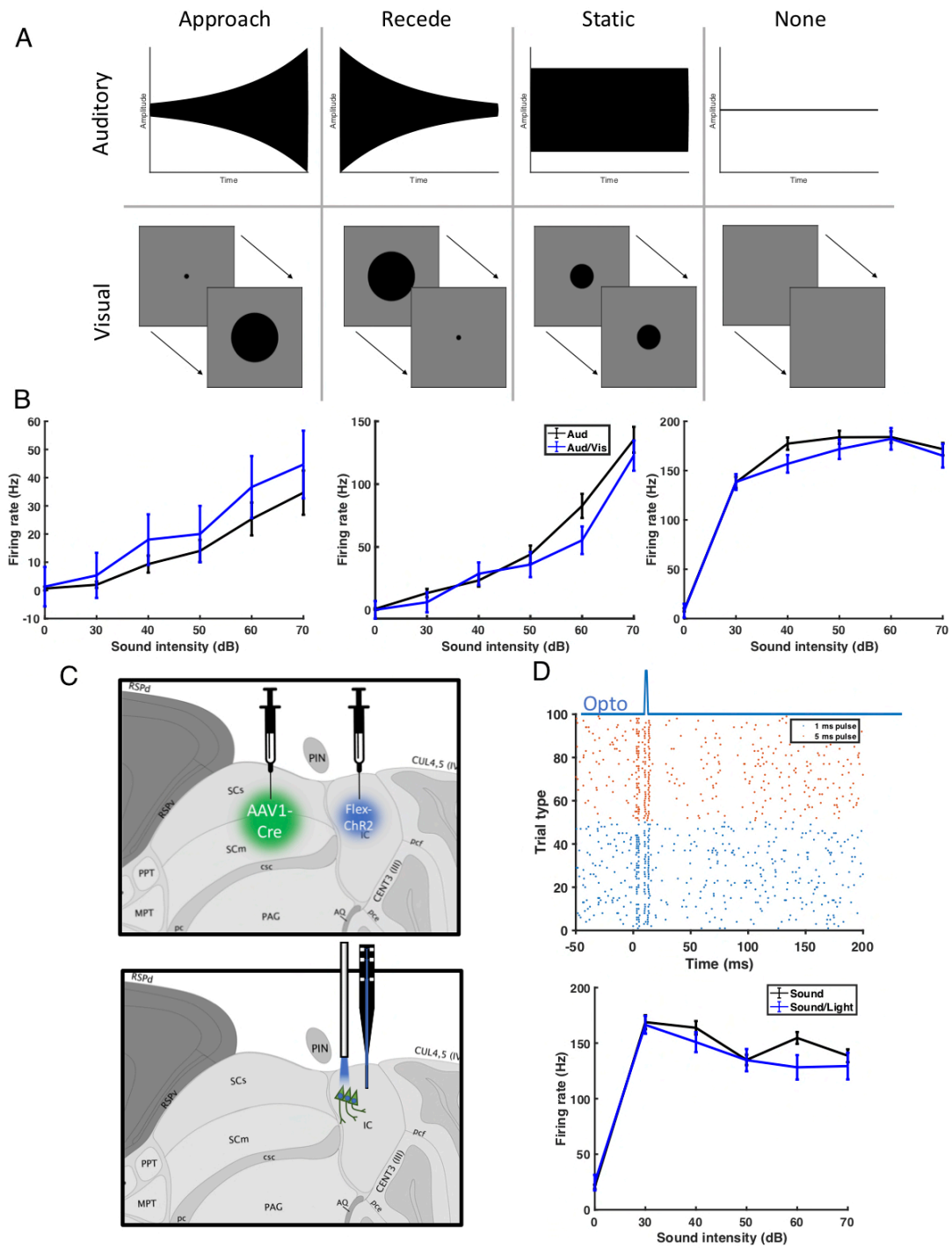
Figure 4.4 | Inferior colliculus neurons are minimally responsive to looming visual stimuli (A) Diagram of the approaching and receding variations of the looming audiovisual stimuli. (B) Quantification of three example IC units initially classified as responsive to visual input, however were not significantly responsive

following multiple comparisons correction. (C) Diagram illustrating the viral injection of AAV1-Cre into the SC and AAV9-Flex-ChR2 into the IC to allow optotagging of IC_{SC} neurons. (D) Optotagging (top) and audiovisual responses (bottom) of example opto-tagged IC_{SC} unit.

iteration we used it to induce expression of ChR2 in the IC (Figure 4.4C). We then performed electrophysiological recordings in the IC, and delivered pulses of 473 nm light to opto-tag these IC$_{SC}$ neurons prior to presenting the same set of audiovisual stimuli described above. Opto-tagging of the IC$_{SC}$ was successful, with laser delivery evoking an extremely low latency, 5-10 ms train of spikes in a small subset of neurons (Figure 4.4D). However, similar to above, while these neurons were sound responsive, they did not exhibit any pronounced visual responsiveness following 2-way ANOVA classification with multiple comparisons corrections (Figure 4.4D). Together, these results indicate that despite anatomic and functional connections between the SC and IC, IC neurons in mice have visual tuning properties that are not tuned for visual drifting gratings or looming stimuli.

DISCUSSION

We were interested in understanding how visual input affected auditory processing in the IC. Beginning from a circuit perspective, we used retrograde tracing to identify the SC as the primary presynaptic visual region that projected to the IC, with additional afferent visual input from V1. And we used an AAV1-mediated transsynaptic anterograde tracing technique to determine that these SC afferents primarily synapse within the external shell of the IC, consistent with other cross-modal inputs to this non-lemniscal region. Transitioning from a circuit to a more physiologic perspective, we found that stimulation of these SC afferents increased both spontaneous and sound-evoked activity in the IC. However, we were unable to identify visual stimuli that consistently evoked changes in neuronal firing rates in the IC, either at baseline or with coincident auditory input.

99

Ultimately these findings demonstrate that despite receiving dense projections from the neighboring SC, the auditory midbrain is relatively insensitive to visual drifting gratings and looming stimuli and remains specific in which visual and audiovisual stimuli it is tuned for.

The tuning properties of sensory brain regions is an important topic within sensory neuroscience. Mapping of the tuning patterns to various stimulus features, e.g. auditory volume, auditory frequency, visual grating direction, etc, demonstrates what information is important to the neuron and therefore potentially the neuronal population and brain as a whole. This is a question that applies to the field of multisensory and audiovisual integration as well. In the present study, we found that neurons in the IC are relatively insensitive to visual input in the form of drifting gratings or looming stimuli. However, this does not preclude this auditory region from responding to other forms of visual input. Neurons in the barn owl IC modulate their spatial tuning properties based on the location of simultaneous visual input (Bergan and Knudsen, 2009). Therefore, it is reasonable to speculate that neurons in the mouse IC are similarly spatially tuned, and therefore do not effectively code for visual grating orientation or looming speed. We did present the auditory and visual stimuli from the same direction, an important component in the barn owl studies, but perhaps it is necessary the vary the spatial origin of the audiovisual stimulus to effectively detect the effects of visual stimuli on the auditory tuning.

It is also reasonable to speculate whether audiovisual integration in the mouse IC is sensitive not to the spatial origin of the stimulus, but rather another unstudied audiovisual stimulus feature. This brings into question the ethological relevance of sensory processing and the experimental designs used to study it. The barn owl, a predatory bird that navigates

3D space to capture prey, relies heavily on its auditory system to locate the spatial origin of sounds with high precision (Carr and Christensen-Dalsgaar, 2015). Conversely, mice may not rely as heavily on the auditory system, or more importantly on visual modulation of auditory tuning, for its regular ethological behavior. Looming stimuli are thought to evoke behavioral and neuronal patterns similar to incoming predators (Yilmax and Meister, 2013). However, perhaps a different set of looming stimulus parameters, or an entirely different visual stimulus, would better evoke neuronal patterns relevant to behavior in the mouse ecosystem.

Finally, it is possible that the projections from SC to the IC in the mouse brainstem are not used for audiovisual integration in the IC. Given the dense bidirectional projections between the superior and inferior colliculi, it is reasonable to consider whether the role of the SC to IC projections is to tune, calibrate, and prune the reciprocal IC-to-SC input. It has been shown, again in the barn owl, that the SC uses the auditory space map in the IC to help learn and sharpen the visual spatial map within the SC itself (Brainard and Knudsen, 1993; Bergan and Knusden, 2009). These mutual connections could therefore be a component of that process, rather than relevant to the auditory system itself. Longitudinal studies of this bidirectional pathway throughout development, the use of objects that warp visual or auditory space, and optogenetic or pharmacologic probing of the circuits would be necessary to fully understand the role of this pathway in that function.

Corticofugal visual projections to the IC

In the present study, we focused on the SC-IC circuitry because we identified the SC as the primary visual input to the region using retrograde tracing. However, the

technique also labeled neuronal cell bodies within V1. The presence of parallel projections from visual regions to the IC is an interesting finding. While it is possible that these projections serve redundant or compensatory roles, it is likely that the subcortical and cortical pathways have distinct functions. Traditionally, cortical circuitry is thought to be more involved in learning associations and pattern recognition, information that is then transmitted to brainstem, spinal, and motor regions to modulate more reflexive or faster behaviors. Whether this subcortical and corticofugal audiovisual circuitry is an example of this distinction in circuit function remains unclear and is worthy of further investigation.

Transsynaptic anterograde tracing in neuronal circuits

The field of neuroscience has benefitted greatly from transsynaptic retrograde tracing techniques, however transsynaptic anterograde tracing techniques have been less widely used. The transsynaptic properties of AAV1 were first reported and used in a cortical circuits that terminate in the SC (Zingg et al., 2017; 2020). In the present study, we were able to successfully use AAV1 to fluorescently label and optogenetically tag downstream neurons in SC-IC and AC-IC circuits. The success of this technique in these various studies raises the question about why transsynaptic anterograde techniques are less well used in systems neuroscience. A potential explanation is differences between cortical and subcortical neurons. Viruses such as AAV1 and other viral serotypes rely on cell surface markers and internal machinery to transduce a neuron, replicate within it, and then jump to neighboring or downstream neurons. Perhaps cell surface makers of subcortical neurons allow easier transfer of AAV1 particles to postsynaptic dendrites than cortical

neurons. It will be useful to consider viral serotype and the circuit of study when performing anterograde, retrograde, and other viral transduction techniques.

METHODS

Mice

All experimental procedures were in accordance with NIH guidelines and approved by the IACUC at the University of Pennsylvania. Mice were acquired from Jackson Laboratories (7 male, 7 female, aged 10-18 weeks at time of recording; B6.Cast-*Cdh23*$^{Ahl+}$ mice [Stock No: 018399]) and were housed at 28°C in a room with a reversed light cycle and food provided ad libitum. Experiments were carried out during the dark period. Mice were housed individually after viral injection or headplate and optic cannula implantation. Euthanasia was performed using CO_2, consistent with the recommendations of the American Veterinary Medical Association (AVMA) Guidelines on Euthanasia. All procedures were approved by the University of Pennsylvania IACUC and followed the AALAC Guide on Animal Research. We made every attempt to minimize the number of animals used and to reduce pain or discomfort.

Surgical procedures

Mice were implanted with skull-attached headplates to allow head stabilization during recording, and skull-penetrating ground pins for electrical grounding during recording. The mice were anesthetized with 2.5% isoflurane. A ~1mm craniotomy was performed over the right frontal cortex, where we inserted a ground pin. Additionally, a craniotomy was performed over the left superior colliculus (0.5-1mm posterior of lambdoid

suture, 1.0-1.5mm lateral of midline), where an optic cannula was inserted. A custom-made stainless steel headplate (eMachine Shop) was then placed on the skull at midline, and the ground pin, the optic cannula, and headplate were fixed in place using C&B Metabond dental cement (Parkell). Mice were allowed to recover for 3 days post-surgery before any additional procedures took place.

Viral injections

During surgery, viral vectors were injected to express fluorophores or opsins in neurons to enable labeling and optogenetic manipulations, respectively. Viral particles were injected (500 nL) unilaterally via glass syringe (30-50μm) using a syringe pump (Pump 11 Elite, Harvard Apparatus) targeted to SC (0.5-1mm posterior of lambdoid suture, 1.0-1.5mm lateral of midline) or IC (2-2.5mm posterior of lambdoid suture, 1.0-1.5mm lateral of midline). The viruses used were retroAAV-hSyn-Cre (Addgene #105553), AAV1-hSyn-Cre (Addgene #105553-AAV1), AAV5-hSyn-ChR2-mCherry (Addgene #26976), and AAV9-CAG-Flex-ChR2-tdTomato (Addgene #18917). For optogenetic experiments, fiber-optic cannulas (ThorLabs, ∅200μm Core, 0.22 NA) were implanted in the craniotomy over the superior colliculus and secured using silicon and C&B Metabond dental cement (Parkell). Craniotomies without hardware implanted were filled using bone wax.

Fluorescent tracing and immunohistochemistry

For tracing experiments, 2-3 weeks following viral injection, mice were anesthetized using a ketamine/dexmedetomidine combination, and perfused with saline.

Brains were removed from the skull and placed in 4% PFA overnight, followed by 48 hours submersion in 30% sucrose. Fixed brains were then sliced into 40μm sections and directly mounted on slides for visualization.

Electrophysiological recordings

All recordings were carried out inside a custom-built acoustic isolation booth. 1-2 weeks following the headplate and ground pin attachment surgery, we habituated the mice to the recording booth for increasing durations (5, 15, 30 minutes) over the course of 3 days. On the day of recording, mice were placed in the recording booth and anesthetized with 2.5% isoflurane. We then performed a small craniotomy above the left inferior colliculus (2-2.5mm posterior of lambdoid suture, 1.0-1.5mm lateral of midline). Mice were then allowed adequate time to recover from anesthesia. Activity of neurons were recorded using a 32-channel silicon probe (NeuroNexus A1x32-Poly2-5mm-50s-177). The electrode was lowered into the cortex via a stereotactic instrument to a depth of 1000-1500μm. Following the audiovisual stimulus presentation, electrophysiological data from all 32 channels were filtered between 600 and 6000 Hz, and spikes belonging to single neurons and multi-units were identified in a semi-automated manner using KiloSort2 (Pachitariu et al., 2016).

Audiovisual stimuli

The audiovisual stimuli were generated using MATLAB (MathWorks, USA), and presented to mice on a 12" LCD monitor (Eyoyo) and through a magnetic speaker (Tucker-Davis Technologies) placed to the right of the mouse. The visual stimulus was generated

using the PsychToolBox package for MATLAB and consisted of square wave drifting gratings 1 s in duration, 4-Hz temporal frequency, and 0.1 cycles/°. The gratings moved in 12 directions, evenly spaced 0°-360°, and were scaled to a range of 5 different visual contrast levels (0, 0.25, 0.5, 0.75, 1), totaling 60 unique visual stimuli. The auditory stimulus was sampled at 400 kHz and consisted of a 1 s burst of 70 dB white noise. The visual grating was accompanied by the auditory noise on half of trials (120 unique trial types, 10 repeats each), with simultaneous onset and offset. The auditory-only condition corresponded to the trials with a visual contrast of 0. The trial order was randomized and was different for each recording. A second set of audiovisual stimuli consisted of a looming or receding black dot on a gray background, with accompanying auditory white noise that increased or decreased in intensity. The dot went grew by 1400%, and the white noise grew by 40 dB. This was supplemented with trials of static visual dot and white noise of intermediate size and intensity. There were 15 trials of each condition.

Optogenetic stimulation and suppression

Optogenetic stimulation of ChR2 was performed using a blue 473nm laser (BL473T3-150). The laser was administered through an optic cannula, and was pulsed at 40 Hz with 50% duty cycle.

Data analysis and statistical procedures

Spiking data from each recorded unit was organized by trial type and aligned to the trial onset. The number of spikes during each trial was input into a 2-way ANOVA to determine main effects of the auditory and visual stimulus, or auditory and laser stimulation, as well

106

as any interaction terms. Neurons that were responsive to both light and sound or had a significant light-sound interaction term were classified as "light-responsive sound-modulated."

Statistics

Figure data are displayed as means with standard error of the mean (SEM), unless otherwise noted. Shapiro-Wilk tests were used to assess normality, and the statistical tests performed are indicated in the text, figures, and Table 1. For multi-group and multivariate analysis (e.g., ANOVA and Kruskal-Wallis tests) in which a significant ($p<0.05$) interaction was detected, we subsequently performed a post hoc Bonferroni-corrected test.

CHAPTER 5: CONCLUSIONS

Organisms use their various sensory systems to receive information about the external world, filter and process this information, and generate appropriate behavioral responses. Multisensory integration is an important component of sensory processing in humans and other animals, improving accuracy of the sensory modalities and contributing to the generation of a smooth and unified sensory percept. In this dissertation, we explored studies that investigated the neuronal mechanisms of audiovisual integration. Chapter 2 focused on the codes that underlie audiovisual integration in the primary visual cortex (V1). We found that sound enhanced both the magnitude and timing of visual responses of individual neurons. These changes in activity led to improved neuronal encoding of the visual stimulus, specifically through the interaction between response magnitude and trial-by-trial coefficient of variation (CV). Furthermore, the study clarified differential roles of sound and movement on visual processing, finding that sound enhanced the magnitude of the onset response, whereas sound-induced movement primarily modulated the sustained portion of the response.

This series of experiments and analyses performed in V1 demonstrate an important coding mechanism by which audiovisual integration is mediated – improved signal-to-noise ratio (SNR) on audiovisual trials relative to visual trials. This improvement in SNR resulted from a natural reduction in CV associated with increased response magnitude, a relationship that derived directly from Poisson-like spiking. This relationship may not exist

for all neurons, e.g. those that display burst-type spiking or stimulus-evoked suppression of spiking activity. However, the relationship was present in the pool of neurons that we recorded from using electrophysiology in V1, in which the majority of visual and audiovisual responses were excitatory. It will therefore be important in future studies to explore how this coding scheme applies to other instances of audiovisual and multisensory integration.

Chapter 3 explored audiovisual integration in V1 from a circuit perspective, with particular focus on the anatomic and functional connection between the auditory cortex (AC) and V1. The study found excitatory connections from AC to V1 that when optogenetically stimulated enhanced visual responses in V1 neurons. However, suppression of these AC projection neurons caused a minimal or paradoxically slightly excitatory effect on V1 visual responses. We proposed disinhibition of superficial inhibitory neurons in V1, or alternatively circuits mediated by the superior and inferior colliculi (SC and IC, respectively), as potential explanations for these findings. And finally, Chapter 4 highlighted subcortical research that studied the relationship between the SC and IC. We labeled SC projections to the IC and found that they terminated in the IC external shell, and optogenetic stimulation of this pathway resulted in enhanced sound-evoked activity in the IC. However, IC neurons failed to exhibit sensitivity to visual drifting gratings and looming stimuli, suggesting specificity in the visual tuning and ethological function of this intercollicular circuit.

The experimental techniques used to virally trace, histologically characterize, and optogentically probe these circuits all have spatial scale ranging from individual to local populations of neurons. This level of analysis allowed us to directly visualize these neurons

109

and describe them as excitatory or inhibitory. And probing these circuits at this level revealed how individual neurons respond to cross-modal stimuli and cross-modal neuronal activity from distant regions. It was the study of the relationship between individual neuronal activity and population activity that enabled improved understanding of the codes that underlie audiovisual integration. However, other experimental techniques with larger spatial scales exist. Diffusion tensor imaging (DTI) is a useful, non-invasive technology that uses the polarity of water molecules to map anatomic connections between regions throughout the entire brain with millimeter resolution. Functional magnetic resonance imaging (fMRI) similarly uses the polarity of hemoglobin molecules to detect changes in blood flow as a proxy for local neuronal activity throughout the brain, similarly at millimeter resolution. The lower spatial resolution of these techniques precludes them from being used to understanding how groups of neurons represent and communicate information, however they are useful for demonstrating which brain regions act together to mediate behaviors and thought patterns and how the brain state affects one's ability to perceive and respond to sensory stimuli (Jones and Callan, 2003; Kaposvári et al., 2015). And the non-invasive nature of these approaches is ideal for studying audiovisual and multisensory integration in humans. Transcranial magnetic stimulation (TMS) is a non-invasive way to exogenously perturb neuronal activity that has already been used to demonstrate causality between activity in certain brain regions and behavioral and sensory experiences (Hamilton et al., 2013).

At the other end of the spatial scale, patch clamp and whole cell recording reveals how an individual neuron responds to sensory input and afferent neuronal activity. This and related techniques are useful for understanding how a neuron's spiking activity results

110

from biochemical impulses and intracellular compartmentalization of incoming signals. At this scale, it becomes clear the physical mechanisms by which an individual neuron represents sensory information and has been used in audiovisual studies (Ibrahim et al., 2016). We chose to focus on neuronal activity and sensory responses at an intermediate spatial scale, and the studies outlined above contributes to the field from that perspective. However, it is essential for future studies to continue to explore audiovisual and multisensory integration at various spatial scales to provide us with the broadest understanding of how this sensory processing is neuronally mediated. Furthermore, it will be beneficial for studies to correlate experiments at different spatial scales, allowing the most in depth understanding of how chemical and biochemical reactions lead to complex behaviors and perceptual experiences.

Attention, brain state, and multisensory integration

The studies described in this dissertation focus on audiovisual integration without any trained association between the auditory and visual stimuli, nor between the audiovisual stimulus and a behavioral reward. We chose this experimental approach to build a framework and foundational understanding of audiovisual integration for future research in our lab and for others in the field. It would be inaccurate to describe this as a "behaviorally naïve" state, since the mice were able to process and learn audiovisual and multisensory associations from their wakeful lives outside of the experimental setup. Nevertheless, specifically teaching these mice an experimentally designed, parameterized association between the audiovisual stimulus and an external reward would broaden the scope of these audiovisual studies.

The use of external rewards, such as water in water deprived subjects or sugar treats, is commonly used to promote animal subjects' attention to the sensory stimulus or increase the salience of a particular component of the stimulus. Attention to sensory stimuli tends to increase neuronal acuity and behavioral performance both in humans and animal models (Remijn and Kojima, 2010; Schröger et al., 2015). And reward-based learning of sensory associations shifts the neuronal representation of relevant stimulus features (Weinberger, 2004; Wood et al., 2020). The use of behavioral readout of task learning, such as spout licking and wheel turning, in animal models can even paradoxically reduce performance on cognitive tasks (Kuchibhotla, 2019). Additionally, sensory processing is modulated and suppressed under anesthesia, which entails broad pharmacological manipulation of brain activity. How these task-oriented and brain state-centered experimental designs affect multisensory processing has been studied (Aggarwal et al., 2019; Zuanazzi and Noppeney, 2020). However, more research will need to be done to fully understand when a cross-sensory input is beneficial for cognition versus when it can be distracting from attention elsewhere (Dean et al., 2017), and the neuronal underpinnings of these contextual differences.

The studies outlined in this dissertation, specifically in Chapter 2, constructed neurometric curves of stimulus encoding accuracy. Future studies can build on this approach by pairing the audiovisual stimulus with training of the mouse subjects to behaviorally report the perceived visual stimulus direction. This expansion to the experimental setup would allow one to correlate neurometric accuracy with behavioral performance, an analysis that would directly demonstrate the relationship between neuronal representation of the stimulus and the sensory perception. Pairing the audiovisual

stimulus with a task is also relevant to how brain regions communicate with each other, explored in Chapters 3 and 4. Connections between the AC and V1 are involved in audiovisual cue learning (Garner and Keller, 2020), and AC input is particularly beneficial during tactile goal-directed behavior (Godenzini et al., 2021). Further experiments using multisensory behavioral tasks would allow mapping how the functional relationship between brain areas shifts during learning and behavioral performance. And the discrepancy between the roles of cortical and subcortical circuits in this context could demonstrate differences in the ethological and evolutionary origins and pressures driving this neuronal processing. Furthermore, sensory integration and global access to shared information is a proposed prerequisite of consciousness, so comparing multisensory integration under awake and anesthetized conditions can improve not only our understanding of how anesthetics work, but also our understanding of perceptual and conscious experience more broadly.

Audiovisual and multisensory integration in disease

Thus far, the research described in this dissertation has investigated a basic neuroscientific question – how neurons and brain regions represent audiovisual information. However, it is also important to consider the translational and clinical relevance of these studies. Sensory and multisensory processing is a brain function that is often impaired in various types of neurological and neuropsychiatric disorders. People with autism spectrum disorders (ASD) differ in the behavioral responses to sensory and multisensory stimuli (Kwakye et al., 2011; Woynaroski et al., 2013) and it has been shown these differences in sensory responses correlate with symptom severity and disease

presence (Feldman et al., 2019). The neuronal mechanisms of this sensory sensitivity and dysregulation in ASD are starting to be revealed (Green et al., 2013). Additionally, sensory processing is affected in cases of sensory loss such as congenital deafness or blindness, and multisensory integration may still be impaired once sensation is restored (Nava et al, 2014; Hauthal et al., 2014). However, it is also in people with sensory loss that we observe heightened acuity in other sensory modalities (Gougoux et al., 2005; Huber et al., 2019). Studying the neuronal mechanisms of multisensory integration at a basic neuroscience level can improve our understanding of the sensory processing deficits and differences in these clinical disorders and develop improved treatment strategies.

For the past several centuries, medicine has largely used a molecular approach to modify the biological activity of cells, tissues, and organs. Progress in recent decades, though, indicates that the future of medicine may lay with more cellular-based therapies. For example, CAR T-cell therapy is an innovative strategy for treating cancer using T-cells derived from the patient themselves, as opposed to the molecular approach of chemotherapy. Neurological treatments are likely headed in a similar direction, with research being performed on how to effectively transplant neuronal tissue to restore function in damaged brain areas. Nevertheless, these approaches are geared towards maintaining and restoring the physical and biophysiological integrity of neural tissue, providing a healthy environment for neurons to process information. Therefore, the studies included in this dissertation can aid in this endeavor by describing how neurons represent sensory and multisensory information. This basic neuroscience research will guide translational and clinical studies by providing neuronal metrics and proxies used to developing prognostics, screening tools, and therapies for neurological disorders.

REFERENCES

Aggarwal, A., Brennan, C., Shortal, B., Contreras, D., Kelz, M.B., Proekt, A. (2019) Coherence of visual-evoked gamma oscillations is disrupted by propofol but preserved under equipotent doses of isoflurane. *Front Syst Neurosci*, 13:19. Doi: 10.3389/fnsys.2019.00019

Ahmadlou, M., Zwifel, L.S., Heimel, J.A. (2018) Functional modulation of primary visual cortex by the superior colliculus in the mouse. *Nat Commun*, 9(1):3895. Doi: 10.1038/s41467-018-06389-6

Arnold, P., Köpsel, A. (1996) Lipreading, reading and memory of hearing and hearing-impaired children. *Scand Audiol*, 25(1):13-20. Doi: 10.3109/01050399609047550

Atilgan, H., Town, S. M., Wood, K. C., Jones, G. P., Maddox, R. K., Lee, A., & Bizley, J. K. (2018). Integration of Visual Information in Auditory Cortex Promotes Auditory Scene Analysis through Multisensory Binding. *Neuron*, *97*(3), 640–655.e4. https://doi.org/10.1016/j.neuron.2017.12.034

Aitkin, L.M., Dickhaus, H., Schult, W., Zimmermann, M. (1978) External nucleus of inferior colliculus: auditory and spinal somatosensory afferents and their interactions. *J Neurophysiol*, 41(4):837-47. Doi: 10.1152/jn.1978.41.4.837.

Bergan, J.F., Knudsen, E.I. (2009). Visual modulation of auditory responses in the owl inferior colliculus. *J Neurophysiol*, 101(6):2924-33. Doi: 10.1152/jn.91313.2008

Berens, P., Ecker, A.S., Cotton, R.J., Ma, W.J., Bethge, M., Tolias, A.S. (2012) A fast and simple population code for orientation in primate V1. *J Neurosci*, 32(31): 10618-26. doi: 10.1523/JNEUROSCI.1335-12.2012

Bigelow, J., Poremba, A. (2016) Audiovisual integration improves monkeys' short-term memory. *Anim Cogn*, 19(4): 799-811. Doi: 10.1007/s10071-016-0979-0

Bigelow, J., Morrill, R.J., Dekloe, J., Hasenstaub, A.R. (2019) Movement and VIP interneuron activation differentially modulate encoding in mouse auditory cortex. *eNeuro*, 6(5); ENEURO.0164-19.2019. doi: 10.1523/ENEURO.0164-19.2019

Bimbard, C., Sit, T.P.H., Lebedeva, A., Harris, K.D., Carandini, M. (2021) Behavioral origin of sound-evoked activity in visual cortex. BioRXiv. Doi: https://doi.org/10.1101/2021.07.01.450721

Borst, A., Theunissen, F.E. (1999) Information theory and neural coding. *Nat Neurosci*, 2(11):947-57. Doi: doi: 10.1038/14731

Brainard, M.S. and Knudsen, E.I. (1993) Experience-dependent plasticity in the inferior colliculus: a site for visual calibration of the neural representation of auditory space in the barn owl. *Neurosci*, 13(11):4589-608. Doi: 10.1523/JNEUROSCI.13-11-04589

Brainard, M.S. and Knudsen, E.I. (1995) Dynamics of visually guided auditory plasticity in the optic tectum of the barn owl. *J Neurophysiol*, 73(2):595-614. Doi: 10.1152/jn.1995.73.2.595.

Carr, C.E., Christensen-Dalsgaar, J. (2015) Sound localization strategies in three predators. *Brain Behav Evol*, 86(1):17-27. Doi: 10.1159/000435946

Churchland, A.K., Kiani, R., Chaudhuri, R., Wang, X., Pouget, A., Shadlen, M.N. (2011) Variance as a signature of neural computations during decision making. *Neuron*, 69(4):818-31. Doi: 10.1016/j.neuron.2010.12.037

Cohen, L., Rothschild, G., Mizrahi, A. (2011). Multisensory integration of natural odors and sounds in the auditory cortex. *Neuron*, 72(2): 357-69. Doi: 10.1016/j.neuron.2011.08.019

Colonius, H., & Diederich, A. (2017). Measuring multisensory integration: from reaction times to spike counts. *Scientific reports*, 7(1), 3023. https://doi.org/10.1038/s41598-017-03219-5

Cooper, M.H. and Young, P.A. (1976) Cortical projections to the inferior colliculus of the cat. *Exp Neurol*, 51(2):488-502. Doi: 10.1016/0014-4886(76)90272-7

Crosse, M. J., Butler, J. S., & Lalor, E. C. (2015). Congruent Visual Speech Enhances Cortical Entrainment to Continuous Auditory Speech in Noise-Free Conditions. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, 35(42), 14195–14204. https://doi.org/10.1523/JNEUROSCI.1829-15.2015

Dardalat M.C., Stryker M.P. (2017) Locomotion enhances neural encoding of visual stimuli in mouse V1. *J Neurosci*, 37(14):3764-75. Doi: 10.1523/JNEUROSCI.2728-16.2017

Dean, C.L., Eggleston, B.A., Gibney, K.D., Aligbe, E., Blackwell, M., Kwakye, L.D. (2017) Auditory and visual distractors disrupt multisensory temporal acuity in the crossmodal temporal order judgment task. *PLoS One*, 12(7): e0179564. Doi: 10.1371/journal.pone.0179564

Denervaud, S., Gentaz, E., Matusz, P.J., Murray, M.M. (2020) Multisensory gains in simple detection predict global cognition in schoolchildren. *Sci Rep*, 10(1):1394. Doi: 10.1038/s41598-020-58329-4.

Deneux T., Harrell E.R., Kempf A., Ceballo S., Filipchuk A., Bathellier B. (2019) Context-dependent signaling of coincident auditory and visual events in primary visual cortex. eLife 8: e44006. doi: 10.7554/eLife.44006

Denison, R. N., Driver, J., & Ruff, C. C. (2013). Temporal structure and complexity affect audio-visual correspondence detection. *Frontiers in psychology*, *3*, 619. https://doi.org/10.3389/fpsyg.2012.00619

Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. *Perception & psychophysics*, *66*(8), 1388–1404. https://doi.org/10.3758/bf03195006

Dräger, U.C., Hubel, D.H. (1975) Responses to visual stimulation and relationship between visual, auditory, and somatosensory inputs in mouse superior colliculus. *J Neurophysiol*, 38(3):690-713. Doi: 10.1152/jn.1975.38.3.690

Fahey, P.G., Muhammad, T., Smoth, C., Froudarakis, E., Cobos, E., Fu, J., Walker, E.Y., Yatsenko, D., Sinz, F.H., Reimer, J., Tolias, A.S. (2019). A global map of orientation tuning in mouse visual cortex. BioRXiv.

Falchier, A., Clavagnier, S., Barone, P., Kennedy, H. (2002) Anatomical evidence of multimodal integration in primate striate cortex. *J Neurosci*, 22(13): 5749-59. Doi: 10.1523/JNEUROSCI.22-13-05749.2002.

Feldman, D.E., Knudsen, E.I. (1997) An anatomical basis for visual calibration of the auditory space map in the barn owl's midbrain. *J Neurosci*, 17(17)6820-6837. Doi: https://doi.org/10.1523/JNEUROSCI.17-17-06820.1997

Feldman, J.I., Kuang, W., Conrad, J.G., Tu, A., Santapuram, P., Simon, D.M. Foss-Feig, J.H., Kwakye, L.D., Stevenson, R.A., Wallace, M.T., Woynaroski, T.G. (2019) Brief report: Difference in multisensory integration covary with sensory responsiveness in children with and without autism spectrum disorder. *J Autism Dev Disord*, 49(1):397-403. Doi: 10.1007/s10803-018-3667-x

Forbes, J., Cronovich, H. (2020). Romberg Test. StatPearls. PMID: 3308533.

Fujisaki, W., Goda, N., Motoyoshi, I., Komatsu, H., Nishida, S. (2014) Audiovisual integration in the human perception of materials. *J Vis*, 14(4):12. doi: 10.1167/14.4.12

Garner, A.R. and Keller, G.B. (2020) A cortical circuit for audio-visual predictions. *BioRXiv*. Doi: https://doi.org/10.1101/2020.11.15.383471.

Gingras, G., Rowland, B. A., & Stein, B. E. (2009). The differing impact of multisensory and unisensory integration on behavior. *The Journal of neuroscience : the*

official journal of the Society for Neuroscience, *29*(15), 4897–4902. https://doi.org/10.1523/JNEUROSCI.4120-08.2009

Gleiss, S., & Kayser, C. (2012). Audio-visual detection benefits in the rat. *PloS one*, *7*(9), e45677. https://doi.org/10.1371/journal.pone.0045677

Godenzini, L., Alwis, D., Guzulaitis, R., Honnuraiah, S., Stuart, G.J., Palmer, L.M. (2021) Auditory input enhances somatosensory encoding and tactile goal-directed behavior. *Nat Commun*, 12(1):4509. Doi: 10.1038/s41467-021-24754-w

Gougoux, F., Zatorre, R., Lassonde, M., Voss, P., Lepore, F. (2005) A functional neuroimaging study of sound localization: visual cortex activity predicts performance in early-blind individuals. *PLoS Biol*, 3(2):e27. Doi: 10.1371/journal.pbio.0030027

Green, S.A., Rudi, J.D., Colich, N.L., Wood, J.J., Shirinvan, D., Hernandez, L., Tottenham, N., Dapretto, M., Bookheimer, S.Y. (2013) Overreactive brain responses to sensory stimuli in youth with autism spectrum disorders. *J Am Acad Child Adolesc Psychiatry*, 52(11):1148-72. Doi: 10.1016/j.jaac.2013.08.004

Gruters, K.G., Groh, J.M. (2012) Sounds and beyond: multisensory and other non-auditory signals in the inferior colliculus. *Front Neural Circuits*, 6:96. Doi: 10.3389/fncir.2012.00096

Gur, M., Beylin, A., Snodderly, D.M. (1997) Response variability of neurons in primary visual cortex (V1) of alert monkeys. *J Neurosci*, 17(8):2914-20. doi: 10.1523/JNEUROSCI.17-08-02914.1997

Hamilton, R.H., Wiener, M., Drebing, D.E., Coslett, H.B. (2013) Gone in a flash, manipulation of audiovisual temporal integration using transcranial magnetic stimulation. *Front Psychol*, 4: 571. Doi: 10.3389/fpsyg.2013.00571

Hammond-Kenny, A., Bajo, V. M., King, A. J., & Nodal, F. R. (2017). Behavioural benefits of multisensory processing in ferrets. *The European journal of neuroscience*, *45*(2), 278–289. https://doi.org/10.1111/ejn.13440

Hauthal, N., Debener, S., Rach, S., Sandmann, P., Thorne, J.D. (2014). Visuo-tactile interactions in the congenitally deaf: A behavioral and event-related potential study. *Front Integr Neurosci*, 8:98. Doi: 10.3389/fnint.2014.00098

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*(9), e25283. https://doi.org/10.1371/journal.pone.0025283

Huber, E., Chang, K., Alvarez, I., Hundle, A., Bridge, H., Fine, I. (2019) Early blindness shapes cortical representations of auditory frequency within auditory cortex. *J Neurosci*, 39(26):5143-52. Doi: 10.1523/JNEUROSCI.2896-18.2019

Ibrahim, L. A., Mesik, L., Ji, X. Y., Fang, Q., Li, H. F., Li, Y. T., Zingg, B., Zhang, L. I., & Tao, H. W. (2016). Cross-Modality Sharpening of Visual Cortical Processing through Layer-1-Mediated Inhibition and Disinhibition. *Neuron*, *89*(5), 1031–1045. https://doi.org/10.1016/j.neuron.2016.01.027

Iurilli, G., Ghezzi, D.,Olcese, U., Lassi, G., Nozzaro, C., Tonini, R., Tucci, V., Benfenati, F., Medini, P. (2012) Sound-driven synaptic inhibition in primary visual cortex. *Neuron*, 73(4): 814-28. Doi: 10.1016/j.neuron.2011.12.026

Izumi, A. and Kojima, S. (2004) Matching vocalizations to vocalizing faces in a chimpanzee (Pan troglodytes). *Animal Cog*, 7(3): 179-184. https://doi.org/10.1007/s10071-004-0212-4

Jones, J. A., Callan, D.E. (2003) Brain activity during audiovisual speech perception: an fMRI study of the McGurk effect. *Neuroreport*, 14(8):1129-33. Doi: 10.1097/00001756-200306110-00006

Kaposvári, P., Csete, G., Bognár, A., Csibri, P., Tóth, E., Szabó, N., Vécsei, L., Sáry, G., Kincses, Z.T. (2015) Audio-visual integration through parallel visual pathways. *Brain Res*, 1624:71-77. Doi: 10.1016/j.brainres.2015.06.036

King, A.J., Schnupp, J.W.H., Thompson, I.D. (1998) Signals from the superficial layers of the superior colliculus enable the development of the auditory space map in the deeper layers. *J Neurosci*, 18(22) 9394-408. Doi: https://doi.org/10.1523/JNEUROSCI.18-22-09394.1998

Knöpfel, T., Sweeney, Y., Radulescu, C. I., Zabouri, N., Doostdar, N., Clopath, C., & Barnes, S. J. (2019). Audio-visual experience strengthens multisensory assemblies in adult mouse visual cortex. *Nature communications*, *10*(1), 5684. https://doi.org/10.1038/s41467-019-13607-2

Knudsen, E.I. (1985) Experience alters the spatial tuning of auditory units in the optic tectum during a sensitive period in the barn owl. *J Neurosci*, 5(11):3094-109. Doi: 10.1523/JNEUROSCI.05-11-03094

Kreifelts, B., Ethofer, T., Grodd, W., Erb, M., Wildgruber, D. (2007) Audiovisual integration of emotional signals in voice and face: an event-related fMRI study. *Neuroimage*, 37(4):1445-56. Doi: 10.1016/j.neuroimage.2007.06.020

Kuchibhotla, K.V., Sten, T.H. Papadovannis, E.S., Elnozahy, S., Fogelson, K.A., Kumar, R., Boubenec, Y., Holland, P.C., Ostoiic, S., Froemke, R. (2019) Dissociating task acquisition from expression during learning reveals latent knowledge. *Nat Commun*, 10(1):2151. Doi: 10.1038/s41467-019-10089-0

Kwakye, L.D., Foss-Feig, J.H., Cascio. C.J., Stone, W.L., Wallace, M.T. (2011) Altered auditory and multisensory temporal processing in autism spectrum disorders. *Front INtegr Neurosci*, 4:129. Doi: 10.3389/fnint.2010.00129

Laurienti, P.J., Kraft, R.A., Maldjian, J.A., Burdette, J.H., Wallace, M.T. (2004) Semantic congruence is a critical factor in multisensory behavioral performance. *Exp Brain Res*, 158(4):405-14. Doi: 10.1007/s00221-004-1913-2

Lee, A.M., Hoy, J.L., Bonci, A., Wilbrecht, L., Stryker, M.P., Niell, C.M. (2014) Identification of a brainstem circuit regulating visual cortical state in parallel with locomotion. *Neuron*, 83(2):455-466. Doi: 10.1016/j.neuron.2014.06.031

Leong, A.T.L., Dong, C.M., Gao, P.P., Chan, R.W., To, A., Sanes, D.H., Wu, E.X. (2018). Optogenetic auditory fMRI reveals the effects of visual cortical inputs on auditory midbrain response. *Sci Rep*, 8. Doi: https://doi.org/10.1038/s41598-018-26568-1

Lesicko, A.M.H., Hristova, T.S., Maigler, K.C., Llano, D.A. (2016) Connectional modulatory of top-down and bottom-up multimodal inputs to the lateral cortex of the mouse inferior colliculus. *J Neurosci*, 36(43):11037-50. Doi: 10.1523/JNEUROSCI.4134-15.2016

Lesicko, A.M.H., Sons, S.K., Llano, D.A. (2020) Circuit mechanisms underlying the segregation and integration of parallel processing streams in the inferior colliculus. *J Neurosci*, 40(33):6328-44. Doi: https://doi.org/10.1523/JNEUROSCI.0646-20.202

Maddox, R. K., Atilgan, H., Bizley, J. K., & Lee, A. K. (2015). Auditory selective attention is enhanced by a task-irrelevant temporally coherent visual stimulus in human listeners. *eLife*, *4*, e04995. https://doi.org/10.7554/eLife.04995

Marlin, B.J., Mitre, M., D'amour, J.A., Chao, M.V., Froemke, R.C. (2015). Oxytocin enables maternal behavior by balancing cortical inhibition. *Nature*, 520(7528): 499-504. doi: 10.1038/nature14402

Marucci, M., Flumeri, G.D., Borghini, G., Sciaraffa, N., Scandola, M., Pavone, E.F., Babiloni, F., Betti, V., Arco, P. (2021) The impact of multisensory integration and perceptual load in virtual reality settings on performance, workload and presence. *Sci Rep*, 11(1);4831. Doi: 10.1038/s41598-021-84196-8.

McClure, J. P., Jr, & Polack, P. O. (2019). Pure tones modulate the representation of orientation and direction in the primary visual cortex. *Journal of neurophysiology*, *121*(6), 2202–2214. https://doi.org/10.1152/jn.00069.2019

McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. Nature 264, 746–748. doi: 10.1038/264746a0

Meijer, G. T., Montijn, J. S., Pennartz, C., & Lansink, C. S. (2017). Audiovisual Modulation in Mouse Primary Visual Cortex Depends on Cross-Modal Stimulus Configuration and Congruency. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *37*(36), 8783–8796. https://doi.org/10.1523/JNEUROSCI.0468-17.2017

Meijer, G. T., Pie, J. L., Dolman, T. L., Pennartz, C., & Lansink, C. S. (2018). Audiovisual Integration Enhances Stimulus Detection Performance in Mice. *Frontiers in behavioral neuroscience*, *12*, 231. https://doi.org/10.3389/fnbeh.2018.00231

Meijer, G. T., Mertens, P., Pennartz, C., Olcese, U., & Lansink, C. S. (2019). The circuit architecture of cortical multisensory processing: Distinct functions jointly operating within a common anatomical network. *Progress in neurobiology*, *174*, 1–15. https://doi.org/10.1016/j.pneurobio.2019.01.004

Métin, C., Godement, P., & Imbert, M. (1988). The primary visual cortex in the mouse: receptive field properties and functional organization. *Experimental brain research*, *69*(3), 594–612. https://doi.org/10.1007/BF00247312

Montijn J.S., Vinck M., Pennartz C.M.A. (2014) Population coding in mouse visual cortex: response reliability and dissociability of stimulus tuning and noise correlation. *Front Comput Neurosci* 8:58. Doi: 10.3389/fncom.2014.00058

Morrill, R. J., & Hasenstaub, A. R. (2018). Visual Information Present in Infragranular Layers of Mouse Auditory Cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *38*(11), 2854–2862. https://doi.org/10.1523/JNEUROSCI.3102-17.2018

Musall, S., Kaufman, M.T., Juavinett, A.L., Gluf, S., Churchland, A.K. (2019) Single-trial neural dynamics are dominated by richly varied movements. *Nat Neurosci*, 22(10):1677-1686. Doi: 10.1038/s41593-093-019-0502-4.

Nava, E., Bottari, D., Villwock, A., Fengler, I., Büchner, A., Lenarz, T., Röder, B. (2014) Audio-tactile integration in congenitally and late deaf cochlear implant users. *PLoS One*, 9(6):e99606. Doi: 10.1371/journal.pone.0099606

Nelson A., Schneider D.M., Takatoh J., Sakurai K., Wang F., Mooney R. (2013) A circuit for motor cortical modulation of auditory cortical activity. *J Neurosci*, 33(36):14342-53. Doi: 10.1523/JNEUROSCI.2275-13.2013.

Niell C.M., Stryker M.P. (2010) Modulation of visual responses by behavioral state in mouse visual cortex. *Neuron*, 65(4):472-9. Doi: 10.1016/j.neuron.2010.01.033.

Noesselt, T., Tyll, S., Boehler, C.N., Budinger, E., Heinze, H., Driver, J. (2010) Sound-induced enhancement of low-intensity vision: multisensory influence on human

sensory-specific cortices and thalamic bodies relate to perceptual enhancement of visual detection sensitivity. *J Neurosci*, 30(41):13609-23. Doi: 10.1523/JNEUROSCI.4524-09.2010.

Okabe, S., Nagasawa, M., Kihara, T., Kato, M., Harada, T., Koshida, N., Mogi, K., Kikusui, T. (2013) Pup odor and ultrasonic vocalizations synergistically stimulate maternal attention in mice. *Behav Neurosci*, 127(3), 432-438. Hhttps://doi.org/10.1037/a0032395

Pagan, M., Urban, L.S., Wohl, M.P., Rust, N.C. (2013) Signals in inferotemporal and perirhinal cortex suggest an untangling of visual target information. *Nat Neurosci*, 16(8):1132-9. Doi: 10.1038/nn.3433

Quak, M., London, R. E., & Talsma, D. (2015). A multisensory perspective of working memory. *Frontiers in human neuroscience*, *9*, 197. https://doi.org/10.3389/fnhum.2015.00197

Pachitariu M., Steinmetz N., Kadir S., Carandini M., Harris K.D. (2016) Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels. *BioRχiv*. Doi: https://doi.org/10.1101/061481

Remijn, G.B., Kojima, H. (2010) Active versus passive listening to auditory streaming stimuli: a near-infrared spectroscopy study. *J Biomed Opt*, 15(3):037006. Doi: 10.1117/1.3431104

Rochefort, N. L., Narushima, M., Grienberger, C., Marandi, N., Hill, D. N., & Konnerth, A. (2011). Development of direction selectivity in mouse cortical neurons. *Neuron*, *71*(3), 425–432. https://doi.org/10.1016/j.neuron.2011.06.013

Schneider D.M., Mooney R. (2018) How movement modulates hearing. *Annu Rev Neurosci*, 41:553-72. Doi: 10.1146/annurev-neuro-072116-031215.

Schöger, E., Marzecová, A., SanMiguel, I. (2015) Attention and prediction in human audition: a lesion from cognitive psychophysiology. *Eur J Neurosci*, 41(5):641-64. Doi: 10.1111/ejn.12816

Shamash, P., Carandini, M., Harris, K. (2018) A tool for analyzing electrode tracks from slice histology. *BioRXiv*. Doi: 10.1101/447995

Shams L, Kamitani Y, Shimojo S (2002) Visual illusion induced by sound. *Brain Res Cogn Brain Res* 14(1):147-52. Doi: 10.1016/s0926-6410(02)00069-1

Shik, M.L., Severin, F.V., Orlovskii (1966) Control of walking and running by means of electrical stimulation of the midbrain. *Biofizika*, 11(4):659:66.

Sparks, D., Rohrer, W.H., Zhang, Y. (2000) The role of the superior colliculus in saccade initiation: a study of express saccades and the gap effect. *Vision Res*, 40(20):2763-77. Doi: 10.1016/s0042-6989(00)00133-4

Stanislaw, H., Todorov, N. (1999) Calculation of signal detection theory measures. *Behav Res Methods Instrum Comput*. 31(1):137-49. Doi: 10.3758/bf03207704

Stein, B. E., Stanford, T. R., & Rowland, B. A. (2020). Multisensory Integration and the Society for Neuroscience: Then and Now. *J Neurosci*, *40*(1), 3–11. https://doi.org/10.1523/JNEUROSCI.0737-19.2019

Stringer, C., Michaelos, M., Tsyboulski, D., Lindo, S.E., Pachitariu, M. (2021) High-precision coding in visual cortex. *Cell*, 184(10);2767-78. Doi: 10.1016/j.cell.2021.03.042

Syka, J. and Radil-Weiss, T. (1973) Acoustical responses of inferior colliculus neurons in rats influenced by sciatic nerve stimulation and light flashes. *Int J Neurosci*, 5(5):201-6. Doi: 10.3109/00207457309149476.

Tye-Murray, N., Spehar, B., Myerson, J., Hale, S., & Sommers, M. (2016). Lipreading and audiovisual speech recognition across the adult lifespan: Implications for audiovisual integration. *Psychology and aging*, *31*(4), 380–389. https://doi.org/10.1037/pag0000094

Tyll, S., Budinger, E., Noesselt, T. (2011) Thalamic influences on multisensory integration. *Commun Integr. Biol*, 4(4): 378-381. Doi: 10.4161/cib.4.4.15222

von Trapp, G., Buran, B. N., Sen, K., Semple, M. N., & Sanes, D. H. (2016). A Decline in Response Variability Improves Neural Signal Detection during Auditory Task Performance. *J Neurosci*, *36*(43), 11097–11106. https://doi.org/10.1523/JNEUROSCI.1302-16.2016

Wang, Y., Celebrini, S., Trotter, Y., Barone, P. (2008) Visuo-auditory interactions in the primary visual cortex of the behaving monkey: electrophysiological evidence. *BMC Neurosci*, 9:79. Doi: 10.1186/1471-2202-9-79

Weinberger, N.M. (2004) Specific long-term memory traces in primary auditory cortex. *Nat Rev Neurosci*, 5(4):279-90. Doi: 10.1038/nrn1366

Williams A.M., Angeloni C.F., Geffen M.N. (2021) Sound improves neuronal encoding of visual stimuli in mouse primary visual cortex. *BioRXiv*, doi: 10.1101/2021.08.03.454738

Wood, K.C., Angeloni, C.F., Oxman, K., Clopath, C., Geffen, M.N. (2020) Neuronal activity in sensory cortex predicts the specificity of learning. *BioRXiv*. Doi: https://doi.org/10.1101/2020.06.02.128702

Woynaroski, T.G., Kwakye, L.D., Foss-Feig, J.H., Stevenson, R.A., Stone, W.L., Wallace, M.T. (2013) Multisensroy speech perception in children with autism spectrum disorders. *J Autism Dev Disord*, 43(12):2891-902. Doi: 10.1007/s10803-013-1836-5

Yilmax, M., Meister, M. (2013) Rapid innate defensive responses of mice to looming visual stimuli. *Curr Biol*, 23(20):2011-15. Doi: 10.1016/j.cub.2013.08.015

Zhao X., Chen H., Liu X., Cang J. (2013) Orientation-selective resposnes in the mouse lateral geniculate nucleus. *J Neurosci* 33(31):12751-763. Doi: 10.1523/JNEUROSCI.0095-13.2013

Zingg, B., Chou, X., Zhang, Z., Mesik, L., Liang, F., Tao, H.W., Zhang, L.I. (2017) AAV-mediated anterograde transsynaptic tagging: Mapping corticocollicular input-defined neural pathways for defense behaviors. *Neuron*, 93(1):33-47. Doi: 10.1016/j.neuron.2016.11.045

Zingg, B., Peng, B., Huang, J., Tao, H.W., Zhang, L.I. (2020) Synaptic specificity and application of anterograde transsynaptic AAV for probing neural circuitry. *J Neurosci*, 40(16):3250-67. Doi: 10.1523/JNEUROSCI.2158-19.2020

Zuanazzi, A., Noppeney, U. (2020) The intricate interplay of spatial attention and expectation: A multisensory perspective. *Multisens Res*, 33(4-5):383-416. Doi: 10.1163/22134808-20201482