University College London

Faculty of Brain Sciences

Ear Institute

Combining Sensory Information in the Mammalian Temporal Lobe: Approaching realism in audiovisual integration, spatial perception, and comparative neurophysiology

A Thesis Presented for the Degree of Doctor of Philosophy

By

Amit Khandhadia

October 18, 2023

Declaration

I, Amit Khandhadia, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signature:

Date: October 18, 2023

Abstract

The temporal lobe of the mammalian brain receives and processes information from the visual and auditory system. In primates, the inferior temporal lobe is particularly important for the processing of high-level visual form information including regions, which respond specifically to faces, termed face patches. However, some of these face patches sit in the superior temporal sulcus (STS), which is a zone of convergence for different kinds of visual and auditory afferent connections. Similarly, visual regions near the auditory cortex in ferrets also contain widespread connections to different kinds of auditory and visual regions. This thesis aims to investigate and further understand the convergence of these sensory inputs and how single neurons in the temporal lobe can combine auditory, visual, social, and spatial information.

Following a table of contents in chapter 1, chapter 2 reviews the literature exploring functions that the temporal cortex subserves across both primates and carnivores including form-based vision, 3D vision, audiovisual integration, and the potential combination of these functions. Chapter 3 describes experiments, which discovered, for the first time, that neurons in an STS face patch showed acoustic modulation of visual signals and even responses to auditory stimulus alone. Chapter 4 focuses on visual processing, which revealed that neurons in the same face patch were tuned not to retinal angle but the physical size of a face, indicating that spatial information combines with form visual information in this region. Chapter 5 combines these two approaches to examine how spatial manipulations influence the audiovisual responses in this region. Chapter 6 then takes a comparative approach detailing experiments that uncovered face and body selective neurons in ferret visual regions. Finally, chapter 7 synthesize these results and details future experiments that further explore the relationship of audiovisual integration and space as well the comparative functions of the temporal lobe.

Impact Statement

How the brain creates a perception of the natural world remains a challenging subject in the field of neuroscience. The brain must combine numerous different senses and different features within those senses to interact with the world. The limits of technology, time, and resources have often required high levels of control in scientific experiments to focus on specific features of sense rather than how neurons integrate multiple features. However, as technology has advanced, stimulus presentation can extend beyond these limits and begin to take small steps towards realism while still maintaining levels of control.

I first examined if a fMRI face-selective region in the macaque temporal cortex showed audiovisual responses usually natural and synthetic stimuli. These regions exist in both macaques and humans but, in macaques, they had not been examined for their role in audiovisual integration. However, we discovered that this specialized region did in fact have neurons that could be modulated by or could respond to acoustic stimuli. These neurons further showed specificity to faces in their audiovisual responses. This discovery highlights this region as having a broader role in communication and expands the fields understanding of audiovisual integration specifically of social stimuli.

I also examined the role of physical size in the tuning of the responses in this region. Most visual studies present stimuli in retinal subtense, the degrees of the surface of the retina occupied by a visual object. But retinal angle is not a feature of our natural three-dimensional world. In order to understand and interact with the natural world, mammals must understand the size of an object or conspecific in metric coordinates i.e the physical size. Using a 3D animatable macaque avatar, we found that neurons in this region do express tuning to physical size rather than retinal angle. This paradigm is a large shift in the way visual neuroscience studies are

conducted, which largely uses retinal angle to determine the tuning of regions across the brain. This experiment also shows that this region of the temporal lobe could combine these different forms of visual information thought to be separated.

I further combined this exploration of spatial information with the exploration of audiovisual responses using a novel 3D dome setup that allowed me to alter the relative position of the auditory and visual stimulus components. This new presentation paradigm unveiled that the spatial positioning of each element could dramatically impact neural spiking revealing space could influence sound and visual processing in complex ways.

Lastly, the final experiment found preliminary evidence of face selectivity in the ferret cortex similar to the face selective neurons found in the macaque. These cells may represent an important evolutionary link between the species and demonstrate that ferrets have face selectivity for the first time. No carnivore species has previously been found to have face selective neurons. If further verified, this discovery could establish the ferret as a potential model for higher order vision, which could enable further exploration of face cells during natural vision.

UCL Research Paper Declaration Form

referencing the doctoral candidate's own published work(s)

Please use this form to declare if parts of your thesis are already available in another format, e.g. if data, text, or figures:

- *have been uploaded to a preprint server*
- *are in submission to a peer-reviewed publication*
- *have been published in a peer-reviewed publication, e.g. journal, textbook.*

This form should be completed as many times as necessary. For instance, if you have seven thesis chapters, two of which containing material that has already been published, you would complete this form twice.

1. **For a research manuscript that has already been published** (if not yet published, please skip to section 2)

   a) **What is the title of the manuscript?**

   Audiovisual integration in macaque face patch neurons

   b) **Please include a link to or doi for the work**

   https://doi.org/10.1016/j.cub.2021.01.102

   c) **Where was the work published?**

   Current Biology

   d) **Who published the work?** (e.g. OUP)

   Cell Press

   e) **When was the work published?**

   10 May, 2021

   f) **List the manuscript's authors in the order they appear on the publication**

   Amit P. Khandhadia, Aidan P. Murphy, Lizabeth M. Romanski, Jennifer K. Bizley, David A. Leopold

   g) **Was the work peer reviewed?**

   Yes

   h) **Have you retained the copyright?**

   The paper was published open access under a creative commons license

   i) **Was an earlier form of the manuscript uploaded to a preprint server?** (e.g. medRxiv). If 'Yes', please give a link or doi)

   No

   If 'No', please seek permission from the relevant publisher and check the box next to the below statement:

   ☒

   *I acknowledge permission of the publisher named under **1d** to include in this thesis portions of the publication named as included in **1c**.*

2. **For a research manuscript prepared for publication but that has not yet been published** (if already published, please skip to section 3)

   a) **What is the current title of the manuscript?**

Click or tap here to enter text.

b) **Has the manuscript been uploaded to a preprint server?** (e.g. medRxiv; if 'Yes', please give a link or doi)

Click or tap here to enter text.

c) **Where is the work intended to be published?** (e.g. journal names)

Click or tap here to enter text.

d) **List the manuscript's authors in the intended authorship order**

Click or tap here to enter text.

e) **Stage of publication** (e.g. in submission)

Click or tap here to enter text.

3. **For multi-authored work, please give a statement of contribution covering all authors** (if single-author, please skip to section 4)

   APK created some stimuli, collected data, analysed results, and wrote the paper. APM helped create stimuli. LMR created stimuli and interpreted results. JKB interpreted data, helped write the paper, and supervised the work. DAL helped write the paper, interpret data, and supervised the work

4. **In which chapter(s) of your thesis can this material be found?**

   Chapter 3

5. **e-Signatures confirming that the information above is accurate** (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work)

Candidate
Date:
18/10/23

Supervisor/ Senior Author (where appropriate)
Date
15.10.23

UCL Research Paper Declaration Form
referencing the doctoral candidate's own published work(s)

Please use this form to declare if parts of your thesis are already available in another format, e.g. if data, text, or figures:

- *have been uploaded to a preprint server*
- *are in submission to a peer-reviewed publication*
- *have been published in a peer-reviewed publication, e.g. journal, textbook.*

This form should be completed as many times as necessary. For instance, if you have seven thesis chapters, two of which containing material that has already been published, you would complete this form twice.

6. For a research manuscript that has already been published (if not yet published, please skip to section 2)

j) **What is the title of the manuscript?**

Encoding of 3D physical dimensions by face selective cortical neurons

k) **Please include a link to or doi for the work**

https://doi.org/10.1073/pnas.2214996120

l) **Where was the work published?**

Proceedings of the National Academy of Sciences (PNAS)

m) **Who published the work?** (e.g. OUP)

NAS

n) **When was the work published?**

February 21, 2023

o) **List the manuscript's authors in the order they appear on the publication**

Amit P. Khandhadia, Aidan P. Murphy, Kenji W. Koyano, Elena W. Esch, David A. Leopold

p) **Was the work peer reviewed?**

Yes

q) **Have you retained the copyright?**

Yes

r) **Was an earlier form of the manuscript uploaded to a preprint server?** (e.g. medRxiv). If 'Yes', please give a link or doi)

No

If 'No', please seek permission from the relevant publisher and check the box next to the below statement:

☒

*I acknowledge permission of the publisher named under **1d** to include in this thesis portions of the publication named as included in **1c**.*

7. For a research manuscript prepared for publication but that has not yet been published (if already published, please skip to section 3)

f) **What is the current title of the manuscript?**

Click or tap here to enter text.

g) **Has the manuscript been uploaded to a preprint server?** (e.g. medRxiv; if 'Yes', please give a link or doi)

Click or tap here to enter text.

h) **Where is the work intended to be published?** (e.g. journal names)

Click or tap here to enter text.

i) **List the manuscript's authors in the intended authorship order**

Click or tap here to enter text.

j) **Stage of publication** (e.g. in submission)

Click or tap here to enter text.

8. **For multi-authored work, please give a statement of contribution covering all authors** (if single-author, please skip to section 4)

A.P.K., A.P.M., and D.A.L. designed research; A.P.K., A.P.M., and E.M.E. performed research; A.P.K., A.P.M., and K.W.K. analyzed data; and A.P.K., A.P.M., K.W.K., and D.A.L. wrote the paper.

9. **In which chapter(s) of your thesis can this material be found?**

Chapter 4

10. **e-Signatures confirming that the information above is accurate** (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work)

Candidate
Date:
18/10/2023

Supervisor/ Senior Author (where appropriate)

Date
15/20/2023

Acknowledgements

I've been extraordinarily lucky to have two wonderful supervisors, Dr. David Leopold and Dr. Jenny Bizley. Thank you to both of you for your faith in me and all the expertise and knowledge you so generously shared. David, thank you for giving me extraordinary freedom to try such new and exciting experiments. I know I could have never had the success I had without the responsibility you let me take on. Jenny, thank you for agreeing to mentor me after only a couple meetings despite a strange plan to bridge research between monkeys and ferrets and for all the patience with my lack of computational skills. I'm so excited to keep working with you. Thank you to the NIMH-UCL program for supporting me through this whole endeavor and even through COVID

I am grateful for all the members of the SCNI, especially Aidan Murphy and Kenji Koyano without whom I could not have done this work. Another thank you to the Bizley lab, specifically Stephen Town and Rebecca Norris, for all the knowledge they've shared helping me get up to speed with ferrets and analyzing complex data. Finally, I want to thank our collaborator Dr. Lizabeth Romanski whose expertise on macaque vocalizations was invaluable

Of course, thank you to all the animals who contributed so much to these projects from the macaques: Spice, Stevie Ray, Matcha, Mochi, Duane, and Wasabi, to the ferrets: Nala, Rajah, Gnocchi, Eclaire, Star Anise, and Cosmo. Thank you also to all the animal care staff at NIH and the RVC

Special thanks to all the friends who had to hear about all my research: Sara, Christa, Elizabeth, Mandy, and Elena W. To my best friends, Aaron and Blane, you both always make me laugh exactly when I need it. Thank you to my family, Muma, Papa, and Neha; you've all

believed in me and supported me so much even when I’m so dramatic. And to Nikki thank you for all the encouragement, support, and love. It means more than I can say.

1. Table of Contents

2. Literature Review

The functional organization of the brain remains a challenging topic. In the temporal lobe, numerous types of sensory information are processed to support perception, but the processing of this information is often studied separately and without attention to how this processing combines to support the functions of the cortex. This thesis presents an investigation of how information combines in the temporal lobe primarily of macaques but also including preliminary data from ferrets. This literature review will first examine the macaque superior temporal sulcus (STS) with a particular focus on the visual specialization for faces and the underexplored role of spatial information. Then it will examine analogous areas of the ferret, particularly area 20b. It will then move to a general discussion of audiovisual integration with some emphasis on the temporal lobe. Finally, this chapter discusses the combination audiovisual information with spatial information and how spatial features can modulate audiovisual responses. Altogether, this thesis will examine the ways the STS and the temporal lobe supports more complex perception of the natural world by integrating audiovisual, social, and spatial information.

2.1 Sensory Perception and the Temporal Lobe

2.1.1 Sensory Functions of the Temporal Lobe

The temporal cortex of the mammalian brain plays a large role in the processing of high-level sensory information for both visual and auditory modalities. In vision, the temporal cortex of primates receives partially processed visual information to further transform in service of high-level form-based vision, particularly in the inferior temporal (IT) cortex (Mishkin, Ungerleider et al. 1983, Kravitz, Saleem et al. 2013). In line with this general function, many neurons and sections of IT show selectivity to numerous specific features including faces, bodies, and colors (Tsao,

Freiwald et al. 2006, Moeller, Freiwald et al. 2008, Popivanov, Jastorff et al. 2012, Lafer-Sousa and Conway 2013, Hesse and Tsao 2020). These responses align with the conventional view of the separation of the visual streams where the dorsal stream subserves spatial information in the parietal regions while the ventral visual stream subserves form and shape information in these temporal regions (Mishkin and Ungerleider 1982, Mishkin, Ungerleider et al. 1983, Goodale and Milner 1992). As visual information proceeds through these various regions of the ventral stream, the information becomes more processed to serve a variety of functions, most prominently to form the core recognition of objects across a variety of positions and changes to lighting (Freiwald and Tsao 2010, Rust and Dicarlo 2010, DiCarlo, Zoccolan et al. 2012). In the realm of audition, the primary auditory cortex and secondary auditory regions of primates all sit in the temporal cortex. Primary auditory cortex receives information from the thalamus and projects to the belt and parabelt regions, secondary auditory regions (Hackett, Stepniewska et al. 1998, Kaas and Hackett 2000). Within these regions, some evidence suggests that these auditory regions organize in a similar way to the visual system, forming two distinct streams, with rostral one dedicated to distinguishing the type of sound and the caudal one dedicated to localizing sounds in space (Romanski, Tian et al. 1999, Rauschecker and Tian 2000, Rauschecker and Scott 2009) (see Figure 2.1A below). Like in the ventral stream of vision, anterior auditory regions contain concentrated regions for voices indicating the formation of circumscribed regions of selectivity may be a principle of brain organization in macaques (Petkov, Kayser et al. 2008). However, despite this organization of the temporal lobe numerous open questions remain of how information intermingles within the temporal lobe.

2.1.2 Form Based Vision and Face Specialization in the Temporal Lobe

The inferior portion of the temporal lobe composes the final portion of the ventral visual stream and serves to process visual information into specific objects and shapes. Lesions of IT cortex in humans and macaques have led to deficits of these functions including the inability to distinguish shapes (Mishkin, Ungerleider et al. 1983). Anatomically, the IT cortex and the STS receive information from the previous parts of the ventral stream, such as V4 (Seltzer and Pandya 1978, Felleman and Van Essen 1991, Ungerleider, Galkin et al. 2008), aggregating more simple visual representations to form percepts of more complicated whole objects. Supporting this organization on the single neuron level, recordings in macaques have found various types of neurons tuned to respond only to specific objects rather than edges or curvatures like in the earlier visual areas (Perrett, Rolls et al. 1982, Desimone, Albright et al. 1984). This tuning, while sometimes view-point dependent, becomes more invariant as it progresses from the posterior to anterior temporal cortex (Rust and Dicarlo 2010). Face processing in the temporal lobe is a clear example of this phenomenon. fMRI in both humans and non-human primates has revealed multiple circumscribed regions within the temporal cortex named for their visual selectivity for certain social stimuli, such as face and body patches (Kanwisher, McDermott et al. 1997, Moeller, Freiwald et al. 2008, Pinsk, Arcaro et al. 2009). In the macaque, face patches contain a high percentage of cells that respond more strongly to visual presentation of faces than to non-face objects and form an interconnected network (Moeller, Freiwald et al. 2008, Grimaldi, Saleem et al. 2016). In fitting with the progression of information towards invariance and more complicated object representations, neurons in earlier face patches can be highly view dependent, only responding to some views of a face but not others. But as neurons arrive at later face patches, neurons become more invariant, responding more equally to all views (Freiwald and

Tsao 2010). This progression is often thought to reflect a shift from early visual coding to object-centered encoding of the world.

2.1.3 Absolute Size and Distance in the Temporal Lobe

One clear pathway that remains unexplored in the temporal lobe is the role of various three-dimensional (3D) features in vision. Many mammals must perceive the world in 3D space to interact with objects and conspecifics. However, the visual system determines spatial perception based on the geometry of the retina, which only maps the world in two-dimensional (2D) space (Hubel and Wiesel 1959, Hubel and Wiesel 1962). To assemble a full 3D perception, the visual system must calculate a new geometry to accommodate the natural world. This transformation happens through the assembly of 2D cues, such as occlusion or parallax, as well as through disparity between the images to both retina called binocular cues (Wheatstone 1962).The integration of various cues to interact with the 3D world has been discovered in numerous species ranging from amphibians to mammals (Nityananda and Read 2017).

The statistics of the world have in turn shaped the mammalian brain adapting its senses to the natural features and geometry. However, many of these natural features are rarely considered in the study of high-level form-based vision. Among these features, physical size, the size of an object in centimeters, and distance from the observer are crucial to understand how to interact with an object or another individual. The judgement of these features is a crucial component of both form-based and spatial vision. These 3D features are rarely explicitly examined in object related vision. Visual experiments performed on animals often restrain and head-fix the subject and present the stimuli on a 2D screen (Leopold and Park 2020). When stimuli are presented in 2D, stimulus size can only be expressed as retinal subtense. This presentation obscures both physical

size and distance as shifts in retinal angle are ambiguous between the two elements. While many areas of earlier vision may largely respond based on retinal angle and depend on this 2D assembly, the percept of objects must develop into a 3D perception for action across the visual system (Srinath, Emonds et al. 2021).

In the primate cortex, studies have predominately assumed that the assembly of this third dimension is served primarily by the dorsal visual stream in the parietal cortex. Conversely, specialized and selective areas in the higher order ventral visual stream often tolerate changes in scale and position (Rust and Dicarlo 2010, DiCarlo, Zoccolan et al. 2012). However, while selectivity may tolerate these spatial transformations, single neurons in many regions associated with the ventral stream and form-based vision often respond uniquely to changes in size or position while maintaining the same overall selectivity (Op De Beeck and Vogels 2000, Freiwald and Tsao 2010). These responses indicate that position information is integrated into some portions of the ventral stream. Even neurons in regions as early as V1 and V2 can show tuning to and dependence on distance without necessarily changing orientation or direction tuning (Dobbins, Jeo et al. 1998). Similarly, V4, a mid-level region of the ventral visual stream, also adjusts neural responses to changes in distance and in shape both in 2D and 3D (Hinkle and Connor 2001, Hinkle and Connor 2002, Pasupathy and Connor 2002, Srinath, Emonds et al. 2021). At later stages of object vision, IT cortex, and the STS similarly contains neural selectivity for various 3D features and shapes including curvature and shape complexity (Janssen, Vogels et al. 1999, Janssen, Vogels et al. 2000, Janssen, Vogels et al. 2000). The STS, in particular, serves as a region of convergence for numerous different types of visual information (Seltzer and Pandya 1994) and may even facilitate a proposed third visual pathway subserving social interaction and combining of features from the other two visual streams (Pitcher and Ungerleider 2021). These connections may facilitate the

combination of these types of information. Concomitantly, the STS contains multiple patches that respond strongly to disparity and 3D elements of stimuli. (Verhoef, Bohon et al. 2015). At the single neuron level, some STS neurons respond to specific 3D shapes and adjust responses across distance and size while maintaining this tuning to shape (Janssen, Vogels et al. 2000). These results all indicate that neurons in the STS may derive their responses from the natural 3D geometry and the statistics of the natural world.

Natural statistics are often reflected in tuning properties for many neurons within temporal lobe visual regions. In line with this observation, a subpopulation of IT cortex neurons are tuned to shapes aligned with gravity representations, directly mirroring the features of the normal world (Vaziri and Connor 2016). Similarly, tuning in the STS preferred smaller retinal sizes in line with the average retinal subtense of objects suggesting that these regions have some internal model of 3D features of spatial features (Vaziri, Carlson et al. 2014). This model may stem from connections between the dorsal and ventral stream. The STS has direct neural connections with parietal regions in the dorsal stream (Pandya 1984, Seltzer and Pandya 1994). The STS may also distribute this information to inferior temporal cortex with through connections to the nearby ventral stream (Seltzer and Pandya 1989). Evidence from fMRI suggests that parietal regions drive responses in the STS during 3D shape perception (Van Dromme, Premereur et al. 2016, Janssen, Verhoef et al. 2018). Similarly, electrical stimulation of parietal regions activated disparity selective regions along the STS and IT cortex while stimulation within those regions did not activate the parietal cortex (Premereur and Janssen 2020). Furthermore, the STS appears to play a causal role in behavioral reporting of 3D curvature information as manipulations of activity lead to changes in curvature reporting (Verhoef, Vogels et al. 2012). This evidence all indicates that information converges in the temporal cortex from shape-based and spatially based visual regions and this

convergence may play a clear role in integrating 3D information with other forms of vision in the temporal lobe. However, despite this evidence, most studies have not continued to assess properties derived from 3D visual processing. The combination of different types of visual information in the STS may also assist in aligning with other modalities including auditory information to further assemble perception in more natural environments and scenes.

2.1.4 Comparing Carnivore and Primate Sensory Systems

Elements of mammalian brain physiology have been highly conserved across different species and clades. Comparisons of carnivore and primate brains can then illuminate common origins for these behaviors or principles. We can then extrapolate them further to different species and potentially illuminate the basis for new functions. The carnivore superior colliculus has served as a model of spatial multisensory integration that has remained consistent with the macaque brain (Stein, Meredith et al. 1993). At the cortical level, early sensory areas across the cortex have also demonstrated consistency that can further elucidate properties that can provide insight across species. Auditory regions have been extensively compared in this regard. Early auditory areas across numerous mammalian species have revealed extensive homology or analogy and a potential basis for the functional organization within the human brain particularly in core auditory regions (Kaas 2011). Similarly, carnivore and primate early visual areas do display certain similarities in organization and carnivore visual areas have provided insight into the function of these early visual areas in primate cortex (Hubel and Wiesel 1965). Areas 17,18, and 21 of the carnivore brain roughly align and show potential homology with areas V1, V2, and V4, the early visual regions of the primate cortex (Manger, Kiper et al. 2002) (Figure 2.1). However, beyond these core regions, the relationship of higher-order association areas between the carnivore and primate species remains unclear, particularly as it pertains to function.

Comparing the temporal lobe of carnivores and primates also proves challenging as the temporal lobe in primates has further elaborated to accommodate new, often visually driven, regions such as the STS (Dell, Innocenti et al. 2019). Similarly, association areas also do not always have direct homologues or analogues between species due to these different elaborations. But, while ferrets lack the visual acuity of macaques and the same extent of higher order visual cortex or different association areas, they do have a potentially homologous region, area 20, which has yet to be fully explored for some of the various functions and specializations of the macaque STS and temporal cortex. Further investigation of this area could provide greater understanding of the origins of the higher-level regions of the elaborated macaque temporal cortex that enable vision specialized for social perception and behavior.

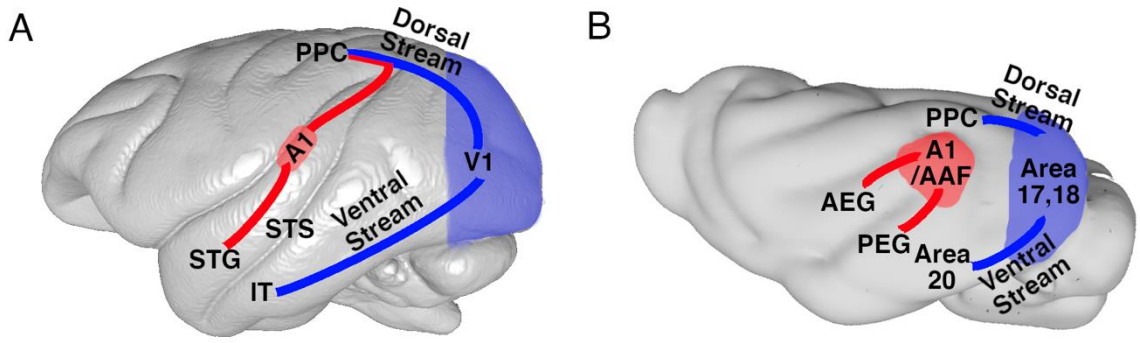

Figure 2.1. Comparison of the Macaque and Ferret Visual and Auditory Systems. A. a schematic of the visual and auditory pathways in the macaque brain where blue represents the visual pathway and red the auditory pathway with labeling of cortical regions involved in both. **B.** a similarly marked schematic of the ferret brain marking the analogous pathways in the ferret cortex. **A, B** were derived from brain atlases based on data from (Rohlfing, Kroenke et al. 2012) (**A**) and (Hutchinson, Schwerin et al. 2017) (**B**)

2.1.5 Ferret Area 20

While ferrets do not display the exact social behavior of macaques, domestic ferrets participate in numerous social interactions that have yet to be explored within the ferret brain (Larrat and Summa 2021). Like macaques, ferrets may have cortical regions that process social stimuli that could provide further insight into how neurons process the real world. Previous work

has suggested ferret area 20 might be an analogous or homologous structure to the STS or other lateral temporal areas (Updyke 1986, Manger, Nakamura et al. 2004). Anatomically, ferret area 20 receives input from high level auditory and visual areas and sits between these areas within the cortex as the STS does in macaques. Tracer studies indicate area 20b, in particular, shares a similar anatomical profile to the STS with connections to high level auditory and visual regions as well as parietal areas that demonstrate multisensory responses (Dell, Innocenti et al. 2019) (Figure 2.2). Anatomical evidence also suggests these functions may include audiovisual integration or even object selectivity.

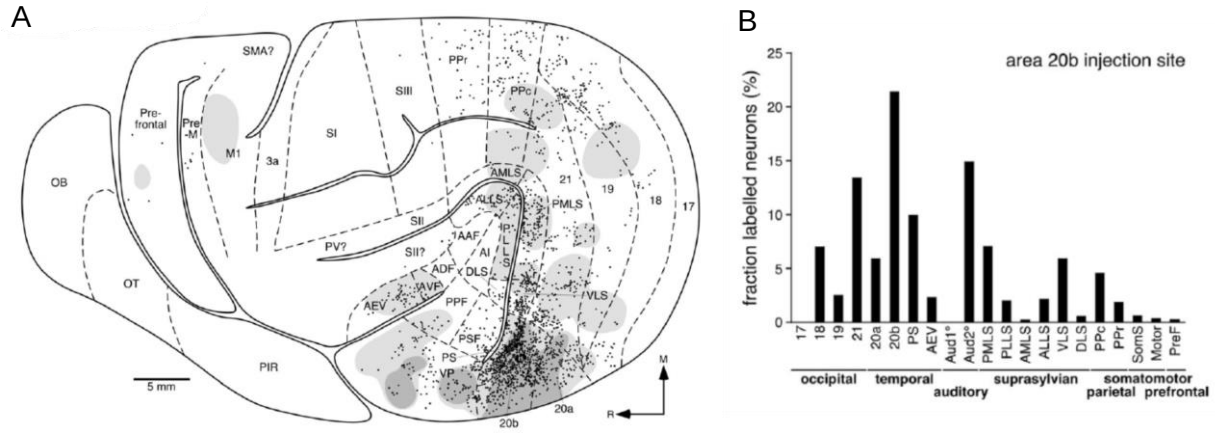

Figure 2.2 Anatomical Connections of Ferret Area 20b. A: diagram of labeled neurons from a retrograde grade tracer placed in area 20b of a ferret. **B:** a bar diagram of the percent of labeled neurons in from this retrograde tracer in different portions of the brain, including wide-spread connections with visual regions and secondary auditory regions. **A, B** adapted with permissions from (Dell, Innocenti et al. 2019)

However, while these anatomical connections have been investigated, few studies have examined the responses of neurons within these areas to determine if area 20b demonstrates any similar functions within the ferret cortex as the STS does within the macaque cortex. Overall, few studies have explored the extent of ferret visual capacity, and few have functionally probed the relationship of area 20b in carnivores and STS within primates. Anatomically, visual areas in the ferret have shown homology with early visual areas in macaque. Visually, both area 20a and 20b reflect properties of lateral temporal areas and STS of the macaque with larger receptive fields and

preferences for certain kinds of motion (Manger, Nakamura et al. 2004). Recently, studies have suggested that ferret spatial processing regions (PMLS, also referred to as the Suprasylvian Sulcal Field, SSY, in the literature (Cantone, Xiao et al. 2006)) can perform similar functions to primate dorsal stream regions (Dunn-Weiss, Nummela et al. 2019, Lempel and Nielsen 2019). These studies provide support for the idea that ferrets cortex may have more homologous or analogous regions with primates in visual processing than previously suspected. Single cell face or object selectivity as well as a large capacity for face recognition has also been demonstrated in sheep (Kendrick and Baldwin 1987), while domestic dogs have shown specialized regions for processing human faces in fMRI (Cuaya, Hernandez-Perez et al. 2016) . While other non-primate species may not have large functional islets such as the face patches, this evidence suggests that face or object selective cells may be present in more species than currently investigated, possibly including carnivores such as ferrets. Domestic ferrets, like sheep, dogs, and macaques, can live in groups with other conspecifics, which may lead to neurons in visual areas dedicated to face processing. Similarly, the process of domestication in mammals may facilitate the specialization of regions or single neurons of the temporal cortex to respond specifically to faces, either of humans or conspecifics. Establishing this kind of selectivity would establish a new comparative connection between these species and expand previous comparative knowledge regarding face selectivity and social perception.

2.2 Audiovisual Integration

While studies have often focused on the functions and organization high-level visual and auditory systems in the temporal lobe, the combination of senses has remained relatively unexplored. In all mammalian species, combining information from different senses is crucial to

survival. The combination of auditory and visual stimuli is particularly important to primates as social animals that must traverse large social scenes with numerous sources of sounds. Even for less social mammals, the proper integration of sounds and sights is crucial to numerous behaviors. In this section, we will discuss the current understanding of the basic principles of audiovisual integration, how these principles currently apply or do not apply to integration in space, and the relationship of social stimuli to these spatial principles across species especially as it pertains to the temporal lobe and the face patches.

2.2.1 Importance of Audiovisual Integration

Each sense has particular features or qualia of a stimulus that cannot translate to another sense (i.e. one cannot see pitch or hear the temperature of a surface). But, while organisms have developed and specialized each sense for different function and different advantages, combining these modalities dramatically increases the amount of available sensory information. As such, the brains of numerous species maximize the combination of different senses to facilitate perception. The combination of auditory and visual information, audiovisual integration, serves an important role among mammalian species. The patterns by which these senses influence one another have been thoroughly studied on the behavioral level broadly in two ways. First, studies utilize multisensory illusions that occur when stimuli from different modality are placed into conflict. Prominent examples include the McGurk effect where a conflicting visual stimulus can shift the perception of a syllable or the sound induced flash illusion where the addition of multiple beeps can shift the perception of the number of visual events (MacDonald and McGurk 1978, Shams, Kamitani et al. 2000). These illusions indicate one sense can shift the final percept of another sense and provide a framework for that transformation. Second, studies compare the performance of a

subject on a task in unisensory conditions compared to multisensory conditions. Combination of auditory and visual stimuli improves detection of stimuli, decreases reaction time, and improves spatial localization compared to either sense alone (Gielen, Schmidt et al. 1983, Stein, Meredith et al. 1993, Stein and Stanford 2008). All these enhancements to sensory abilities indicate the importance of multisensory integration in behavior. However, the relationship of these behavioral changes and improvements with the responses of brain remains challenging to fully understand and study of it on a neurophysiological level remains nascent. While all these aspects of audiovisual integration are important, this chapter will primarily focus on the spatial elements and how audiovisual integration improves spatial perception.

2.2.2 Principles of Integration at the Single Cell Level

Audiovisual integration often occurs even at the level of a single neuron across numerous regions of cortex and subcortical structures. Previous criteria have defined multisensory integration at the single cell level as a difference in response for the combination of sensory stimuli relative to the summation of the response to unisensory stimuli. In the case of audiovisual integration, this definition means a multisensory neuron responds differently to audiovisual stimuli as compared to visual or auditory stimuli alone. These responses can exhibit as either an increase in spiking response relative to the unisensory components referred to as multisensory enhancement or a decrease referred to as multisensory depression (Stein and Stanford 2008). Studies have often assumed, and some have shown that multisensory enhancement indicates a stronger effect and strengthens behavioral advantages while multisensory depression weakens the behavioral advantages conferred by multisensory stimuli. However, this multisensory modulation can serve a variety of purposes across numerous brain regions and do not always match this exact description

(Perrodin, Kayser et al. 2015). Early research of single neurons focused on the superior colliculus of cats and has developed three rules of single cell integration: the temporal rule, the spatial rule, and the principle of inverse effectiveness. The temporal rule states that single neurons are more activated and respond more strongly to multisensory stimuli when they are temporally coincident, the spatial rule indicates that neurons are activated more when multisensory stimuli originate from spatial locations that overlap a cells receptive fields, and the principal of inverse effectiveness states that multisensory modulation of a neuron is most effective when unisensory stimuli only weakly activate a neuron (Stein, Meredith et al. 1993, Stein and Stanford 2008). While these rules have developed, their relevance to behavior and their scalability across the brain remains controversial. These rules were primarily developed in anesthetized recordings of animals utilizing simple stimuli. As with the effect of multisensory modulation, different regions of the brain express various elements of these principles while eschewing others and can express selectivity for particular stimuli rather than simply pairing together any visual and auditory stimuli (Perrodin, Kayser et al. 2014). These combinations indicate that understanding of how single neurons respond to multisensory stimuli requires further investigation in new paradigms and with a better understanding of the function and relationships of a particular region.

2.2.3 Locations of Audiovisual Integration

The investigation of the cat superior colliculus has provided the basis for early audiovisual and multisensory research. The superior colliculus of cats has served as the model and the early rules of multisensory integration have derived from electrophysiology studies of its responses (Stein, Meredith et al. 1993, Stein and Stanford 2008). On the level of cortex, the regions of multisensory integration range from primary sensory areas to so-called higher order association

areas receiving projections from multiple sensory areas (Ghazanfar and Schroeder 2006, Cappe, Rouiller et al. 2009). Despite the predominance of a single sense, primary sensory cortical regions across species show long range connections with multisensory or regions processing other modalities, and some have demonstrated multisensory responses (Figure 2.3B). In cortical areas of the macaque, evidence indicates visual information modulates auditory responses in core, belt, and parabelt auditory regions (Ghazanfar, Maier et al. 2005, Kayser, Petkov et al. 2007, Kayser, Logothetis et al. 2010). Like in primates, primary auditory cortex of ferrets has repeatedly shown single unit responses and modulation by visual signals indicating a parallel between mammalian species (Bizley and King 2009). While auditory cortex has received some examination for the effect of visual signals, outside of rodents, visual cortex has rarely been examined for the same multisensory cells. In cats, primary visual cortex itself demonstrated multisensory responses to auditory stimuli in a spatially specific manner providing some evidence that the mammalian brain has the capability of processing multiple senses (Morrell 1972, Fishman and Michael 1973).

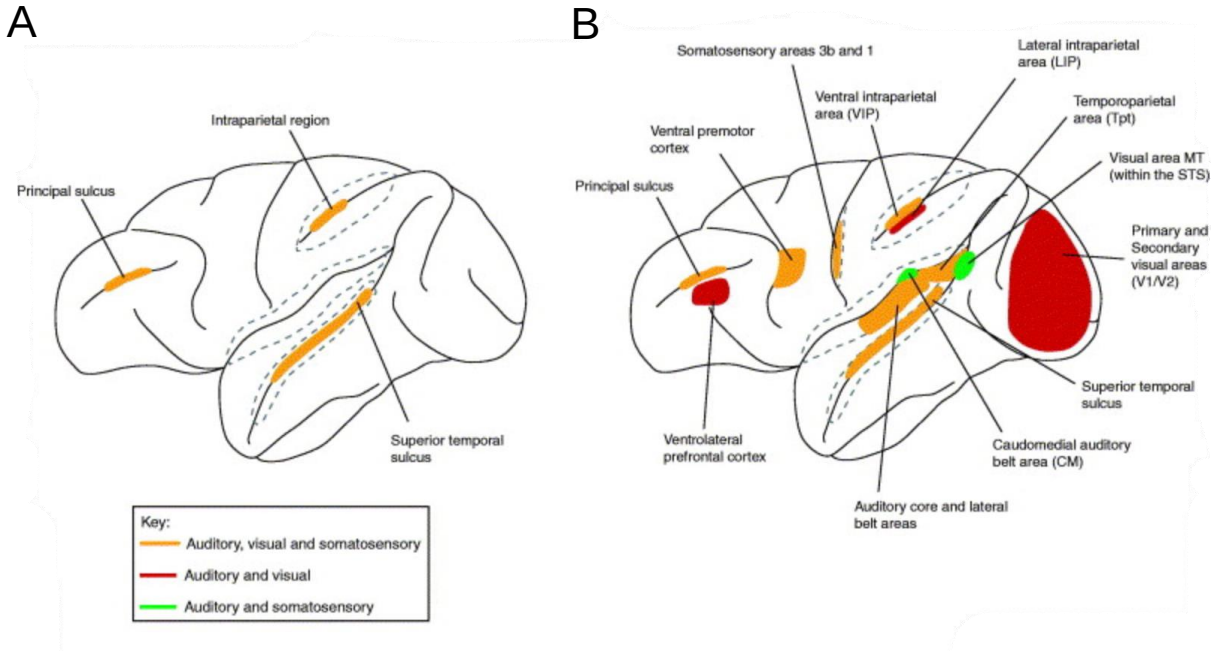

Figure 2.3 Multisensory Regions of the Macaque Brain. A: A schematic representation of historical multisensory association areas in the macaque brain. **B:** A similar schematic including more primary sensory areas and multisensory areas. **A,B** taken with permission from (Ghazanfar and Schroeder 2006)

Electrophysiology in humans has also indicated that audiovisual modulation occurs early in visual processing (Molholm, Ritter et al. 2002) and anatomical tracers in macaques indicate visual cortex receives some input from auditory and multisensory regions of cortex (Falchier, Clavagnier et al. 2002, Rockland and Ojima 2003). Despite these connections, early and many later visual cortical regions of primates have not been examined very heavily for the influence of auditory stimuli.

Beyond primary cortical areas, secondary association areas lie between high-level regions of different modalities and receive cortical input from multiple modality specific regions. These areas contain various mixtures of multisensory cells that perform different functions (Figure 2.3B). In the macaque, these regions include parts of every lobe of the brain and exhibit specific properties depending on their position. Association regions of carnivore cortex have also been explored to pinpoint their exact positions but rarely in the context of their functions. Within the carnivore cortex, multisensory integration has often been found a region named anterior ectosylvian sulcus (AES) that responds to auditory, visual, and somatosensory stimuli, lateral rostral suprasylvian area (LRSS), which responds to auditory and somatosensory stimuli (Meredith, Allman et al. 2009). But the precise relationship of these association regions in the carnivore brain and those in the macaque brain remains mysterious. Interestingly, more visual lateral temporal areas of the ferret brain have yet to be explored for their audiovisual integration despite widespread anatomical connections (Dell, Innocenti et al. 2019), consistent with the lack of multisensory studies of more visual areas of the brain. These areas include area 20b, which, as described above, may serve as a homologous or analogous region to the STS. Among the visual association areas within the primate brain, the STS is among the few that has been examined for audiovisual integration. In humans and other primate species, audiovisual integration plays an important role in social communication, for example during the perception of a conspecific's vocalization and concomitant

facial behavior (Barraclough and Perrett 2011). The temporal cortex, and particularly the STS, contain zones of convergence for high-level sensory signals (Barraclough, Xiao et al. 2005, Ghazanfar and Schroeder 2006). In the macaque, the STS fundus borders high-level visual and auditory cortex (Seltzer and Pandya 1978, Pandya and Seltzer 1982, Hackett, Stepniewska et al. 1998, Kaas and Hackett 2000) and exchanges connections with other multisensory areas, including ventrolateral prefrontal cortex and intraparietal cortex (Pandya and Seltzer 1982, Seltzer and Pandya 1994, Romanski, Bates et al. 1999). At the single cell level, neurons within portions of the STS respond to visual and auditory stimuli, as well as their combination (Benevento, Fallon et al. 1977, Baylis, Rolls et al. 1987, Barraclough, Xiao et al. 2005). STS responses also exhibit elements of selectivity as integration is only exhibited to temporally coherent combinations of actions.

In addition to reciprocal connections with high level auditory and visual regions of the brain, the fundus of the STS also shares numerous connections with multisensory regions of the brain including the ventrolateral prefrontal cortex (VLPFC) of macaques. Neurons in the VLPFC selectively combine audiovisual stimuli involving faces and vocalizations (Sugihara, Diltz et al. 2006). This face information stemming from the STS may influence and enable this selectivity within the VLPFC (Romanski, Bates et al. 1999). Some neurons in STS have also expressed responses to multisensory stimuli that includes faces (Barraclough, Xiao et al. 2005). Of the numerous face patches in macaques described above, a subset of these face patches lies in the STS fundus overlapping with the regions previously found to be multisensory (Figure 2.3A). However, whether the neurons within these functionally localized fundus face patches respond to or are modulated by auditory stimulus remains unexplored.

Face selective regions and the STS of macaques show crucial homology with face patches and the STS in humans (Tsao, Moeller et al. 2008). Face selective regions in human cortex, by

contrast, have been examined for multisensory integration with fMRI demonstrating responses to certain auditory and tactile stimuli (von Kriegstein, Kleinschmidt et al. 2005, Ratan Murty, Teng et al. 2020). While these areas predominately respond to visual stimuli, this multisensory modulation indicates neurons within face-selective regions may be modulated by auditory stimuli.

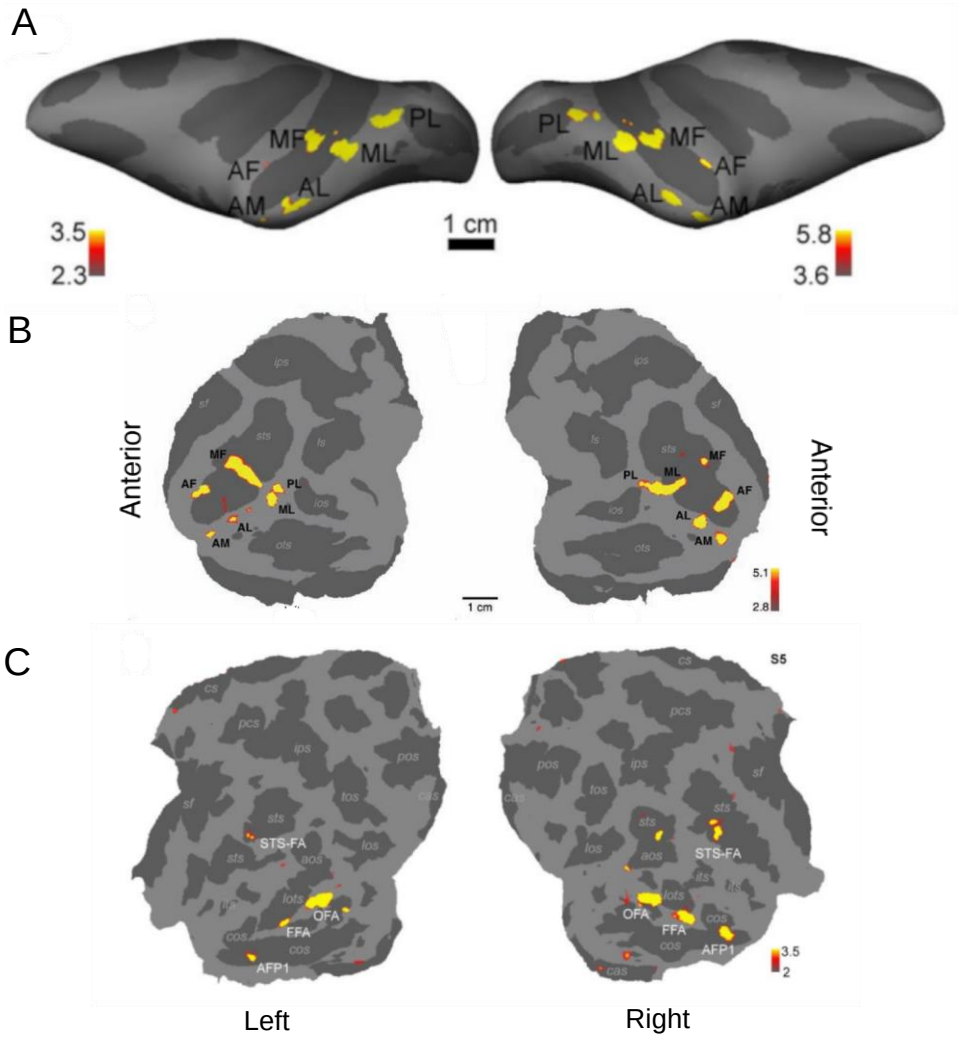

Figure 2.4 Macaque and Human Face Patch Systems. A: Inflated fMRI maps of the macaque brain displaying the face patches. **B:** Flat maps fMRI maps of the macaque face patches. **C:** flat of the human brain displaying the corresponding face patches. **A, C** taken from (Tsao, Moeller et al. 2008) with permission (Copyright 2008 National Academy of Science, **B** taken from (Moeller, Freiwald et al. 2008). Reprinted with permission from AAAS

Face responsive regions between human and macaque cortex also demonstrate potential homology in arrangement and response properties, including in the STS, though they do not directly mirror one another (Figure 2.4B). Beyond face selectivity, the posterior STS of humans,

like the STS of macaques, also combines and responds uniquely to audiovisual stimuli often in an action or object specific manner (Beauchamp, Argall et al. 2004, Beauchamp, Lee et al. 2004) and STS face selective regions show a preference for moving faces over static faces (Pinsk, Arcaro et al. 2009) (Figure 2.4).

But unlike in macaques, human STS face-selective regions have been examined for their audiovisual properties and have been shown to respond to specific pairings of audiovisual stimuli with specific attention to facial features (Zhu and Beauchamp 2017). While these human studies have relied on fMRI and not interrogated single-unit responses, they indicate that macaque face patches within the STS could likely follow similar patterns. An important aspect of the human STS face patch is its responses to motion and role in social communication. Similar to this specialized visual organization for faces, macaques have a specialized auditory region for social stimuli, voices. This voice region was similarly discovered and defined through fMRI, contrasting macaque voices with calls from other species of animal (Petkov, Kayser et al. 2008). This voice region rests in the superior temporal plane, anterior to even the parabelt regions of the macaque auditory cortex suggesting it lies in a higher order auditory region (Figure 2.5A).

This specialization indicates this region may mirror the face patches in the auditory realm. The two regions do not precisely overlap as extracellular recordings within this region discovered a much lower proportion of voice selective neurons in voice patches than face selective neurons typically found in the face patches (Perrodin, Kayser et al. 2011). However, this similar functional specialization indicates this region may provide further insight into how more specialized areas process multisensory information. Indeed, this voice patch, while predominately auditory, is modulated by the addition of visual stimuli (Perrodin, Kayser et al. 2014, Perrodin, Kayser et al. 2015). Interestingly, the visual responses do not need to precisely match the auditory stimuli in

order to affect the neural response (Figure 2.5), suggesting that more specialized areas show primary sensitivity to one modality and less to a secondary stimulus (Perrodin, Kayser et al. 2014). Altogether, this region, specialized for auditory social stimuli, indicates that the some of the face patches with similar specialization could also be modulated by auditory stimuli and provides clues as to how single neurons in face patches might respond to audiovisual stimuli.

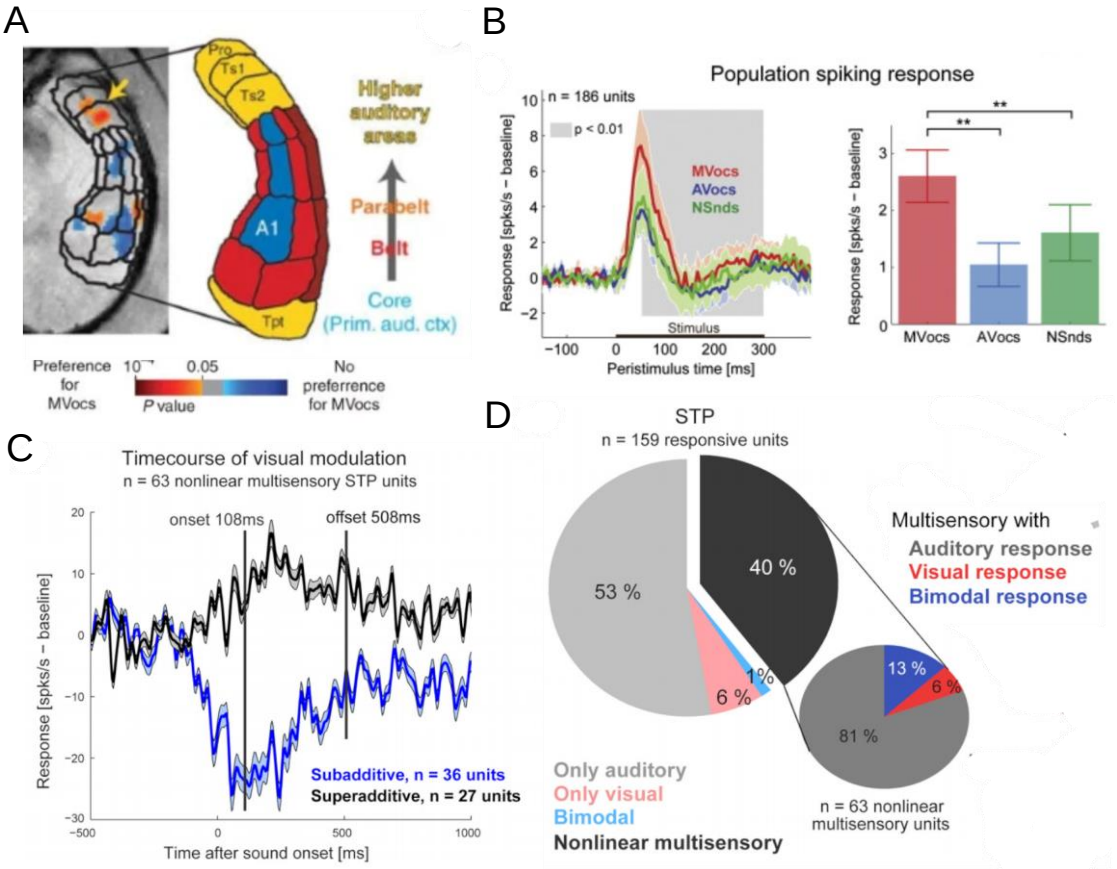

Figure 2.5 The Macaque Voice Patch System. A: a fMRI mapping of the voice patch in the superior temporal plane and a schematic representation of the voice patch in the auditory processing hierarchy. Taken from (Petkov, Kayser et al. 2008) with permission. **B:** the population response of voice patch neurons to vocalizations and other sound stimuli. Taken from (Perrodin, Kayser et al. 2011) with permission. **C:** Population response of voice patch neurons to audiovisual stimuli separated by superadditive and subadditive responses. **D:** a pie chart of the proportion of cells in the voice patch, which display audiovisual integration. Taken from (Perrodin, Kayser et al. 2014) with permission

2.3 Audiovisual Integration in Space

However, while multisensory areas have been located in the mammalian, their functions in behavior and perception beyond processing multiple modalities remains mysterious. Importantly,

these regions like in the STS, overlap with many regions that depend on spatial elements. The advantages of audiovisual integration depend on synchronizing and aligning these different sensory components in space. Despite its importance, the spatial aspects of audiovisual integration remain relatively unexplored and have rarely been examined in naturalistic contexts, which include multiple sound sources and naturalistic stimuli.

2.3.1 Spatial Overlap and Spatial Information

The combination of the auditory and visual modalities into a single audiovisual object in space remains an important underexplored aspect of audiovisual integration. Both senses encode space in dramatically different ways as the visual system determines spatial position through a retinotopic map projected directly from its sensory epithelium while the auditory system calculates space based on interaural timing and phasic differences between its two cochleae. Despite these differences in spatial calculation and reference frame, the brain can link spatial location of single source of these two modalities into a single percept. This linkage is part of a phenomenon referred to as binding where coherence, both temporal and spatial, results in the combination of multisensory components into one single class of object (Stein and Stanford 2008, Bizley, Maddox et al. 2016). This binding has been found to improve localization of targets in space over either unisensory stimulus in repeated studies (Alais and Burr 2004). However, behavioral effects of multisensory information do not always directly result from binding and the spatial principles that maximize neural responses are not always necessary for behavioral enhancements. Importantly, auditory and visual stimuli can influence and even shift the perception of spatial cues of one another without fully binding together. For example, the ventriloquist effect, wherein a temporally synchronized audiovisual motion can capture slightly spatially mismatched auditory cues to the

visual position (Alais and Burr 2004). This effect demonstrates a case wherein binding can misalign but when modalities are first localized separately the shift in spatial position produced by the illusion weakens (Kording, Beierholm et al. 2007). Auditory stimuli can also affect the ability to of the visual system to detect targets in more eccentric space and the addition of auditory stimulus to visual stimuli even without precise alignment can improve spatial detection (Gleiss and Kayser 2013, Spence 2013).

Audiovisual integration can also vary with position in 3D space. In human behavior, audiovisual integration appears more effective at greater distances. Distance also shows interactions with other spatial factors as it can show greater facilitation for stimuli of smaller retinal subtense (Van der Stoep, Van der Stigchel et al. 2016). Similarly, audiovisual looming stimuli, stimuli that appear to approach rapidly from distance, also show multisensory facilitation in reaction times beyond the facilitation common for multisensory stimuli (Cappe, Thut et al. 2009, Cappe, Thelen et al. 2012). This preference for audiovisual looming is preserved across species as macaques prefer to look at audiovisual looming signals (Maier, Neuhoff et al. 2004). Importantly, this facilitation is improved when stimuli are presented with binocular depth cues and mimic the spatial frequency of naturalistic scenes (Conrad, Kleiner et al. 2013). These results indicate that audiovisual integration in 3D space may also depend on the effects of natural statistics. However, few studies of 3D space in audiovisual integration have evaluated the role of physical size. Studies that have investigated physical size have primarily focused on the McGurk effect, finding that the illusion salience weakens with smaller sizes under certain conditions (Colin, Radeau et al. 2005) but the neural signals underlying this change remain mysterious. Similarly, the facilitation of audiovisual integration in physical size has rarely examined with 3D stimuli, a potentially salient or important dependence for integration.

More recent work on human behavior has suggested that the pattern of resolving space in multisensory stimuli can also vary based on task requirements and context indicating that the understanding of audiovisual space remains far from complete. Binding and the strategies by which the brain binds together complete multisensory objects in space may play a more important role in more crowded naturalistic environments with multiple sound sources and full visual environments. Real world environments, especially for primates, often include numerous conspecifics, sound sources, and competing stimuli. In these more crowded scenes, matching and pairing visual and auditory stimuli becomes more challenging and more crucial. Audiovisual integration can improve the ability of individuals to discriminate between features of competing sounds by forming a single percept even in challenging scenarios (Maddox, Atilgan et al. 2015). This integration can help disambiguate features of the competing stimuli even when the stimulus of the other modality does not directly relate to the targeted feature of the stimulus it binds with, making a stronger percept of the whole object (Bizley, Maddox et al. 2016). This phenomenon has rarely been studied at the at the spatial level. But those studies that have examined how spatial elements affect stimulus competition and vice versa determined even audiovisual effects that did not initially depend on spatial features became more sensitive to spatial overlap and coincidence when stimulus competition was introduced (Bizley, Shinn-Cunningham et al. 2012, Fleming, Noyce et al. 2020). Multisensory spatial binding may then serve a stronger role in scene segmentation in ways that remain unexplored at the neural level.

2.3.2 Spatial Integration of Audiovisual Signals in the Brain and Single Cells

To successfully combine space from the auditory system and visual system, the brain must also bridge reference frames. The auditory system encodes space relative to the head, while the

visual system encodes space relative to the eyes. When integrating audiovisual stimuli, single neurons bridge this gap through numerous patterns across the cortex. This early work on multisensory space has primarily utilized carnivores and like much of multisensory research, focused on the cat superior colliculus rather than cortical areas. The superior colliculus contains topographic maps of auditory and visual space. Responses in this region depend on overlapping of these spatial receptive fields of both auditory and visual space for single neurons. These neurons show enhanced responses when the auditory and visual stimuli spatially overlap into the aligning receptive fields, no modulation when auditory and visual stimuli do not overlap, and even depressed responses when auditory and visual stimuli are misaligned into so-called inhibitory zones (Stein, Meredith et al. 1993). These receptive fields do not necessarily directly overlap in space, so auditory and visual stimuli do not have to originate from the exact same point source to elicit multisensory enhancement. This multisensory enhancement has primarily been studied in the context of orienting and has been shown to improve reaction times and localization across space. Interestingly, superior colliculus neural responses auditory stimuli in space evolve over time in the macaque, transforming from initial presentation to response. The neurons initially encode auditory stimuli in a hybrid reference frame between an eye and head centered response before shifting to an eye-centric reference frame before a response, highlighting its role in visual orienting (Lee and Groh 2012). Interestingly, the auditory dominated inferior colliculus has yet to be examined for reference frame but has shown some influence from the position of the eye (Groh, Trause et al. 2001). However, how strongly these principles translate to cortical audiovisual processing remains unclear.

On the cortical level, different regions have expressed elements of these principals but predominately utilizes different methods for spatial integration. Cortical processing may also reflect a difference in function of audiovisual space compared to the orienting based responses of superior colliculus. In the ferret auditory cortex, previous studies have demonstrated the importance of visual influence to enhancing and improving spatial perception. Mutual information and spike rate analysis has established that visual stimuli provide information regarding space to in ferret auditory cortex neurons (Figure 2.6 A,B) rather than forming overlapping maps of space

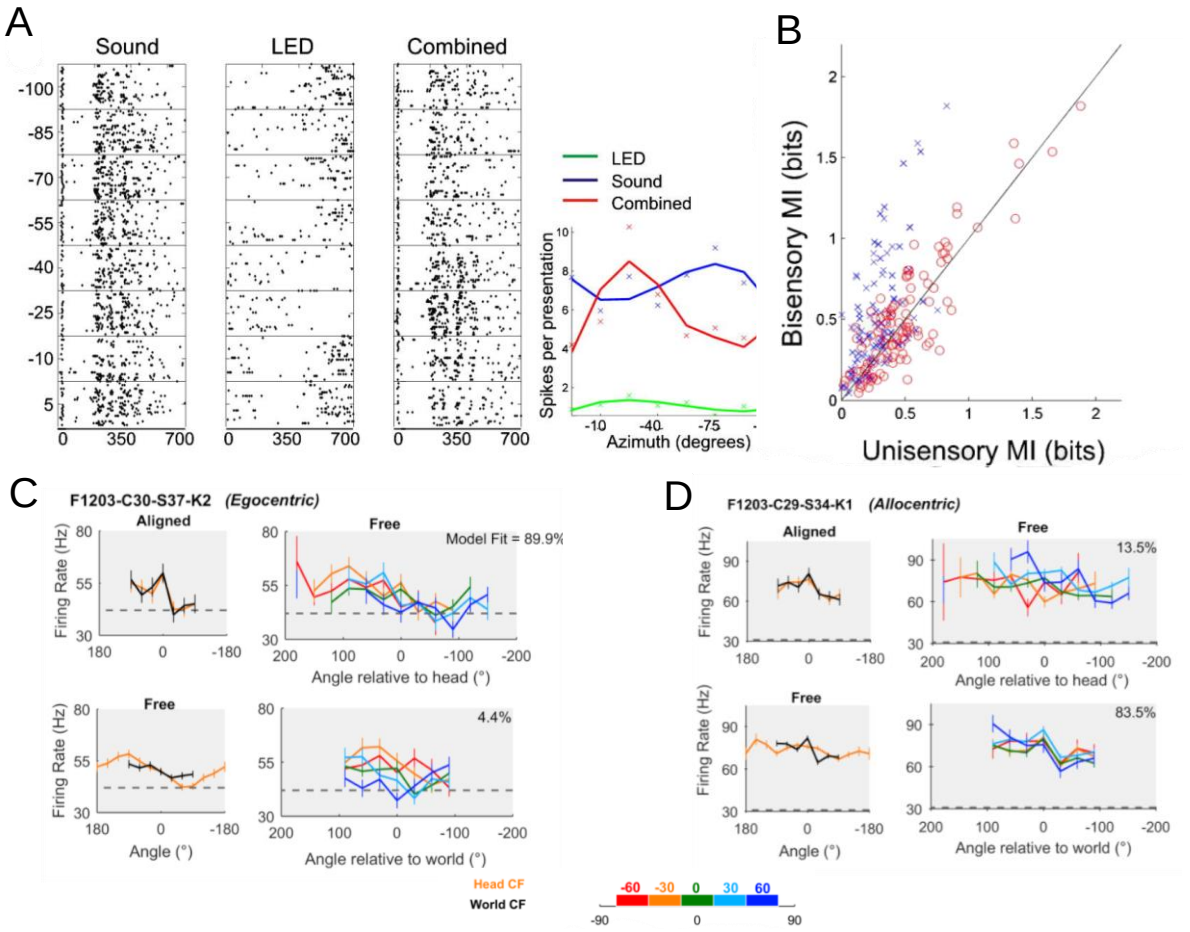

Figure 2.6 Spatial Tuning in Auditory Cortex. **A:** a single neuron example in the auditory cortex that showed a sharper spatial tuning when visual stimulus was added. **B:** a scatter plot displays the shift in mutual information from unisensory auditory (blue) and unisensory visual (red) stimuli to multisensory stimuli indicating visual stimuli tended to increase information when added to auditory stimuli: **A, B** taken from (Bizley and King 2009) with permission. **C-D** Single neuron example of ego and allocentric reference frames in auditory cortex. **C:** an example of an egocentric neuron with tuning curves for when reference frames were aligned (top left) and when they are allowed to vary (bottom left) as well as to the angle of the head (top right) compared to angle relative to the world (bottom right). **D:** an example of an allocentric neuron with the same plots. **C,D** taken from (Town, Brimijoin et al. 2017) with permission

as in the colliculus (Bizley, Nodal et al. 2007, Bizley and King 2008). In contrast with the more eye aligned coding in the superior colliculus, the ferret auditory cortex encodes space in both head centric and world centric coordinate systems meaning some cells encode auditory space (Figure 2.6 C,D) invariant to the direction of the head (Town, Brimijoin et al. 2017). Beyond this study, few have examined the coordinate frames by which auditory cortex in any species encodes space. Within macaque auditory cortex, a proportion of neurons in the macaque auditory regions are modulated by the position of the eyes (Werner-Reiss, Kelly et al. 2003) suggesting but not showing that neurons in auditory cortex may also process space in eye-centric patterns. However, due to limitations of macaque studies, whether cells in macaque auditory cortex can encode space in world-centric coordinates remains unclear.

In the macaque, those studies that have examined of audiovisual space in the cortex have primarily focused on posterior parietal regions of cortex as these areas correspond with parts of the dorsal visual stream. Within the posterior parietal regions, particularly the intraparietal areas, cells seem to perform important operations in shifting reference frames and aligning spatial elements of audiovisual stimuli. Many audiovisual neurons in these regions exclusively exhibit modulation when audiovisual stimuli are spatially aligned. These areas, while they contain neurons that respond in head-centric and eye-centric manners, primarily respond in a complex not clearly head or eye centric frame but in a complex hybrid of the two (Mullette-Gillman, Cohen et al. 2005, Mullette-Gillman, Cohen et al. 2009). These responses also match with reports that the position of the eyes influences neural responses found in auditory cortex and inferior colliculus (Groh, Trause et al. 2001, Werner-Reiss, Kelly et al. 2003). With the strong connections between posterior parietal cortex and the STS (Pandya 1984), these results in parietal areas may provide clues as to the principles of integration in the STS but studies of audiovisual space have rarely examined the

STS (Figure 2.7A). Visual studies have often classed or assumed that the STS follows principals of the ventral visual stream and, in that pattern, coordination of temporal aspects or even semantic aspects influences the multisensory response of single neurons in STS (Barraclough, Xiao et al. 2005).

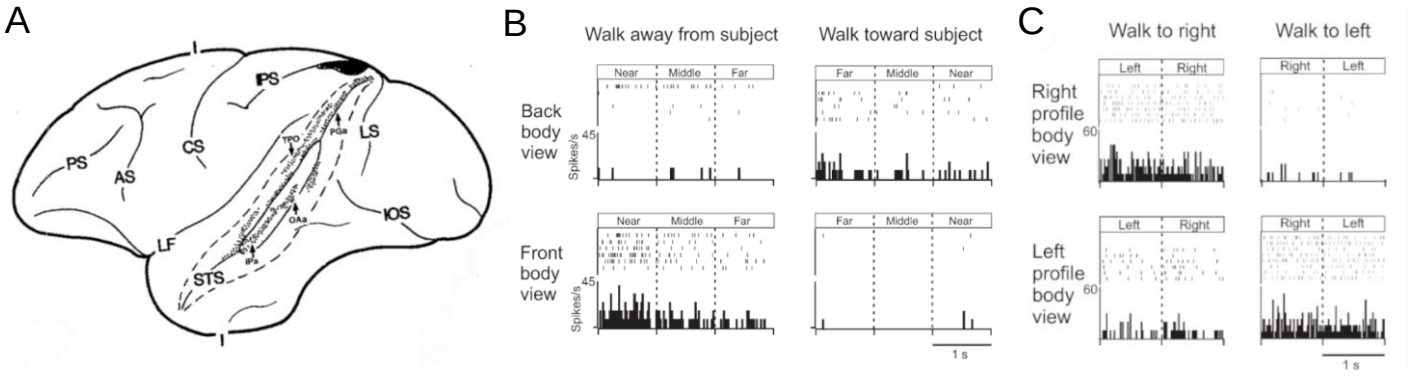

Figure 2.7 Parietal Connections and Spatial Feature Selectivity in STS. A: labeling in the STS (shown in hatch marks) produced by an anterograde tracer injection in the intraparietal sulcus (marked in black), a spatial processing region of the macaque brain. Taken from (Seltzer and Pandya 1984) with permission **B, C**: Peristimulus time histograms (PSTH) of the response of STS neurons to different spatial stimuli. **B:** PSTH of the response of a neuron to frontal and back views of a body and forward and backwards motion at different distances. **C:** PSTH of the response of a neuron to left and right views of a body and movement to the left or right from different locations. Taken from (Jellema, Maassen et al. 2004) with permission

But initial studies of the ventral stream in the macaque did not include the STS and STS neurons often respond to certain spatial principals including locations in space (Jellema, Maassen et al. 2004), particularly forwards and backward motion (Figure 2.7B,C). Concomitantly, audiovisual looming increases coherence of neural responses in the STS indicating 3D space may play an important role in audiovisual integration in this region (Maier, Chandrasekaran et al. 2008). Single neurons within STS also respond to biological motion, which may assist in binding auditory and visual stimuli through temporal coherence (Jastorff, Popivanov et al. 2012). Moreover, previous examination of electrophysiology in humans has suggested that spatial properties of multisensory responses may only appear during stimulus competition. These functions combined suggest the STS may play a role in combining spatial elements in audiovisual integration.

2.3.4 Active Sensing

Across species, experiments examining space in both multisensory integration and individual senses have often been conducted with head-fixed restrained animals. While these previous studies of the spatial component of perception have discovered numerous important features, they have often neglected the role of active sensing. Active sensing, in biology, refers to the process of attentionally or mechanically sampling an environment to gather information (Schroeder, Wilson et al. 2010). Perception of space, in particular, requires the ability to interact and probe the space using different senses to segment scenes into individual components. By restraining these animals, experiments of senses often attempt to reduce the ability of the subject to properly examine these properties. Even amongst studies of vision, subjects are often trained to fixate and passively view stimuli, which has yielded numerous important discoveries but does not actively engage natural eye movements or natural sampling of an environment (Leopold and Park 2020). Experiments that have examined these self-directed and attentional components of perception have demonstrated these behaviors reveal new cognitive properties previously unexplored. Among its benefits, active sensing can resolve stimulus competition among many stimulus sources. The Cocktail Party phenomenon provides a clear example wherein attention and active sensing augment the response to a particular auditory stimulus over background noise or competing sounds (Cherry 1953). Using active attentional selection, the brain can parse complex scenes to pinpoint a desired stimulus among many. Active sensing can also interplay with multisensory integration gaining benefits in resolving competition in scenes while also enabling more multisensory facilitation (van Atteveldt, Murray et al. 2014) (Figure 2.8A, B). Studies in humans have shown the resolution of different auditory stimuli in environments with multiple stimuli and the Cocktail phenomenon benefits greatly from the addition of temporally aligned visual stimuli (Zion Golumbic, Cogan et al. 2013).

On the neural level, the influence of active sensing has mainly been discovered in LFP responses and current source density (CSD) of early neural areas. In macaque V1, CSD oscillation phase in different frequency bands varied depending on attention to visual or auditory stimuli (Lakatos, Karmos et al. 2008). Primary auditory regions across species have also been found to disambiguate competing stimuli at the single cell level through these oscillation (Schroeder, Lakatos et al. 2008, Zion Golumbic, Ding et al. 2013). The Cocktail Party phenomenon can also benefit from visual input to enhance representation of auditory signals to improve the active elements of auditory stimulus selection (Zion Golumbic, Cogan et al. 2013) (Figure 2.8C, D). However, these experiments have still often neglected the role of free movement and spatial perception in understanding these mechanisms. In macaques, enabling simple free-viewing of naturalistic movie stimuli rather than strict fixation revealed visual regions of the brain that were invariant to movement as compared to others (Russ and Leopold 2015, Russ, Kaneko et al. 2016). Free movement and active sensing may also figure into questions of reference frames as these systems of perceiving space evolved in concert with motor systems to perform actions. Enabling free movement revealed neurons in the ferret auditory cortex that computed auditory space in world-centric frames (Town, Brimijoin et al. 2017). How active sensing plays a role in combining these modalities in space remains poorly understood. Free movement may elucidate more principals of audiovisual space and enable better understanding of the role of multisensory integration in resolving stimulus competition in more naturalistic environments.

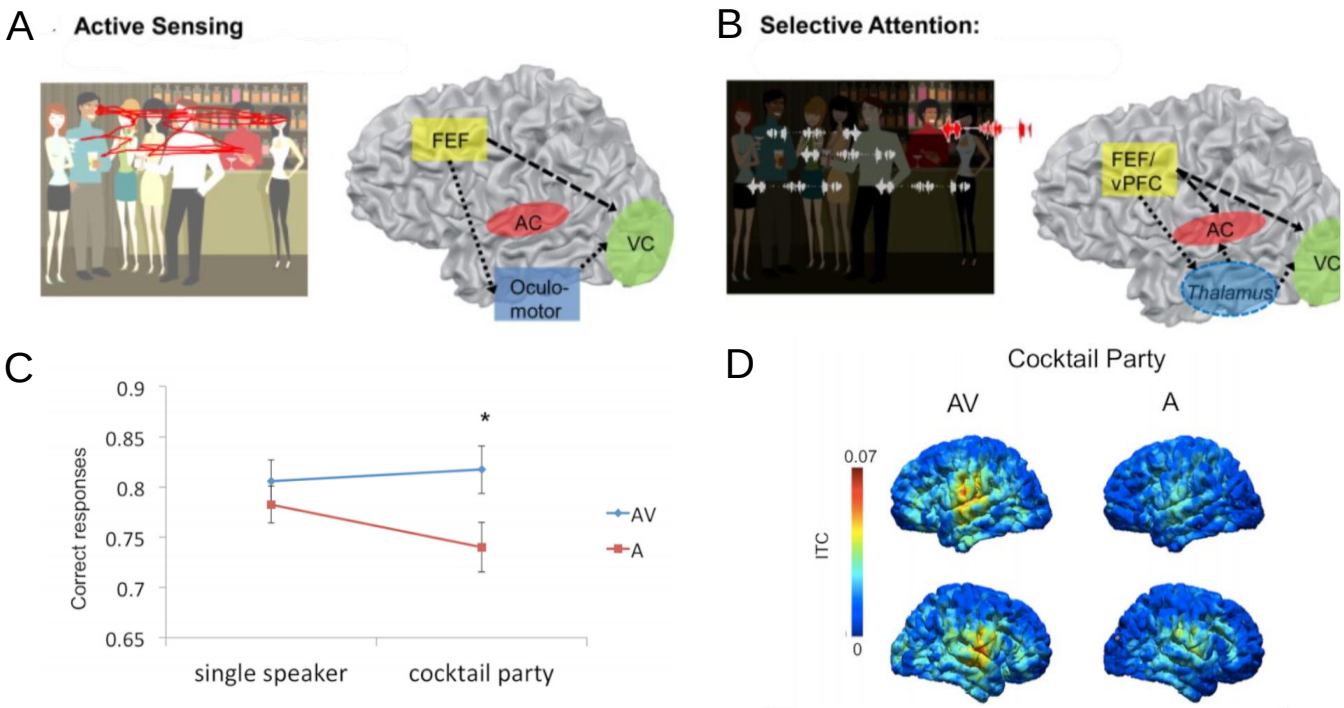

Figure 2.8 Interaction of Active Sensing and Multisensory Integration. A, B: Schematic representations of active sensing and perceptual selection and its processing in the brain. **A:** A schematic representation of searching an environment with the eyes and the brain regions involved. **B:** A schematic representation of active sensing and multisensory processes enabling isolation and magnification of a particular stimulus in a crowded environment and the brain regions engaged. **A, B** adapted from (van Atteveldt, Murray et al. 2014) with permissions. **C:** Ratio correct of distinguishing speakers of a single speaker as compared a cocktail phenomenon in audiovisual and auditory conditions showing AV can assist in distinguishing between competing stimuli. **D:** MEG mapped phase changes for auditory stimuli during the cocktail phenomenon with and without visual stimuli. These results indicate stronger phase tracking in auditory cortex for AV conditions that A conditions during competing stimuli: Adapted from (Zion Golumbic, Cogan et al. 2013)

2.4 Project Outline

Despite the diversity of functions and information that arrives in the temporal lobe of both animals, few of these functions have been examined together. While audiovisual integration, spatial perception, social perception, and comparative aspects of the primate and ferret brain have each received some individual attention they have rarely been examined within these temporal lobe regions. However, real world scenes often involve many spatially separate sources of stimuli, involving spatial assessments in 3D, and multisensory stimuli. To fully understand how the

mammalian brain interacts with the real world these seemingly separate concepts must receive further analysis and investigation as a whole.

The goal of this project is to further understand the various ways the temporal lobe combines information. To explore these functions, we conducted neural recordings on different animals, with a focus particularly on single neuron responses. While the population doctrine of neural activity has gained popularity recently, understanding population responses requires assumptions of the feature space being mapped (Saxena and Cunningham 2019). Single neuron responses remain useful to determine the variety of responses without these assumptions and can illuminate a diversity of functions.

This project breaks into four distinguishable sections that each detail important aspects of audiovisual integration in space. First, I will examine and explore the effects of audiovisual integration in macaque face patches to establish it as a model region for these behaviors. Second, I will explore the tuning of a face patch to 3D spatial features, particularly physical size. Third, I will examine spatial aspects of audiovisual integration in both macaque STS. Finally, we will probe the selectivity for naturalistic and social elements within ferret temporal regions. We hypothesize that multisensory responses will enable these neurons to spatially segment scenes with multiple stimuli.

2.4.1 Care of Animal Subjects

Because these investigations target single neural activity deep in the temporal cortex, these investigations required the use of the animal models. Before discussing the project, we here discuss the steps taken to provide care and well-being to animal subjects. To ensure ethical standards, these experiments were in accordance with the principles of the 3Rs. Macaques and ferrets were repeatedly used across different experiments to ensure the fewest number of animals received

electrode implants. Similarly, experiments in both macaques and ferrets used chronic electrode implants to ensure the lowest chance of adverse side effects and the lowest interference with the daily life of the animals involved. Further, we worked to ensure all animals received the utmost care to during and outside of experiments. The wound margins around all implants were regularly cleaned to minimize discomfort and disruption of normal activity. Macaques also received regular MRI scans to check for infection and both sets of animals were consistently monitored by experimenters and animal care staff for injuries or changes in behavior that indicated ailments or discomfort. Outside of experiments, animals were group-housed and allowed to play with conspecifics to ensure social well-being. Additionally, animals received multiple pieces of enrichment to further ensure mental well-being. The health of all animal subjects was prioritized, dangerous infections or injuries were immediately treated, and subjects were not used in experiments until expressly allowed by veterinary staff. With these actions, we attempted to maintain a strong culture of care for both macaques and ferrets while ensuring the project could obtain the best possible data.

2.4.2 Electrophysiological Recordings of Face Patches

Fundus face patches, similar to the STS at large, respond more to facial motion than face patches in more inferior temporal cortex (Figure 2.9A), making the fundus face patches strong candidates for binding acoustic and visual stimuli (Fisher and Freiwald 2015). The connections with the posterior parietal regions of the brain may also convey spatial information into these regions, which may enable fundus face patches to participate in the spatial operations of audiovisual integrations (Figure 2.9B). Work on face selective neurons in STS has indicated spatial elements that play a role in social interaction, such as head direction, gaze direction and location,

can influences spiking responses (Perrett, Hietanen et al. 1992, Jellema, Maassen et al. 2004). Of the fundus face patches, the anterior fundus (AF) face could likely integrate these spatial visual signals with auditory signals.

The AF face patch occupies a high-level position in the face patch system, according to current hierarchical models of face patch function but has previously shown distinct responses from other high-level face patches Importantly, these responses have included a particular sensitivity to spatial components of faces such as size or closeness during naturalistic movies

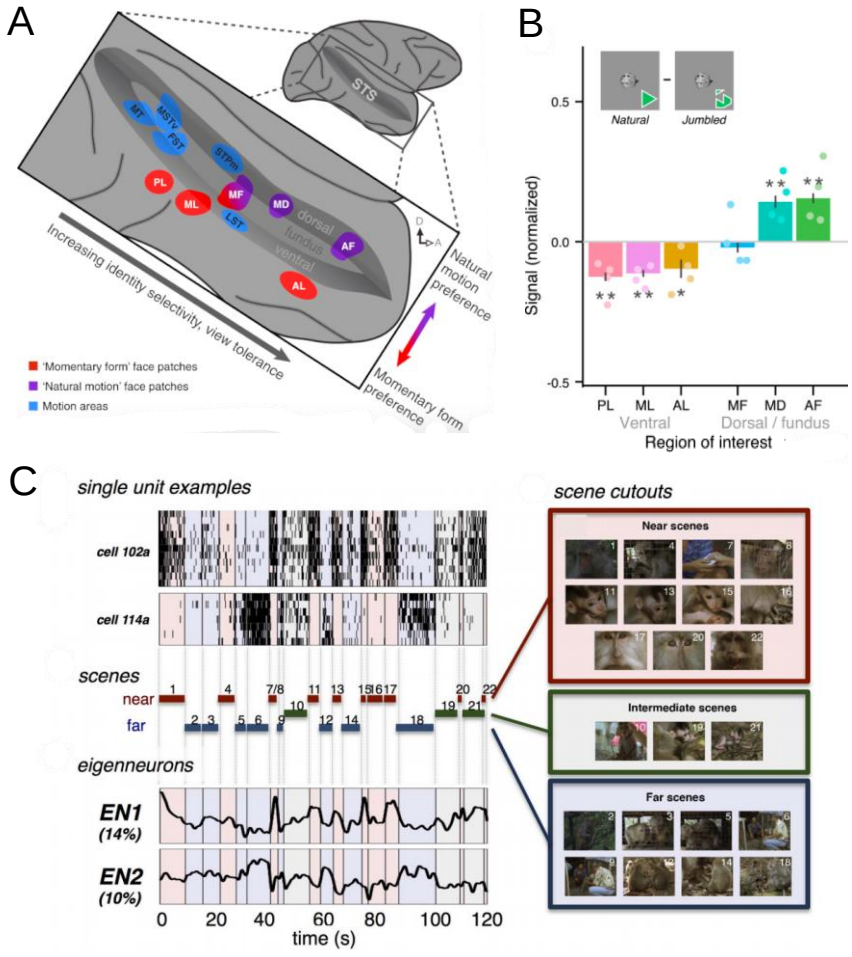

Figure 2.9 Spatial and Motion Responses in AF Face Patch. A: A schematic of the face patch system indicating that more dorsal and fundus face patches may be part of a facial motion processing system. **B:** A comparison of the signal change in fMRI for natural motion as contrasted with jumbled motion of a face in face patches demonstrating more dorsal face patches such as AF show a greater preference for natural motion. **A, B** adapted from (Fisher and Freiwald 2015) with permission. **C:** Responses and eigenneurons of single neurons in AF patch that demonstrate a preference of certain distances, near and far, in naturalistic movies. Taken from (McMahon, Russ et al. 2015).

(McMahon, Russ et al. 2015) (Figure 2.9C). Moreover, it lies near to disparity tuned regions of the macaque cortex that may influence its responses to spatial properties (Verhoef, Bohon et al. 2015). Thus, the AF face patch provides a model region to explore naturalistic, multisensory, and spatial elements of perception that may elucidate these principals for association regions across species. How these three features integrate within this single region and indeed across the brain, remain poorly understood. But the properties of audiovisual integration in the STS as well as lateral temporal areas in all species require further examination, particularly in relation to audiovisual space.

To explore these facets of AF face patch, we first examined audiovisual integration directly within macaque face patches. We recorded from the AF face patch located in the fundus of the STS to determine if fundus face patch cells respond uniquely to audiovisual stimuli and compare it with more inferior temporal face patches. Macaques underwent an fMRI localizer to find the position of these patches and surgically target them for electrophysiology recordings. Subjects were surgically implanted with a microwire bundle, which enabled long term recording from a single neuronal population (McMahon, Jones et al. 2014). These microwire bundles enabled long term recordings without repeated electrode penetrations. While recording from these patches, we presented audiovisual movies of unfamiliar macaques performing a range of vocalizations and examine the effect of audiovisual stimuli on the spiking response of face patch neurons. We also utilized a series of control stimuli including calls by an animated 3D macaque avatar derived from computed tomography images of real macaques that removes effects of identity and contrast of the original clips (Murphy and Leopold 2019). We compared the single cell responses to this face-like stimulus to an expanding dot, and temporally matched broad band noise to evaluate audiovisual modulation and some of the principals that determine these responses. Moreover, given the face-

selective audiovisual responses of connected areas, these control stimuli will evaluate the effect of visual selectivity on audiovisual integration in the temporal lobe. We expected that AF face patch would show audiovisual responses with face specificity.

2.4.3 Absolute Size Tuning in AF Face Patch

We also examined spatial elements in AF face patch, particularly the tuning of neurons to absolute size over retinal angle tuning. Recording neurons from the face patch with the same method, we presented stimuli using a 3D TV and place polarized lens in front of the subject to create stereo 3D. Using the avatar, we could then manipulate stimuli in 3D space and render them at a variety of physical sizes and virtual distances. We first, rendered stimuli at nine sizes and nine distance, matched to produce a subset of stimuli that had identical retinal angle. Through this set of stimuli, we could deconvolve tuning for absolute size from tuning for retinal subtense by examining the variation of single neuron spike rate across both factors. We hypothesized that the physical size would shape tuning more than the retinal angle. We then rendered stimuli at a variety of sizes from unrealistically small to unrealistically large to examine the extent of physical size tuning compared to more average face sizes, with the expectation that the average sizes would yield the strongest responses. Finally, we explored if absolute size tuning in face patches extend to non-face objects including unfamiliar objects such as a fork or house, familiar objects such as apples or bananas, and unfamiliar animals such as a goat and an elephant. Through these stimuli, we could examine if tuning for size extends beyond faces in the STS.

2.4.4 Effects of Spatial Manipulations on Audiovisual Integration in AF Face Patch

Then to combine these elements, we manipulated a series of spatial elements of both the auditory and visual stimulus. We initially transformed visual spatial element utilizing the 3D macaque avatar, including head angle, virtual distance, and absolute size by presenting stimuli in stereo. Combining these elements, we evaluate how these spatial elements impact audiovisual responses in single neurons. Then, we employed a 3D virtual reality dome that enabled a wider and larger field of view for visual stimuli. We attached an array of fifteen speakers to the dome before covering them with an acoustically transparent screen. We again recorded from AF face patch and utilized the macaque avatar along with other naturalistic stimuli. With this setup, we introduced spatial shifts between the sensory modalities of the stimulus and investigate how these spatial misalignments affect neurons in this region. Visual stimuli were presented at a variety of positions relative to the subject and auditory stimuli were presented at locations that could align or misalign with these visual components. Using this combination, we could more easily examine visual and auditory space as well as the principals, which determine their combination at single cell level.

2.4.5 Comparison of Carnivore and Primate Temporal Lobe

The comparative neuroscience of audiovisual integration and social perception within the mammalian cerebral cortex remains mysterious. Comparative anatomy and physiology can begin to outline overall principles of multisensory integration and further decode how these principles function in multiple species. While studies have investigated numerous species individually, few have directly compared these principles by which different species combine multisensory information or social perception. Previous comparisons between the carnivore and primate brain have principally examined the relationship of the cat and the macaque brain. Neurons in cats and

macaque cortex exhibited similar principles of multisensory integration between numerous regions including portions of macaque STS (Stein, Meredith et al. 1993). However, this work has yet to examine audiovisual integration or naturalistic stimuli, particularly across space. Ferret and cat visual areas have also demonstrated strong homology in visual areas indicating these comparisons may still apply within the ferret brain and provide a guide to comparison between the ferret and macaque (Homman-Ludiye, Manger et al. 2010).

Predominately visual regions of the ferret brain, similar to the macaque, have also received relatively little examination for multisensory properties. Among visual areas examined, area 21, thought to be a carnivore homologue or analogue of area V4 in the macaque, showed little audiovisual response despite strong anatomical connections to auditory areas, indicating connections may not always elicit response (Allman, Bittencourt-Navarrete et al. 2008). Similar connections have been examined in V1 of macaques but have similarly yielded few functional results (Falchier, Clavagnier et al. 2002). However, these results may further reflect homology between the carnivore and primate brain that remains unexplored. PPr and PPc of the ferret cortex similar to ventral intraparietal cortex (VIP) of the macaque responds to visual and touch information (Avillac, Ben Hamed et al. 2007, Meredith, Allman et al. 2009). This multisensory integration in an anatomically similar region again indicate homology in multisensory regions across species and may indicate similarity between the ferret and macaque brain. Area 20b has shown extensive connections with nearby auditory regions particularly with posterior ectosylvian gyrus (PEG) and preceding visual regions including areas, 18,19, and 21 (Bizley, Nodal et al. 2007, Dell, Innocenti et al. 2019) much like the macaque STS (Figure 2.10). Along with area 20b's

connections to visual and auditory regions, it also connects with PPr and PPc providing more multisensory information and further connects with PMLS, which also indicate that area 20b likely receives some level of visual spatial information. Combined, this evidence suggests that area 20b, like the STS, may express audiovisual modulation in a manner that favors more naturalistic visual or auditory stimuli and include certain spatial elements.

Thus, we hypothesized that ferret area 20 may have face selective neurons. To properly evaluate the tuning of ferret area 20 and to compare it with the macaque STS, we trained ferrets to maintain a position in front of monitor while passively viewing a set of visual stimuli. Few carnivore species have been examined for any kind of visual or naturalistic selectivity, so these stimuli included ferret faces, human faces, ferret bodies, objects, and scenes. We recorded from

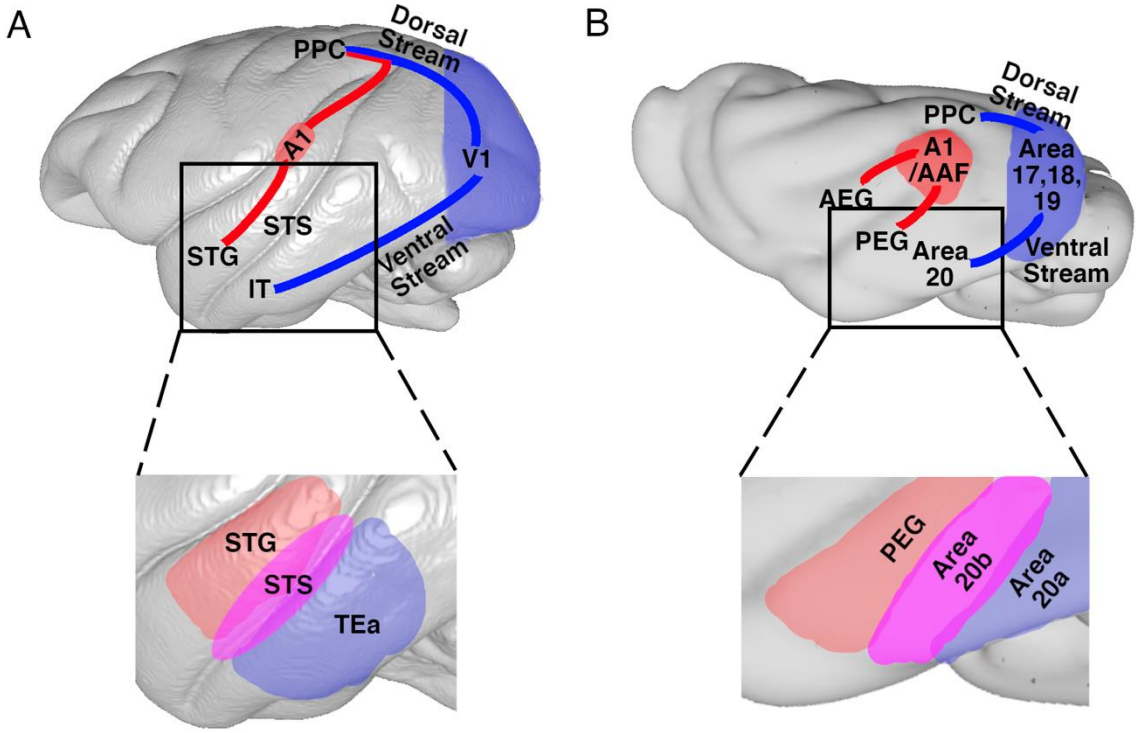

Figure 2.9 Comparison of Audiovisual Temporal Lobe of Ferret and Macaque. A. Schematic of macaque visual and auditory pathways with a cutout of the temporal lobe including the multisensory region of the STS. **B.** schematic of analogous regions of the ferret cortex with the cutout emphasizing the potentially multisensory region of area 20b. **A, B** were derived from brain atlases based on data from (Rohlfing, Kroenke et al. 2012) (**A**) and (Hutchinson, Schwerin et al. 2017) (**B**)

ferret area 20 and other nearby temporal cortex visual regions to evaluate if these regions could show visual selectivity for these naturalistic stimuli. This experiment investigated important comparative elements of social processing and can further establish the ferret as a model for evaluating vision or multisensory integration in more naturalistic conditions.

3. Audiovisual Integration in Unexamined Areas

3.1 Introduction

This work has been published in Current Biology: Audiovisual Integration in macaque face patch neurons, (Khandhadia, Murphy et al. 2021)

This chapter aims is to establish the combination of audiovisual information with high-level visual information in the temporal lobe. In humans and other primate species, audiovisual integration plays an important role in social communication, for example during the perception of a conspecific's vocalization and concomitant facial behavior (Ghazanfar and Santos 2004, Barraclough and Perrett 2011). The temporal cortex, and particularly the superior temporal sulcus (STS), contain zones of convergence for high-level sensory signals (Beauchamp, Argall et al. 2004, Beauchamp, Lee et al. 2004, Barraclough, Xiao et al. 2005, Ghazanfar and Schroeder 2006). In the macaque, the STS fundus borders high-level visual and auditory cortex (Seltzer and Pandya 1978, Pandya and Seltzer 1982, Hackett, Stepniewska et al. 1998, Kaas and Hackett 2000) and exchanges connections with other multisensory areas, including ventrolateral prefrontal cortex and intraparietal cortex (Pandya and Seltzer 1982, Seltzer and Pandya 1994, Romanski, Bates et al. 1999). At the single cell level, neurons within portions of the STS respond to visual and auditory stimuli, as well as their combination (Benevento, Fallon et al. 1977, Bruce, Desimone et al. 1981, Baylis, Rolls et al. 1987, Hikosaka, Iwai et al. 1988, Barraclough, Xiao et al. 2005). Functional MRI (fMRI) investigation of the macaque temporal cortex has also revealed a number of operationally defined regions named according to their visual category selectivity, such as face and body patches (Perrett, Hietanen et al. 1992, Kanwisher, McDermott et al. 1997, Tsao, Freiwald et al. 2003, Tsao, Moeller et al. 2008, Fisher and Freiwald 2015). In macaques, face patches are replete with cells that respond more strongly to faces than to other

categories of images (Tsao, Freiwald et al. 2006, Bell, Malecek et al. 2011, Aparicio, Issa et al. 2016) and form an interconnected network (Moeller, Freiwald et al. 2008, Grimaldi, Saleem et al. 2016). A subset of these patches lies along the STS and are coextensive with known multisensory regions in the fundus (Baylis, Rolls et al. 1987, Barraclough, Xiao et al. 2005). However, despite intensive study of neurons within the visually defined face patches, it is presently unknown whether or not they participate in multisensory integration.

Here, we investigated audiovisual single-unit responses in two fMRI-defined face patches. The anterior fundus (AF) and anterior medial (AM) patches were selected as key candidate regions for investigation, as they are both thought to occupy high-level positions in the face-processing hierarchy but are situated in distinct portions of the temporal cortex (Freiwald and Tsao 2010, Fisher and Freiwald 2015). Area AF is located in the STS fundus, within regions known to contain multisensory neurons while AM is located on the undersurface of the temporal lobe surface adjacent to, and interconnected with, the perirhinal and parahippocampal cortices, which also receive multisensory information (Grimaldi, Saleem et al. 2016, Miyashita 2019). After identifying these patches based on their selective visual fMRI responses to faces, we recorded the activity of individual neurons within each patch to brief movie clips of macaque vocalizations, including the full audiovisual stimulus as well as the visual and auditory components alone. The results demonstrate that auditory information prominently influences the responses of AF neurons but has little effect on the responses of AM neurons. We then further evaluated the audiovisual modulation in AF with control experiments. These control experiments demonstrate that auditory modulation is specific to faces and depends on the temporal, rather than spectral, structure of the acoustic stimulus. We discuss the findings in relation to the layout

of the macaque face patch network, macaque temporal lobe, and its intersection with known audiovisual cortical areas.

3.2 Methods

3.2.1 Subjects

Four rhesus macaque monkeys designated SP (Monkey 1, 9 kg, female), SR (Monkey 2, 10 kg, male), W (Monkey 3, 11kg, male), and M (Monkey 4, 9 kg, male), were implanted with a single chronic microwire bundles fixed within a custom MRI-compatible chambers and microdrive assembly. The electrode bundles were advanced post-surgically to achieve proper depth of recording. The electrodes in Monkey SP were located in the right hemisphere, whereas those in monkeys SR, M, W were in the left hemisphere. In accordance with the principles of the 3Rs, the fewest number of animals possible were used and the sampling technique of the microwire bundles enabled to repeated recording from the same neurons rather than repeated penetrations, which would cause more neural tissue damage. All procedures were approved by the Animal Care and Use Committee of the National Institute of Mental Health.

3.2.2 fMRI

Functional and anatomical magnetic resonance imaging (MRI) was conducted in the Neurophysiology Imaging Facility Core (NEI, NIMH, NINDS) using a vertical 4.7T Bruker Biospin scanner. For all subjects, hemodynamic responses were enhanced by injection with monocrystalline iron-oxide nanoparticles (MION). Details of scanning and stimulus presentation are described further in (McMahon, Jones et al. 2014). Briefly, Monkey SP underwent a standard block design localizer consisting of 24-second blocks of images of static macaque faces contrasted with blocks of images of non-face objects. In monkeys SR, M, and W, the blocks

consisted of short movie clips of macaques making facial expressions contrasted with short movie clips of moving scenes and moving objects. Subjects received a juice reward for maintaining fixation every 2s. All fMRI data was analyzed with AFNI (Cox 1996) and custom software developed in MATLAB (Mathworks, Natwick, MA).

3.2.3 Experiment Design

All subjects performed a viewing task. The subject initiated the trial by fixating on a 0.7° crosshair within a window of 2° visual angle for between 200-300ms. A stimulus was then presented in either an audiovisual, visual only, or audio only format. For the audiovisual and visual conditions, a visual stimulus appeared in a square 10° visual angle window, and the subject was allowed to freely view any part of the stimulus. For the audio only condition, the fixation marker remained on the screen and the subject was required to maintain fixation to ensure the subject remained engaged in the task and maintained eye position in a similar position to the other conditions. An infrared camera (Eyelink II, SR Research) monitored the subject's gaze as it performed this task and trials were aborted if the subject looked outside the window for longer than 100ms. All visual stimuli were presented on an OLED 4k Monitor 95cm from the subject using a graphical user interface (GUI) derived from PLDAPS (further described here (Eastman and Huk 2012)) in MATLAB. All auditory stimuli were presented in mono from two Tannoy Reveal Speakers placed on the edges of the monitor to create the percept the sound originated from the center of the screen. All auditory stimuli were projected at 65-80 dB SPL, verified using a Brüel and Kjaer (Denmark) sound level meter and at the full frequency range available.

3.2.4 Stimuli

In Experiment 1, stimuli consisted of short movies of monkeys vocalizing. The short movie clips featured macaque calls of varied acoustic structure and with a range of referential meaning and valence including affiliative coos, aggressive pant-threats, barks, and bark-growls, and agonistic/submissive screams (Gouzoules, Gouzoules et al. 1984, Hauser 1991, Hauser and Marler 1993, Romanski, Averbeck et al. 2005). Of these calls, the agonistic and coo calls used here had tonal/harmonic elements, while the remaining calls were broadband (Romanski, Averbeck et al. 2005).These movies featured three individual monkeys at a variety of head positions (Figure 3.1C).

For Experiment 2, we selected five vocalization movie exemplars that represented each of the different call types described above. We then matched the mouth movements of the computer-generated animated macaque avatar (Murphy and Leopold 2019) to the vocalization onset and envelope in each call to create new audiovisual movies (Figure 3.1D). The avatar enabled us to hold the basic visual appearance of a macaque face constant, including its identity and 3D head orientation, while its facial actions and mouth movements were animated and synchronized with the true macaque vocalizations. For this, we independently controlled features such as size of the mouth opening and amount of lip motion from the original movies using a GUI developed in MATLAB and animated the macaque avatar to follow the same patterns of motion. Movies of these avatar-vocalizations were rendered and compiled using the software Blender (the Blender Foundation). Frames were added to the avatar clips to extrapolate starting from or returning to a neutral facial expression before or after the vocalization.

In addition to the avatar stimuli, the monkey subject was presented with two categories of other control stimuli to determine the selectivity of audiovisual responses. Both sets of controls maintained the original temporal dynamics of the call structures. The first set controlled for

visual motion. We replaced the macaque movie video with a dynamic disc whose instantaneous size was matched to the amplitude of the movements of the monkey's mouth in the original movie (Ghazanfar, Maier et al. 2005, Sugihara, Diltz et al. 2006). This control was designed to determine whether auditory modulation is face-specific or would be found with any temporally synchronized visual stimulus. In the second set of controls the audio track was manipulated by replacing the original spectral content with broadband noise convolved with the envelope of the original vocalization. This control evaluated the contribution of the spectral information on the acoustic modulation of neural responses.

To remove the contribution of known transient responses following the initial onset of a visual stimulus, all stimuli were introduced with a 500 ms static image prior to the onset of the movie movement and acoustic vocalization. Thus, the onset of the vocalization movie began after the face had already been on the display for 500 ms. This paradigm enabled us to deconvolve visual transient effects from the response to motion and addition of audio as well as approach a more naturalistic paradigm.

3.2.5 Electrophysiology Recording

Following fMRI localization of the relevant face patches, subjects were implanted with 64 channel NiCr microwire bundle arrays fabricated by Microprobes for extracellular recording. Monkeys SR and SP received implants in face patch AF (an overlay of functional activation in Monkey SP is shown in Figure 3.1A). Monkeys M and W were implanted in face patch AM (functional overlay of Monkey W is shown in Figure 3.1B). All recordings were conducted in a radio shielded room (ETS-Lingreen) with a RZ2 BioAmp processor (Tucker-Davis Technologies) with a 128-channel capacity collecting a broadband signal of 0.5Hz-20KHz.

One feature of the microwire arrays is the capacity for long-term longitudinal recordings of individual neurons, verified through similarity in waveform and selectivity fingerprints across days. All spike sorting was performed offline. Spikes were sorted using the wave_clus spike sorting package (Quiroga, Nadasdy et al. 2004) and utilized the computational resources of the NIH HPC Biowulf cluster (http://hpc.nih.gov). To ensure cells were consistent across days, monkeys viewed a "fingerprinting" stimulus set consisting of 60 images containing sets of face categories (monkey faces and human faces) and non-face categories (objects and scenes). Face cells maintain selective responses across days (McMahon, Jones et al. 2014); therefore, by evaluating selectivity to a consistent stimulus set we can combine cell responses across days and months for the same cell. Responses were matched principally based on the pattern of selectivity to the "fingerprinting" stimulus set as well as the spike waveform and basic distribution of interspike intervals. Responses to these stimuli were also used to compute a face selectivity index (FSI) to quantify the strength of selectivity (equation 1): FSI=(response_face – response_nonface)/ (response_face + response_nonface)

3.2.6 Data Analysis

Following spike sorting and concatenation of individual cell responses across days, data were analyzed using custom software also created in MATLAB. We isolated a response window between 500ms (the end of the still frame and the beginning of motion) and 100ms after the conclusion of the movie stimulus and a baseline window between -300 and 0ms before the onset of the still frame. These windows were used for all further stimulus analysis. We conducted a two-way analysis of variance test (ANOVA) comparing response for the presence and absence of each component stimulus against the baseline response (further described in (Sugihara, Diltz et al. 2006)) Neurons were classed as visual only or auditory only if they showed a significant

response to either of the component stimuli alone, linear multisensory if they showed a significant main effect of both modalities, and non-linear multisensory if they had a significant interaction term. We also combined all stimuli into a single combined ANOVA again with each modality as a factor to assess the effect of each modality overall. Additionally, for the spectral controls, we conducted a pairwise comparison using a Tukey-Kramer test between the spectral controls, natural vocalization, and the silent condition to directly compare the effect of different auditory components.

For both Experiments 1 and 2, the main analysis compared the response to the audiovisual stimulus to the corresponding response to the visual stimuli alone. To this end, we calculated an index of modulation (equation 2): Index of modulation = AV-V/AV+V. All rates were computed following baseline subtraction. Here, AV is the mean response to an audiovisual stimulus for a particular cell, whereas V is the mean baseline subtracted response to the visual only counterpart of that stimulus. This index enabled us to quantify the magnitude of modulation induced by auditory stimuli with an index between 1 and -1 with a positive index indicating that the addition of acoustic stimuli enhanced the response and a negative response indicating that acoustic stimuli suppressed the response.

Finally, we created a linear mixed-effect model to determine the effects of face patch independent of individual cell responses. Taking the average across all stimuli for each modality, we examined the effect of face patch, presence or absence of visual stimulus, and presence or absence of auditory stimulus on the average spiking rate. Each of these factors served a fixed-effect variable whereas the different cells were classed as a random effect. Through this model, we could isolate the exact effect of face patch and its interaction with stimulus type independent of variance between cells.

3.3 Results

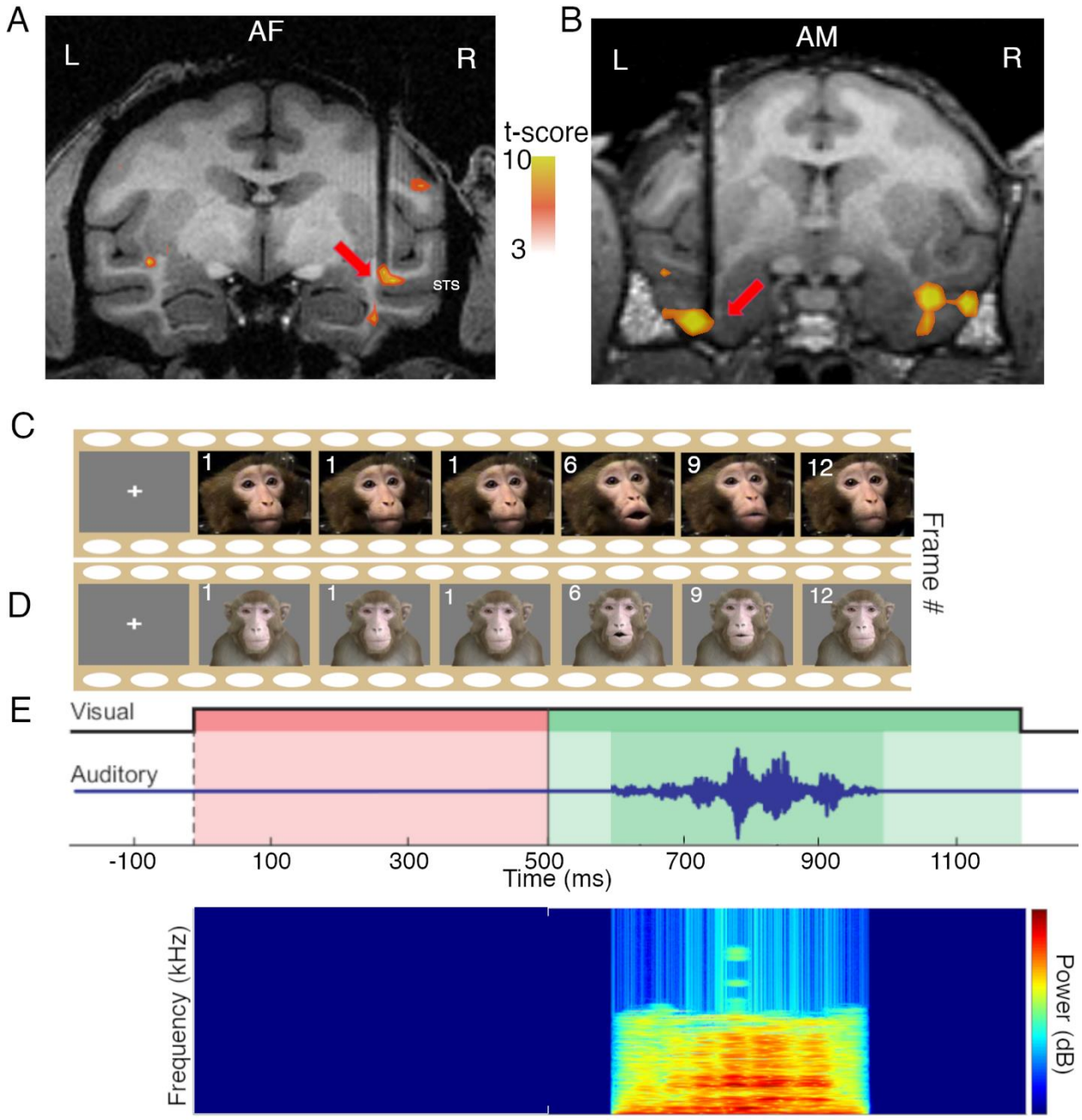

Figure 3.1. **Localization of recording sites in the STS and IT cortex. A, B**. Functional overlays of AF and AM from Monkey Sp and Monkey W, respectively, of an fMRI contrast of faces vs. objects. The tract of the electrode is indicated with the red arrow targeted to the desired areas of recording. **C.** Pant-threat vocalization from an unfamiliar macaque. **D.** The same vocalization as performed by the avatar both including the 500ms still frame indicated by the frames labelled 1. **E.** Presentation timeline of stimulus indicating the onset of the still frame indicated in red and the onset of the movie in green as well as the auditory stimulus including a spectrogram

We conducted extracellular recordings in fMRI-defined face patches in four adult macaque monkeys. Based on an initial fMRI mapping of face patches (see Methods), we targeted a single chronic 64 channel microwire electrode bundle into the centers of the AF or AM face Figure patches (Figure 3.1A, B). Each macaque received a single implant into a recorded face patch. We have previously demonstrated that this recording method supports longitudinal, stable recordings from the same cells over multiple sessions (Bondar, Leopold et al. 2009, McMahon, Jones et al. 2014). We recorded from 295 neurons in face patches of four monkey subjects: 240 from AF (125 from Monkey SP 115 from Monkey SR) and 55 neurons from AM (49 from Monkey W, 6 from Monkey M). In addition to the main experimental conditions featured in the study, subjects viewed a short "fingerprinting" stimulus set of static images each day, which included human and monkey faces, objects, and scenes. This daily dataset allowed us both to determine each neuron's face selectivity index (see STAR Methods) and to verify its identity across successive sessions (Bondar, Leopold et al. 2009).

Consistent with previous studies, the majority of neurons in both AF and AM were face selective. Specifically, 84.1% of all neurons (198/240 of AF neurons and 50/55 of AM neurons) responded to flashed faces with a face selectivity index (FSI) absolute value of greater than to 0.333, a criterion that has previously been used to categorize neurons as face-selective (Tsao, Freiwald et al. 2006, Aparicio, Issa et al. 2016) and both face patches show a distribution of FSI greater than zero ($t_{(118)}$=11.375, p=1x10^{-20} for AF, $t_{(54)}$=11.702, p=2x10^{-16} for AM). During the main electrophysiological experiments, the animals were required to maintain their gaze anywhere within the visual stimulus or, in the case of auditory only presentation, upon a small fixation marker. The dynamic component of the video was always preceded by a 500 ms static image of the face, corresponding to the first frame of the movie video. This presentation was

incorporated to diminish the contribution of abrupt visual transients during the period of audiovisual integration under study (Figure 3.1C, D, E). Subjects experienced 20-40 repetitions of each stimulus, receiving a juice reward after completion of each presentation.

3.3.1 Experiment 1: Multisensory Responses of AF and AM Face Patch Neurons

The goal of the first experiment was to determine whether the addition of the auditory component of the vocalization influences the responses of neurons in the two face patches. Subjects were presented with fifteen dynamic natural movie clips of three unfamiliar monkeys issuing five different call varieties of differing emotional valence. The call types included eight affiliative coos, two agonistic tonal screams, two aggressive pant-threats, two barks, and one bark-growl (Figure 1C) (Gouzoules, Gouzoules et al. 1984, Hauser 1991, Hauser and Marler 1993, Romanski, Averbeck et al. 2005). The acoustic structure ranged broadly, with coos and agonistic calls having more tonal elements and barks, pant-threats, and bark-growls having a broadband and atonal structure (Romanski, Averbeck et al. 2005). Trial sequences consisted of randomly interleaved presentations of each original audiovisual movie, the visual component only (i.e. silent movie), and auditory component only.

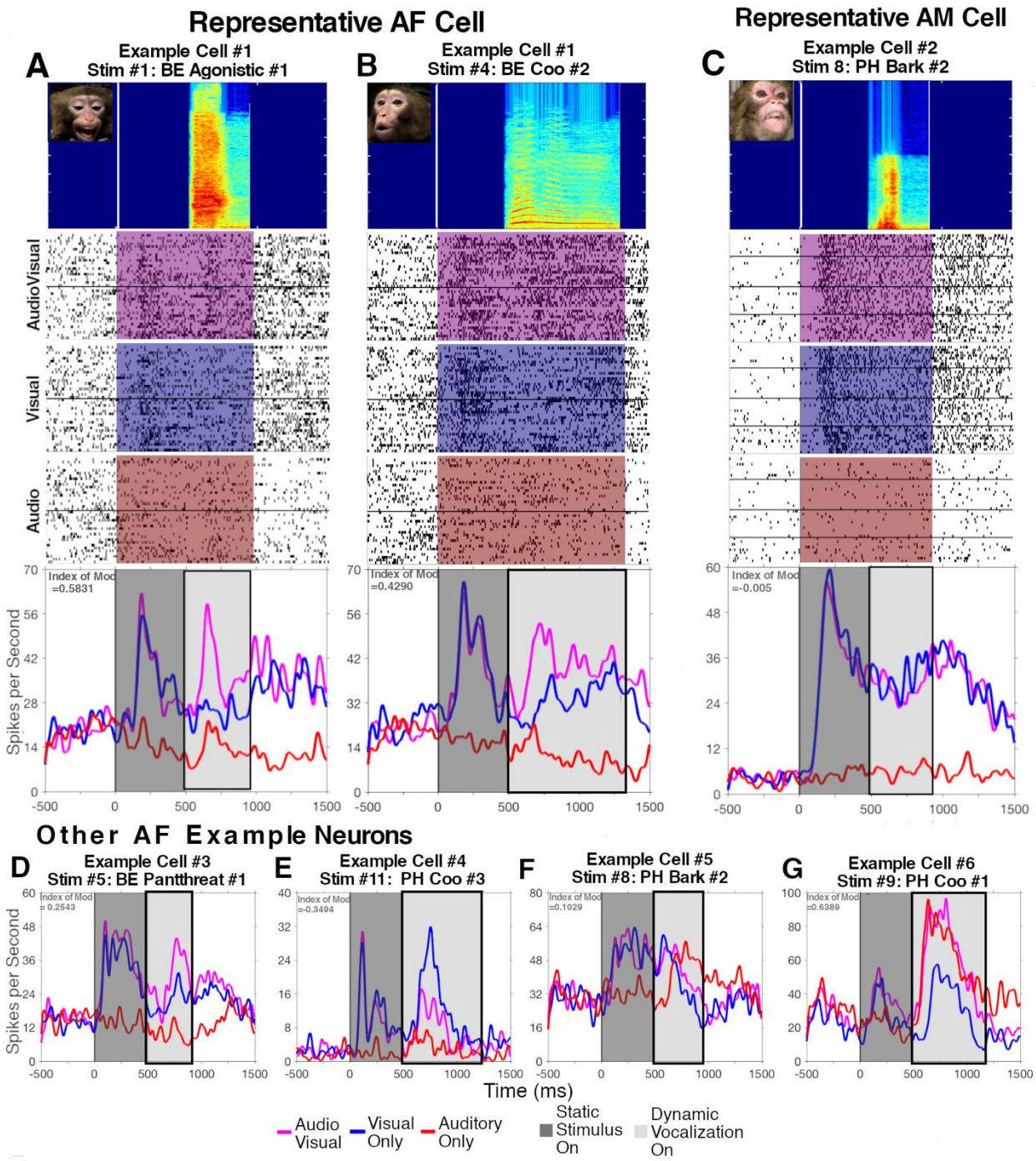

Figure 3.2. **Example Responses from AF and AM**. The dark gray panel indicates the static frame while the light gray indicates the audiovisual movie stimulus (magenta), silent movie (blue), or vocalization (red). **A,B.** Typical enhancement of AF neuron's response for two different stimuli (2-way ANOVA, A. p Vis<0.0001, p Aud =0.3401, p Int<0.0001; B. p Vis<0.0001, p Aud=0.8686, p Int<0.0001). The horizontal black line within the rasters delineate the different recording sessions for the presented neurons**. C.** Typical AM neuron's response with little or no auditory modulation (p Vis<0.0001, p Aud=0.8014. p Int=0.2002) **D, E, F, G** Additional example AF neuron responses. **D** Another typical AF non-linear multisensory enhanced response (p Vis<0.0001, p Aud=0.7110, p Int<0.001). **E-G** Different profiles of audiovisual integration also expressed by neurons in AF. **E** A cell with a nonlinear suppression of spiking in response to the audiovisual condition compared to the visual only condition (p Vis<0.0001, p Aud=0.0086, p Int=0.001). **F,G** Bimodal responses, where the response to the audiovisual movie mirrored the response to a unimodal condition (visual in **F**, p Vis=0.0396, p Aud<0.0001, p Int=0.05, and auditory in **G**, p Vis=0.7750, p Aud<0.0001, p Int=0.2888) along with a response to the other unimodal stimulus. Response types were determined by two-way ANOVA considering the presence or absence of the audio and visual stimulus components and their interaction. The Index of modulation is shown in the corner of each SDF. See also Figure 3.3.

AF Face Patch Neurons: The majority of neurons in AF exhibited a significant auditory modulation of their visual responses in response to one or more of the vocalization stimuli. In addition, some AF neurons responded to the auditory component alone. The influence of acoustic information on AF responses took multiple different forms, which we qualitatively separated based on the characteristics of their response. The most commonly observed pattern was *multisensory enhancement* of the visual response (Figure 3.2A,B, D). For neurons in this category, the auditory stimulus alone did not elicit a significant response but did elevate the neurons' response to the visual movie. The prominence of this pattern across the population was evident in the auditory enhancement observed in the grand average activity across all AF cells and all stimuli (Figure 3.3A). A smaller number of neurons exhibited *multisensory suppression* (Figure 3.2E), where the auditory stimulus diminished the neurons' visual response. Finally, a relatively small subset of neurons did respond to one or more *auditory stimuli alone* (Figure 3.2F,G). These neurons were generally bimodal, meaning that they responded to both the

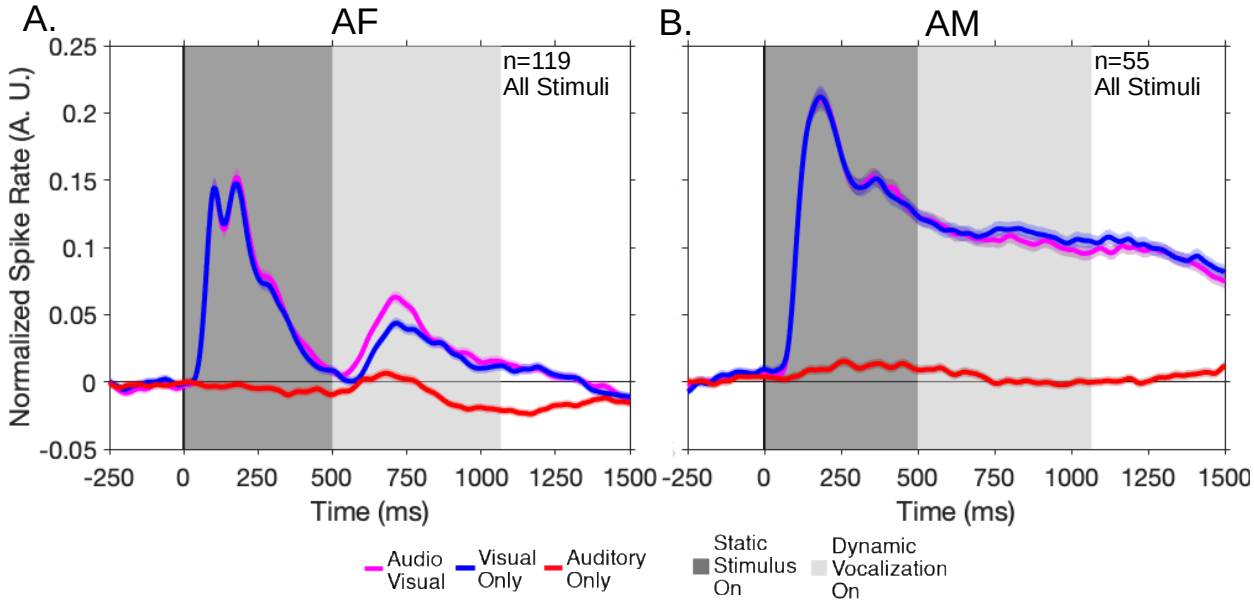

Figure 3.3 Grand average responses across all cells and all stimuli for face patches. A, B SDF plots of the grand average for AF, AM respectively (the SEM for each response is shown in a cloud surrounding each line with a very low SEM for AF)

auditory stimuli and the visual stimuli. For such neurons, the magnitude of their response to an audiovisual stimulus typically matched that to the visual stimulus alone, though a few matched that of the auditory stimulus alone.

To quantitively evaluate auditory responses and audiovisual interactions, we determined the average spike rate during the dynamic period of the movie, beginning at 500ms, after the static frame presentation and ending 100ms after the termination of the movie clip, which varied between stimuli. We conducted a two-way analysis of variance (ANOVA) on the spike rates for each movie separately or collapsed across all calls to determine the auditory or visual contributions, along with their interaction. Neurons were classified as visual if they showed a significant main effect only of the visual stimulus, auditory if they showed a significant main effect only of the auditory stimulus, linear multisensory if they showed a significant main effect for both the auditory and the visual stimulus, and non-linear multisensory if they showed a significant interaction term.

Based on this analysis across all calls, 76.0% of the 119 neurons recorded from the AF face patch were multisensory and exhibited a significant influence of the auditory component of the vocalization (2-way ANOVA $P<0.01$, Figure 3.4B top bar). Most prominently, 57.7% of neurons were classified as nonlinear multisensory, most often showing a significant modulation of the visual response by the auditory component. Another 14.4% of neurons were classified as linear multisensory, as they exhibited a significant response to both auditory and visual stimuli presented alone, together with a roughly additive effect during the audiovisual condition. Finally, 3.8% of the neurons responded *only* to the auditory stimulus. For each individual vocalization movie, auditory modulation was observed in a subset (24.2-50.6%) of neurons (Figure 3.4B, bars 1 - 15).

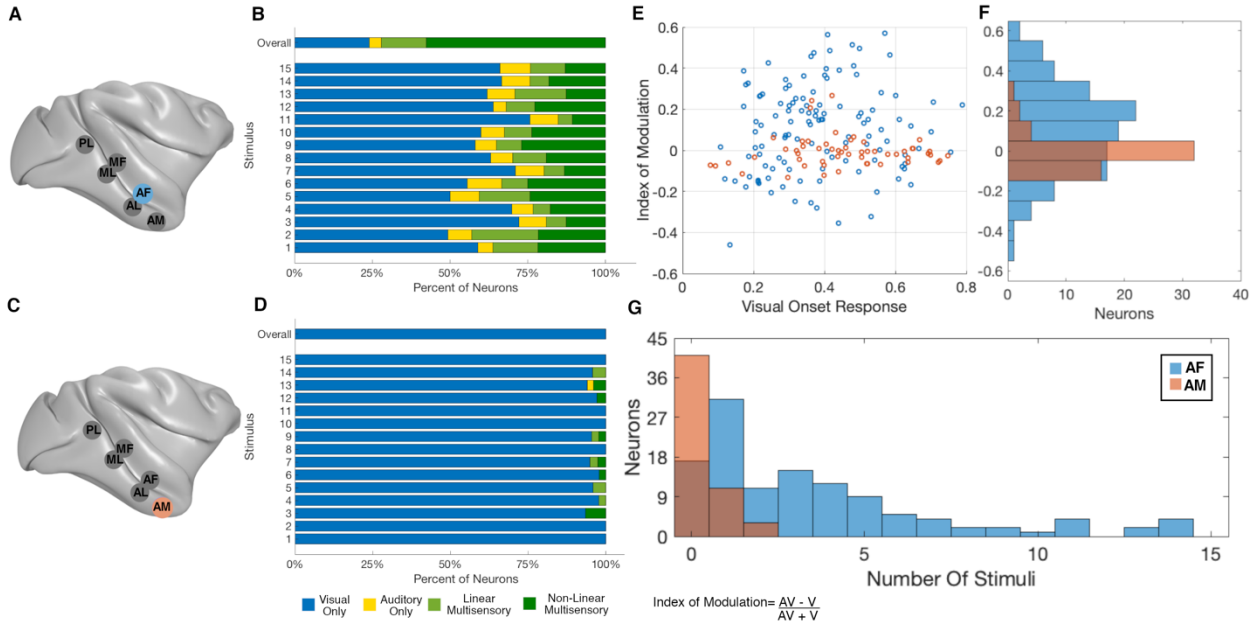

Figure 3.4 Comparison of population responses to audiovisual stimuli of AF and AM neurons. **A,C.** Schematic representations of the relative positions of all the face patches specifically marking AF (**A**) and AM (**C**). **B, D.** Plot of the proportions of neurons with significant modulation to each modality or the combination of modalities as calculated by 2-way ANOVA for all stimuli (top row) and each stimulus analyzed independently (lower rows) (AF, n=119; AM, n=55). E A scatter plot comparing the initial response the appearance of the still frame to the index of modulation for both AF and AM neurons **F** Distribution of the mean index of modulation for each neuron for AF (Blue) and AM (Orange); the black line marks 0 while the dashed blue line indicates the median of the AF distribution (0.1290) and the dashed red line indicates the median for the AM distribution (-0.0150). **G.** Distribution of neurons for which a given number of stimuli demonstrate auditory modulation. See also Figure S2 and Figure S3.

Multisensory cells showed considerable variation in the proportion of the 15 movie stimuli which elicited a response. Approximately one third of neurons (31/102, 30.5%) exhibited auditory modulation to only a single stimulus, whereas another one third of neurons (33/102, 32.5%) showed modulation to five or more stimuli (Figure 3.4G). To quantify the auditory effect on the visual responses, we calculated an index of auditory modulation (see STAR Methods), collapsing values across all fifteen stimuli for each neuron (Figure 3E,3F). The index values range from -1 to 1, where a negative index indicates an auditory suppression of the visual response and a positive index indicates an enhancement. The distribution of collapsed audiovisual index values for AF neurons centered around a median of 0.12, indicating a predominately enhanced spike rate modulation ($t_{(118)}$=5.317, p=5x10^{-7}). The index of modulation

revealed no strong preference for any particular stimulus across the population, although there were some differences between stimuli on average, indicating that affect or identity alone did not determine the responses of these regions. Further, the magnitude of acoustic modulation was not systematically related to visual responsiveness of the neuron (Figure 3.4E) and showed a non-significant relationship with the face selectivity index (Spearman correlation ρ=0.1110, p=0.2296). Together, these analyses demonstrate a prominent auditory modulation of visual responses to macaque vocalizations among AF face patch neurons, with the net effect being enhancement of selective visual responses across the population.

AM Face Patch Neurons: We performed the same analyses for neurons recorded from the AM face patch, which is known to receive direct input from AF as well as from other multisensory regions(Grimaldi, Saleem et al. 2016). In stark contrast to AF, few AM neurons were affected by the auditory component of the vocalization movies and the grand average across all AM cells and stimuli showed little if any audiovisual modulation (Figure 3.2C, 3.3B). This contrast was most clearly reflected in the ANOVA analysis across all stimuli of AM neurons revealing no significant auditory modulation of any neuron (Figure 3.4D). The audiovisual index across the population had a median of -0.02, which was not significantly different from zero ($t_{(54)}$=-0.6271, p=0.5332, Fig 3E, F). For individual movies, only a few AM neurons (n=14) showed significant auditory modulation. Of these neurons, 11/14 showed such modulation to a single stimulus with the remaining 3 cells showing significant modulation to two stimuli (Figure 3.4G). These results indicate auditory modulation in area AM is rare and, when present, highly selective for particular stimuli or very weak. The difference in auditory contribution to the AF and AM face patches was underscored by the results of a linear mixed-effects model that included the face patch and modality as its variables (Table 1).

Factor	Beta Coefficient (A.U)	T-stat	Degrees of Freedom	P Value
Presence of Auditory Stim.	0.0186	1.9288	518	P=0.0543
Presence of Visual Stim.	0.1321	14.4616	518	P<0.0001
Interaction of Face Patch and Auditory Stim.	**0.0339**	**3.87636**	**518**	**P=0.0001**
Face Patch	-0.0674	-3.8036	518	P=0.0002

Table 3.1. Display of the results of Linear Mixed-Effects model evaluating the effect of each factor on the average spike rate. The model shows a significant effect for the interaction of face patch and auditory stimuli with the positive beta indicating that AF neurons respond more strongly than AM neurons to the addition of auditory stimulus. See also Figure S3.

In summary, the results indicate that two high-level anterior face patches, AF and AM, differ sharply in their modulation by the auditory component of macaque vocalizations. The auditory influence in AF was conspicuous, widespread, and often extended to multiple stimuli, whereas that in AM was virtually nonexistent in our recordings. We next focused on the observed audiovisual modulation in face patch AF, and in particular, the requisite auditory and visual components of our stimuli.

3.3.2 Experiment 2: Investigation of Multisensory Responses using Macaque Avatar

To examine audiovisual processing in the AF face patch further, we used a realistic macaque avatar stimulus(Murphy and Leopold 2019), whose facial movements were programmed to mimic real facial actions during the specific vocalizations. The macaque avatar allowed for the investigation of particular aspects of audiovisual integration while maintaining the same face identity, head angle, and other visual stimulus properties. For Experiment 2, the avatar was animated to match five different calls (coo, agonistic, pant-threat, bark, and bark-growl) based on the original macaque movie clips. Now with this more controlled visual component of the stimulus, we investigated two questions related to the specific features important to the observed audiovisual responses of AF neurons.

Critical Role of the Face. We first asked whether the observed auditory modulation would differ if the visual stimulus were a face, now in avatar form, versus a surrogate non-face stimulus. Specifically, we compared responses elicited by a vocalizing avatar (Figure 3.5A) to those found when the face was replaced by expanding and contracting dynamic disk, whose movements were matched to the changing mouth size and synchronized with the auditory track (Figure 3.5B).

Neural responses to the avatar, including the modulation by the corresponding auditory stimulus, were broadly similar to the original movies, albeit with a smaller fraction of neurons demonstrating multisensory responses (Figure 3.5C). We recorded 121 AF neurons in this experiment, an independent population from those recorded in Experiment 1, and again used a two-way ANOVA to establish significant responses to each sensory modality. Of these neurons, 99/121 (81.8%) responded to at least one of the visual or auditory stimuli, while the remaining 22 were unresponsive to the experimental stimuli and excluded from further analysis. 26/99

(26.3%) neurons exhibited a significant response to the auditory component or significant auditory modulation to at least one of the five vocalization-movie call types (Figure 3.5D).

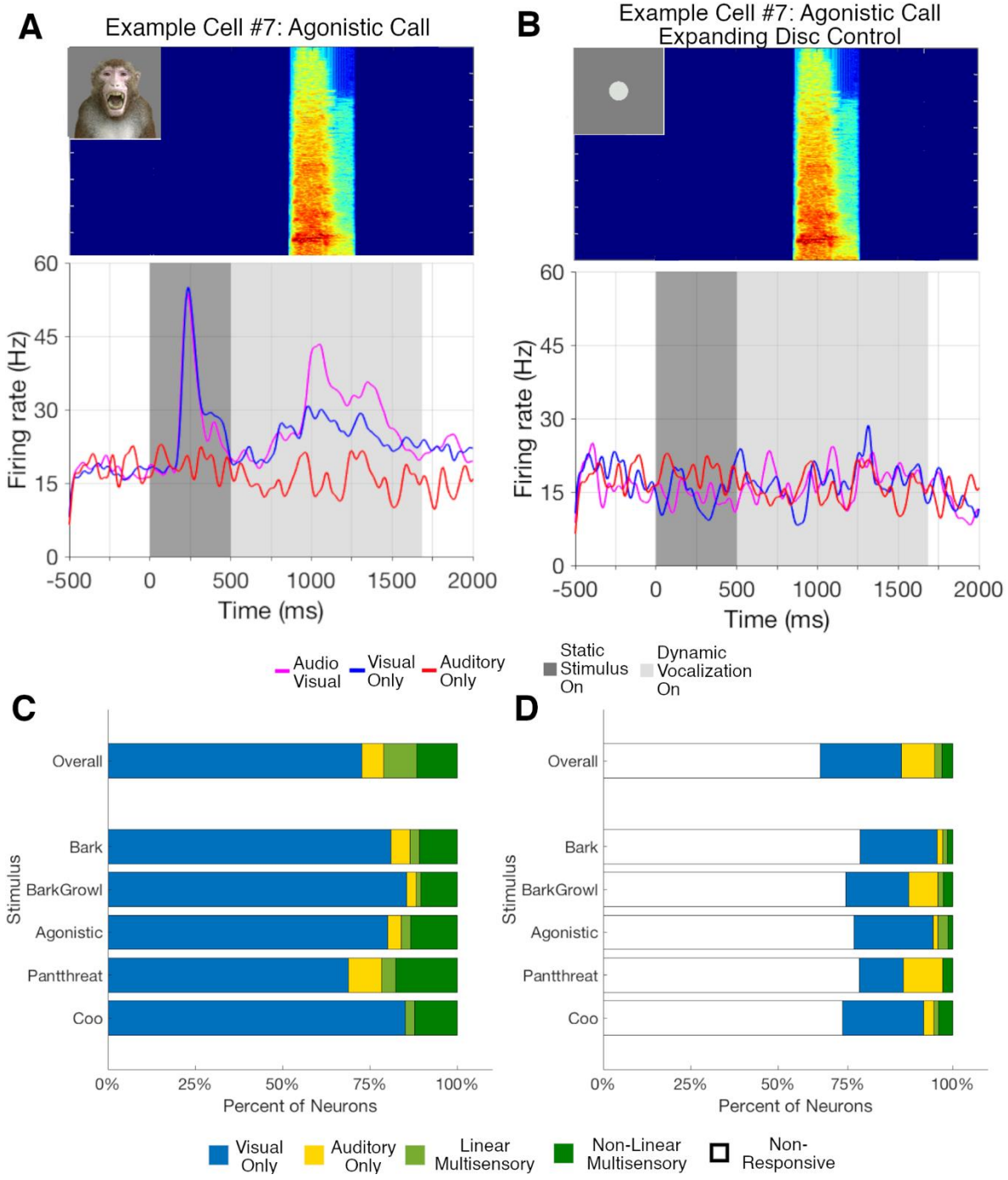

Figure 3.5. Responses to Visual Control Stimuli. **A, B**. Single cell example of responses to the different versions of the agonistic call stimulus. **A** portrays the average responses to the avatar producing an agonistic call while **B**, which shows the response of the same cell to the expanding disk stimulus matched to the same vocalization. **C,D**. Population response of AF to audiovisual avatar stimuli comparing **C,** the selectivity of cell responses to the audiovisual avatar stimuli to **D,** the selectivity of cell responses to the audiovisual expanding disk control stimuli.

The reduced proportion compared to the original faces likely reflects the imposition of a single avatar facial identity, as well as the lower overall number of stimuli. Importantly, very few neurons showed significant auditory modulation when the dynamic face was replaced with the dynamic disk (Figure 3.5B). Of the 99 neurons that responded to the avatar movie stimuli, only 5/99 (5.1%) neurons showed any linear or non-linear multisensory interaction the disk movie control (Fig 3.5D). However, the overall number of responsive neurons reduced from 99 to 46 neurons. To examine if this overall reduction contributed to the reduction in audiovisual neurons, we conducted a permutation test where we randomly selected 46 neurons from the neurons that responded to the avatar stimuli bootstrapping 1000 times to compare to the results of the dot stimulus presentation. This test showed only 1.2% of random samples yielded the same or fewer neurons with audiovisual effects. These responses suggest that the observed auditory modulation does not reflect a general temporal or spatial synchronization with visual movement, but instead depends upon viewing facial structure.

Critical Acoustic Parameters. We next used the same avatar stimulus to investigate the relative importance of spectral acoustic parameters in the modulation of visual responses. To this end, we repeated the experiment by pairing the avatar stimuli with temporally patterned broadband noise (BBN) by applying the temporal envelopes of the original calls to carrier noise (1-20,000 Hz frequency range), such that the temporal structure was preserved but the spectral content was disrupted.

Most AF neurons responded similarly to both normal audiovisual avatar stimuli and the matching audiovisual noise avatar stimuli (example shown in Figure 3.6A,B). Cells still responded to or were modulated by matched auditory noise despite the lack of detailed spectral information and did not differ significantly from the response to the normal vocalization (Figure

3.6C). A similar percentage (25/99, 25.3%) of neurons showed linear or non-linear modulation to the matched noise and 5.1% responded to the noise stimulus alone. To directly compare these different auditory conditions, we conducted an ANOVA with the natural vocalization and

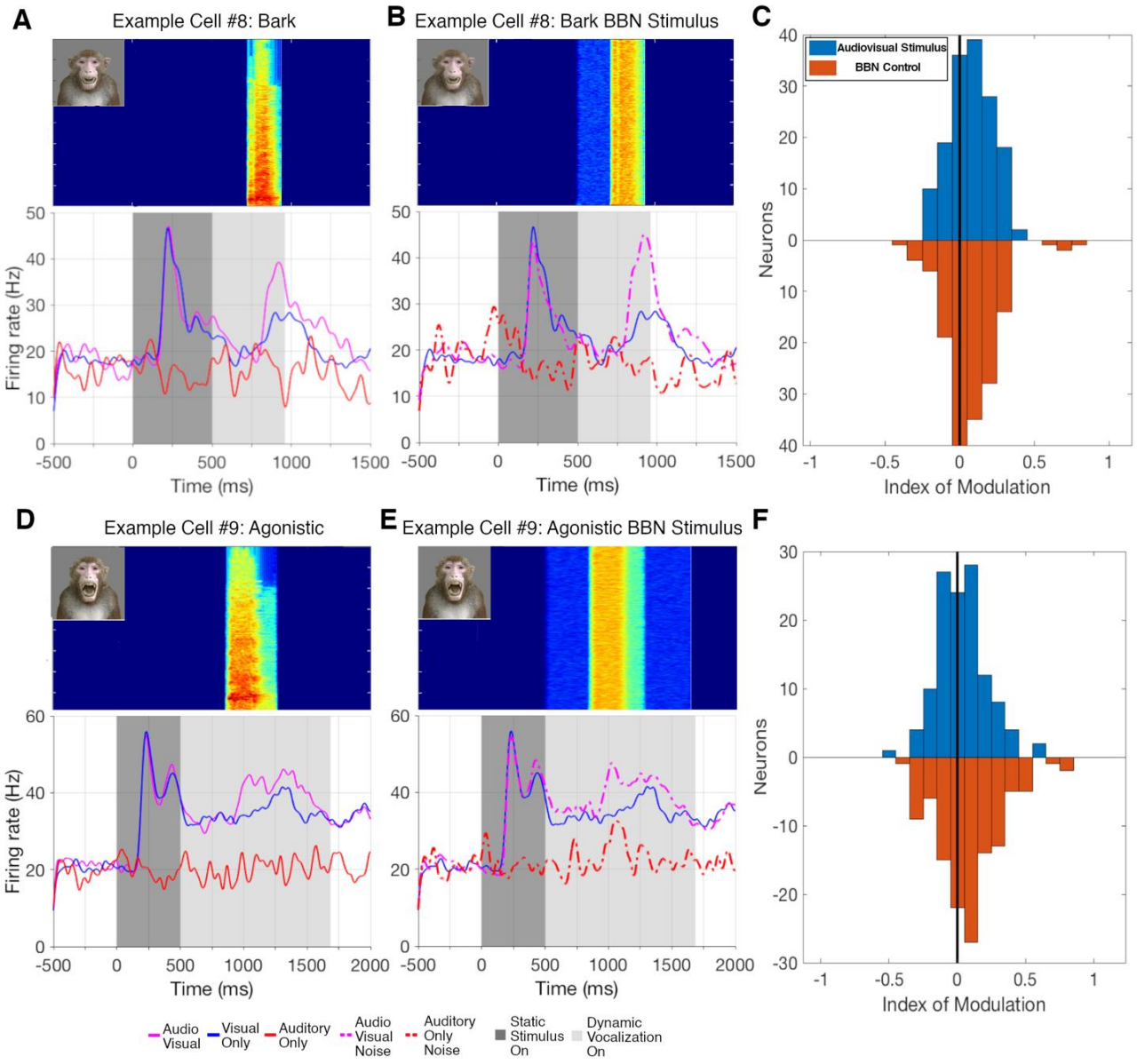

Figure 3.6. **Responses to Acoustic Control Stimuli**. **A, B**, A single cell example of responses to the different versions of the bark stimulus with **A** the average response of a single cell to the avatar bark stimulus and **B**, the response to the avatar when a temporally modulated broadband noise (BBN) stimulus replaced the bark vocalization. **C,** The distribution of the index of modulation for all calls across the population for both the avatar audiovisual stimuli and the avatar BBN control stimuli **D, E** A single cell example of the response to different versions of the agonistic call with **D** the cell response to the avatar agonistic stimulus and **E** the response to the avatar agonistic BBN stimulus. **F** The distribution of index of modulation to the tonal coos and agonistic calls.

matched noise as a factor and performed a post-hoc pairwise comparison with a Tukey-Kramer test. Only 7/121 (5.8%) of neurons exhibited a significant difference between the matching noise and the natural vocalization.

To ensure this similarity was not driven solely by sensitivity to broadband vocalizations, we compared responses to the tonal coo and agonistic calls with their corresponding broadband controls. Even in cases of harmonic calls, the neurons continued to show similar responses between the audiovisual conditions and the matched BBN conditions (example in Figure 3.6D,E), at similar levels across the population (Figure 3.6F). Similar proportions of neurons exhibited multisensory modulation to both the harmonic calls (29.4%) and their matched noise controls (35.4%). These results suggest that, despite sensitivity to fine visual features, multisensory modulation in AF face patch is principally determined not by the fine spectral details of a vocalization but by another feature, such as temporal or spatial structure, of an audiovisual stimulus.

3.4 Discussion

3.4.1 Audiovisual Modulation in Face Patches

These results indicate that most AF face patch neurons are affected by concomitant auditory stimulation during the viewing of macaque vocalizations. While previous research indicates that the anterior STS is a multisensory region (Benevento, Fallon et al. 1977, Bruce, Desimone et al. 1981, Beauchamp, Argall et al. 2004, Beauchamp, Lee et al. 2004) that contains cells with selective audiovisual responses to faces (Ghazanfar, Chandrasekaran et al. 2008, Dahl, Logothetis et al. 2009, Perrodin, Kayser et al. 2014), our data demonstrate, for the first time, this pattern is observed within a visually defined face patch. Though the predominant responses of

the recorded AF cells were visual, most showed some level of auditory modulation to movies within our limited stimulus set, and some cells responded to one or more auditory vocalizations in the absence of any visual stimulus.

Previous explorations of the STS organization in monkeys and humans have indicated a patchy spatial organization across primate species, with unisensory regions for each modality and audiovisual regions clustering together (Beauchamp, Argall et al. 2004, Dahl, Logothetis et al. 2009). In this context, the AF face patch might have been a good candidate for a visual-only region. However, our results instead suggest that this face patch may participate in audiovisual integration. It is possible that eye movement during the audiovisual condition may drive these responses but the similarity of the neural response during the static frame suggests that differences between the audiovisual and visual condition were driven by the auditory component. Previous investigation of face patches has indicated that face patch neurons remain consistent in response despite difference in eye position and movement, further reinforcing this conclusion, though we cannot rule of the role of attention. The diverse expression of multisensory responses was striking. For example, some neurons responded to both auditory and visual stimuli alone but, when presented with the combined audiovisual stimulus, responded as if only one or the other unimodal stimulus had been presented. Other cells responded to static faces and then responded only to the vocalization presented but not the moving face. Some of these results might be due to a high selectivity for individual identities, expressions, or other parameters of the movie, suggesting that estimates of multisensory responses would be greater with larger testing sets (Sugihara, Diltz et al. 2006).

The near absence of auditory modulation observed among AM neurons suggests that audiovisual modulation is expressed differentially among face patches. The contrast between AF

and AM is particularly striking given that AM receives direct anatomical projections from AF and responds when AF receives electrical microstimulation (Moeller, Freiwald et al. 2008, Grimaldi, Saleem et al. 2016). It bears mention, however, that the focal nature of our electrophysiological sampling means that our sampling of AM was limited and that we therefore cannot rule out a stronger multisensory component in other portions of AM that were missed in the two monkeys tested. It is also possible that AM neurons may be responsive to other sensory stimuli through the inputs they receive from neighboring perirhinal and parahippocampal areas, which are known to carry somatosensory information (Miyashita 2019). In contrast to AM, the connections of the AF patch have not been directly assessed with retrograde tracers, so its specific connections are unknown. In general, the STS fundus receives input from multisensory areas, such as intraparietal and prefrontal regions, as well as unisensory association areas, including high-level auditory belt and parabelt cortex as well as visual inferior temporal TE and TEO cortex, both directly and indirectly through lateral regions of the STS (Seltzer and Pandya 1978, Seltzer and Pandya 1984, Seltzer and Pandya 1989, Seltzer and Pandya 1989, Seltzer and Pandya 1994, Hackett, Stepniewska et al. 1998, Sugihara, Diltz et al. 2006, Cappe, Rouiller et al. 2009). Our results, combined with previous anatomical and electrophysiological finding, thus suggest that the AF face patch participates in multisensory integration that is typical for neighboring areas of the STS fundus.

Whether neurons in other faces patches integrate auditory information in a manner similar to AF remains to be seen. The specific pattern of interconnections among face patches, and their arrangement into one or more hierarchies, is presently a matter of inquiry (Freiwald 2020). Auditory sensitivity adds a new property to the response selectivity of AF neurons, whose response profiles and covariation with other brain areas is already quite varied(Park, Russ et al.

2017). Based on the known layout of the temporal cortex and its relationship to audiovisual responses, one might guess that other face patches lying in the fundus (MF) or upper bank (MD, AD) of the STS might be good candidates for audiovisual integration. Notably, these face patches, like AF, generally exhibit a sensitivity to facial motion (Fisher and Freiwald 2015), which may be central to the synchronization of visual and auditory information during a vocalization. By contrast, area AM on the ventral surface of the temporal lobe, is more commonly associated with processing of individual facial identities (Freiwald and Tsao 2010). It, like the recently described perirhinal (PR) and temporal pole (TP) areas involved in face familiarity (Landi and Freiwald 2017), may be less governed by dynamic facial behaviors and more by facial features. Frontal lobe face patches, (PO, PA, and PL) are known to respond to expressive faces similar to fundus patches (Tsao, Schweers et al. 2008, Taubert, Japee et al. 2020), and are coextensive with prefrontal cortical areas that have been shown to be responsive to vocal stimuli and to their combination with facial gestures, including many of the same stimuli used in the present study (Sugihara, Diltz et al. 2006, Diehl and Romanski 2014). These patches may also integrate audiovisual signals in a way that is yet to be elucidated. Further study of these functionally defined regions is needed, including investigation of their specific anatomical interconnections, their participation in multisensory integration, and their roles in reciprocal social communication.

3.4.2 Audiovisual Selectivity

Given AF neurons' selectivity for particular movies, the macaque avatar allowed for a controlled examination of key variables. The virtual absence of auditory modulation for the temporally synchronized dynamic disk stimulus is consistent with the assumed specialization for faces within the face patch network. These responses indicate that AF neurons specifically

combine auditory stimuli with facial information, rather than any temporally synchronized visual object. In previous studies, the temporal congruence between a visual stimulus and its auditory pair has been an important feature in multisensory integration in the STS while call type or spectral detail has shown little effect (Dahl, Logothetis et al. 2010, Perrodin, Kayser et al. 2014). Though we cannot confirm that these neurons are solely sensitive to temporal synchrony of the visual and auditory elements, we found that temporally synchronized auditory stimuli, even when applied to broadband noise, was sufficient to elicit auditory modulation of visual responses to a face. Thus, the relative unimportance of auditory spectral content compared to visual input may thus be a characteristic of multisensory integration in the fundus of the STS. Interestingly, nearly the converse was observed in an fMRI defined voice-specific area, a high-level auditory area on the supratemporal plane of the macaque temporal lobe. In that area, audiovisual neurons expressed selectivity for acoustic vocalizations while visual modulation of the acoustic response exhibited little selectivity to the visual stimulus (Petkov, Kayser et al. 2008, Perrodin, Kayser et al. 2011, Perrodin, Kayser et al. 2014). Together, these results suggest an overall principle of multisensory integration in high-level sensory areas, wherein highly stimulus-selective responses for the primary modality can be modulated by a relatively broad range of temporally synchronized stimuli presented in the secondary modality.

3.4.3 Broader Implications

This chapter establishes that this high-level visual area also combined with auditory responses in the temporal lobe. The face-dependent multisensory responses of AF neurons to vocalizations indicate that this area may participate in a larger cortical network in the service of social communication. For example, regions of the macaque auditory cortex also exhibit multisensory modulation tied to visual presentation of faces and face information, with the visual

response component likely arising from well-known reciprocal connections with the STS (Seltzer and Pandya 1978, Seltzer and Pandya 1994, Ghazanfar, Maier et al. 2005, Romanski, Averbeck et al. 2005, Kayser, Petkov et al. 2007, Kayser, Logothetis et al. 2010). Our results suggest that, within the STS, the AF face patch may be an important region involved in this processing. Projections between these areas may reflect a conserved multisensory pathway. Anatomical and physiological interaction between face-voice areas in humans are thought to mediate vocal communication (von Kriegstein, Kleinschmidt et al. 2005, Blank, Anwander et al. 2011). The STS also feeds forward to and receives feedback projections from the VLPFC, which also contains high proportions of face-selective audiovisual neurons (Sugihara, Diltz et al. 2006, Romanski and Hwang 2012, Diehl and Romanski 2014) and is itself thought to draw upon visual, auditory, and multisensory areas (Romanski, Bates et al. 1999, Romanski, Tian et al. 1999). Further studies are required to establish the specific anatomical and functional connections of AF with other multimodal, affective, and voice-selective regions and more importantly what role it plays within the larger range of multisensory areas. At present, the results draw attention to a well-known face-selective area in the temporal lobe, whose integration of vocal auditory signals into its visual analysis, makes it a likely contributor to primates' advanced skills in the domain of multisensory social perception. In Chapter 4, we further investigate how different kinds of visual information combine in AF face patch.

4. Perception of Absolute Size

4.1 Introduction

This work has been published in Proceedings of the National Academy of Sciences (PNAS): Encoding of 3D physical dimensions by face-selective cortical neurons (Khandhadia, Murphy et al. 2023)

This chapter aims to examine the influence of 3D object information in the temporal lobe and understand how spatial and form vision can combine in the temporal lobe. We experience the world in three-dimensional space, perceiving and interacting with objects and individuals in a scene. For humans and other primates, much of this experience is served by vision, with broad stretches of the cerebral cortex ostensibly devoted to making visual sense of the world. For example, individual neurons throughout the IT cortex of the macaque respond selectively to meaningful objects, with neurons of similar response properties often aggregated together in functional clusters (Lafer-Sousa and Conway 2013, Conway 2018, Hesse and Tsao 2020). One striking finding about the visual selectivity of IT neurons is its tolerance to natural image transformations, such as scaling and translation (Desimone, Albright et al. 1984, Sary, Vogels et al. 1993, Ito, Tamura et al. 1995, Wallis and Rolls 1997, Janssen, Vogels et al. 2000, DiCarlo and Cox 2007, Popivanov, Jastorff et al. 2015, Zhivago and Arun 2016, Taubert, Van Belle et al. 2018). Namely, if neural responses to stimuli are ranked in terms of that relative strength, that ranking often remains even if stimuli are scaled up several-fold in size. The preservation of selectivity is thought to reflect the capacity of the brain to compute a conceptual or abstracted representation of the retinal image separate from metric details. While the mechanism underlying this apparently intrinsic feature of ventral stream visual processing is poorly understood, it is thought to be critical

for image-based object recognition (Dittrich 1990, Wallis and Rolls 1997, Hung, Kreiman et al. 2005, Gothard, Brooks et al. 2009, Freiwald and Tsao 2010, Rust and Dicarlo 2010).

At the same time, even as the rank-ordering of neural responses is preserved, the absolute firing rate of a neuron can change dramatically with image scaling (Rolls and Baylis 1986, Ashbridge, Perrett et al. 2000, Op De Beeck and Vogels 2000). For example, when a set of stimuli is increased in size, a given neuron may double its firing rate to all stimuli, preserving its relative responses and hence its selectivity profile but changing its spiking output and thus its effect on downstream areas. The mechanistic basis of this rate modulation is also poorly understood and is seldom considered explicitly. One relatively unexplored possibility is that object selective neurons explicitly encode physical size. For example, the metric encoding of objects might serve to shape the brain's analysis of scene geometry and thus play a significant role in the visual operations underlying interactive with the environment. The brain may also store information about the typical sizes of objects (Flanagan, Bittner et al. 2008, Konkle and Oliva 2011). This internal knowledge can aid in perceptual judgements and manual actions that support primate behaviors such as foraging and social interaction (Leopold and Park 2020).

The visual encoding of geometric information is most frequently associated with parietal regions of the dorsal visual pathway, where coordinate transformations are thought to transform metric information about objects into effector actions (Andersen and Cui 2009). However, a few previous studies have suggested that neurons in the macaque ventral visual pathway are sensitive to some aspects of 3D geometry. At an early stage, neurons in area V4 adjust their responses to a given retinal image based on the absolute distance to the display (Dobbins, Jeo et al. 1998) and the volumetric 3D shape (Srinath, Emonds et al. 2021). At later stages, many neurons in the superior temporal sulcus (STS) are sensitive to volumetric shape, potentially reflecting their interplay with

intraparietal areas concerned with 3D visual geometry (Seltzer and Pandya 1978, Webster, Bachevalier et al. 1994, Janssen, Vogels et al. 2000, Van Dromme, Premereur et al. 2016, Janssen, Verhoef et al. 2018). These studies indicate that elements of spatial information may combine with object related information in high-level regions of the ventral visual stream. While these findings have enriched our understanding of 3D shape encoding, many open questions remain about how object-selective cortical regions contend with the complexity of natural vision, including the spatial and geometric components inherent in any real-world scene (Leopold and Park 2020). For example, to what extent are selective IT neurons shaped by the 3D size of an object and the spatial context in which it is encountered?

The present study investigates the sensitivity of macaque IT neurons to the physical size and distance parameters of an object. We recorded from a well-studied face-selective region of the STS known as the anterior fundus (AF) face patch (Moeller, Freiwald et al. 2008), where neurons are both selective for faces and highly sensitive to their scale (McMahon, Russ et al. 2015). We asked whether such scale-selective responses are driven primarily by the 2D image subtense of faces on the retina or the 3D physical geometry of the face and head in the world. In most visual electrophysiology experiments, retinal and physical geometry cannot be cleanly separated, as stimuli are presented as flat images on a display or fixed distance. Image scaling varies both retinal and physical geometry in concert. Moreover, absent other explicit depth cues, the physical correspondence of given retinal image is underspecified or ambiguous. To independently assess the contributions of retinal and physical geometry, we utilized a 3D macaque avatar stimulus (Murphy and Leopold 2019). We stereoscopically rendered a set of macaque faces at a selected range of size/distance combinations, including a subset size/distance combinations selected to yield images of the same retinal subtense (Figure 4.2A, B). We hypothesized and found that the

pronounced size modulation of AF neural responses principally reflected the physical 3D size of the face and head rather than the 2D size of the retinal image projection. We also hypothesized that neurons would prefer the most common face size but discovered that, in contrast to intuition, but consistent with some models of predictive coding, AF neurons responded most strongly to extreme sized faces that were much smaller or larger than a usual macaque face size. We discuss the potential role of physical size and natural geometry on the internal visual representation object structure in the primate brain.

4.2 Methods

4.2.1 Subjects

Two rhesus macaque monkeys, designated SP (Monkey 1, Female, 18yrs, 9 Kg) and SR (Monkey 2, Male, 6yrs, 10 kg), were implanted with a chronic microwire electrode bundle in AF face patch held by a custom MRI compatible chamber and microdrive. Electrodes were advance towards the target recording depth post-surgically to achieve stable recordings. Monkey SP was implanted in the right hemisphere while Monkey SR was implanted in the left hemisphere. All procedures were approved by the Animal Care and Use Committee of the National Institute of Mental Health and were conducted in accordance with the National Academy of Sciences Guide for the Care of Laboratory Animals and the NIH Animal Research Advisory Committee (ARAC) Guidelines. The NIH Animal Care and Use Program is accredited by AAALAC, International.

4.2.2 fMRI

All functional and anatomical magnetic resonance imaging (MRI) was conducted in the Neurophysiology Imaging Facility Core (NEI, NIMH, NINDS) in a 4.7T Bruker Biospin scanner. Subjects underwent a face-patch localizer to enable targeting of face patches. For all functional

scans, subjects received an injection of monocrystalline iron-oxide nanoparticles (MION) to augment hemodynamic response. Monkey SP underwent a standard block design consisting of 24 s blocks composed of images of static macaque faces contrasted with blocks of images of non-face objects (Koyano, Jones et al. 2021). Monkey SR viewed a dynamic localizer, which contrasted blocks of clips of macaques making facial expressions with clips of moving scenes or moving objects (Russ and Leopold 2015). The subject was rewarded every 2s for consistent fixation. All fMRI data were analyzed using AFNI and custom software created in MATLAB we have used in previous studies.

4.2.3 Stimuli

The face stimuli consisted of a 3D realistic macaque avatar presented at different sizes and stereoscopically defined distances (see Murphy and Leopold 2019). Briefly, the avatar was developed using computed tomography (CT) volumes of rhesus macaque monkeys averaged into a single avatar, which received computer generated (CG) texture and realistic fur in the 3D animation software and were rendered in Blender 2.79 (https://www.blender.org/). For the present study, we used custom Python scripts in Blender to systematically manipulate scale of the avatar with the virtual environment and its distance of the from the virtual camera This enabled us to present the stimuli at multiple physical locations and multiple 3D sizes, in many cases holding the retinal angular subtense constant.

We also presented a series of objects and animals to examine if tuning to sizes was exclusive to faces. We obtained meshes from online repositories that could be manipulated in the Blender software and preformed similar size manipulations to match sizes presented with the face stimuli.

4.2.4 Experimental Design

Figure 4.1. 3D object stimuli. The object stimuli matched to the size of the macaque face (from crown to chin) and presented to subjects as part of the same image set. The animal stimuli with faces were also matched in size with the macaque face. These stimuli were all presented in 3D to examine if different images or objects would similarly show tuning for the larger sizes of the macaque face.

All subjects performed a passive-viewing task. An infrared camera (EyeLink II, SR Research) monitored the position of the subject's gaze. Subjects began trials by fixating on a 0.7° fixation dot within a 3° window for between 200-300ms. The rendered face or object stimulus was then presented for 500ms with 500ms inter-stimulus interval in trials of 3 stimuli and all trials were aborted if the subject broke fixation for more than 100ms. The virtual avatar was positioned such that the cyclopean eye (the midpoint between the eyes) always remained at the center of the screen All stimuli were presented in side-by-side stereoscopic 3D on an OLED 3D TV (LG). A pair of 3D printed goggles holding polarized lenses was positioned in front of the subjects to create disparity and enable 3D presentation. All stimuli were presented using a graphical user interface (GUI) modified and derived from PLDAPS (Eastman and Huk 2012) in MATLAB.

When presenting the macaque avatar, we initially positioned the 3D TV screen at a viewing distance of 95cm from the subject. We initially presented 9 absolute sizes ranging from 8.7 cm, approximately two-thirds the size of an average macaque face, to 17.3 cm, approximately four-

thirds the size of an average macaque face separated by equal increments. Each of these sizes was presented at 9 virtual distances, ranging from 63.3cm to 126.7cm for a total of 81 stimuli. Importantly, these sizes and distances were matched such that 9 stimuli would subtend the same retinal angle but differ in their physical properties, enabling us to disentangle physical size from retinal size (Figure 4.2A, 4.2B). Extending this approach further in the second set of experiments, we created 20 absolute sizes of the avatar ranging from 1.3cm (one tenth the size of an average macaque face) to 26 cm, (double the size at increments of 1.3. The OLED monitor was positioned at 90cm in front of the animal, with stereoscopic cues dictating the virtual distance of each stimulus as well as its size and volumetric shape.

Aside from the macaque avatar, we also evaluated tuning of AF neurons for several additional 3D rendered animals and objects. These included familiar objects (banana, apple), unfamiliar objects (fork, cluster of rocks, soda bottle, house, watermelon) and unfamiliar animals (elephant, goat, butterfly) (Figure 4.1). These objects were selected to provide a range of shapes and real-world sizes. Each object and animal, including the. macaque avatar was stereoscopically displayed at nine physical sizes: 1.3cm, 5.2cm, 7.8cm, 10.4cm, 13cm, 15.6cm, 18.2cm, 20.8cm, and 26cm. For this experiment, stimuli were rendered volumetrically and presented at a single virtual distance of 90cm.

In addition to main experimental stimuli, we presented "fingerprinting" stimuli at the beginning of each session, containing 60 images with face (human faces and monkey faces) and non-face (scenes and objects) categories.

4.2.5 Electrophysiological Recordings

Following fMRI localization, the AF face patch in both monkeys received an implant of a 64 channel NiCr microwire bundle fabricate by Microprobes for extracellular recording. Following

implantation, the microwire was advanced towards AF face patch and its positioning confirmed with further MRI scans (Figure 4.2C). All recordings were conducted in a radio shielded room (ETS-Lingreen) with a RZ2 BioAmp processor (Tucker-Davis Technologies) with a 128-channel capacity collecting a broadband signal of 0.5Hz-20KHz.

4.2.6 Data Analysis

All data was analyzed in custom software designed in MATLAB. All spike sorting was conducted offline. All spike data was automatically sorted using the wave_clus (Quiroga, Nadasdy et al. 2004) spike sorting package with the resources of the NIH HPC Biowulf cluster (http://hpc.nih.gov). We calculated the mean spike rate for each neuron for each stimulus in a window from 50ms to 500ms after stimulus onset and subtracted a baseline value calculated from the mean spike rate between 200ms and 50ms before the presentation of stimulus and used it for further analysis. For normalized plots including tuning curves and spike density functions, we divided all responses for each neuron by the peak responses across all stimuli. For the initial set of nine sizes and distances, we used these responses to conduct a two-way ANOVA for each cell to determine the significance of the physical factors and to a conduct a one-way ANOVA to determine if the responses of neurons to the equal retinal angle stimuli changed significantly between stimuli of the same retinal subtence.

For this set, we also used these responses to calculate the model fit and model preference for each neuron to assess the relative contribution of size and retinal angle to the responses of neurons. To examine this, we first calculated the deviance of a linear multiple regression analysis of each cells including both physical size and retinal angle as factors. We then contrasted this value with the deviance of independent models for physical size and retinal angle with the following formula (equation 1):

$$\text{Model Fit} = 100\% * (D_{cons} - D_{test})/(D_{cons} - D_{Full}) ,$$

where D_{cons} is the deviance of a constant model of neural responses, D_{Full} is the deviance of the full multiple regression with physical size and retinal angle as factors, and D_{test} is the deviance of a single-factor regression with either physical size or retinal angle as a factor. From the model fit of both factors, we calculated the model preference (equation 2):

$$\text{Model Preference} = \text{Model Fit}_{phys} - \text{Model Fit}_{dva},$$

where Model Fit_{phys} is the fit relative to the physical size parameters and Model Fit_{dva} is the fit to the degrees visual angle parameters. A positive Model Preference indicates physical size explains most of the variance of the model whereas a negative Model Preference indicates retinal angle explains more of the variance.

Each neuron was separately evaluated for its face selectivity index. (FSI): (equation 3):

$$FSI = (R_{face} - R_{nonface})/(R_{face} + R_{nonface}),$$

where R_{face} is the average spiking rate response to face images and $R_{nonface}$ is the average spiking rate response to nonface images (for a plot of these look to Figure 4.2D) FSI was computed based on the response to a fixed set of 60 face and non-face "fingerprinting" stimuli, described above.

4.3 Results

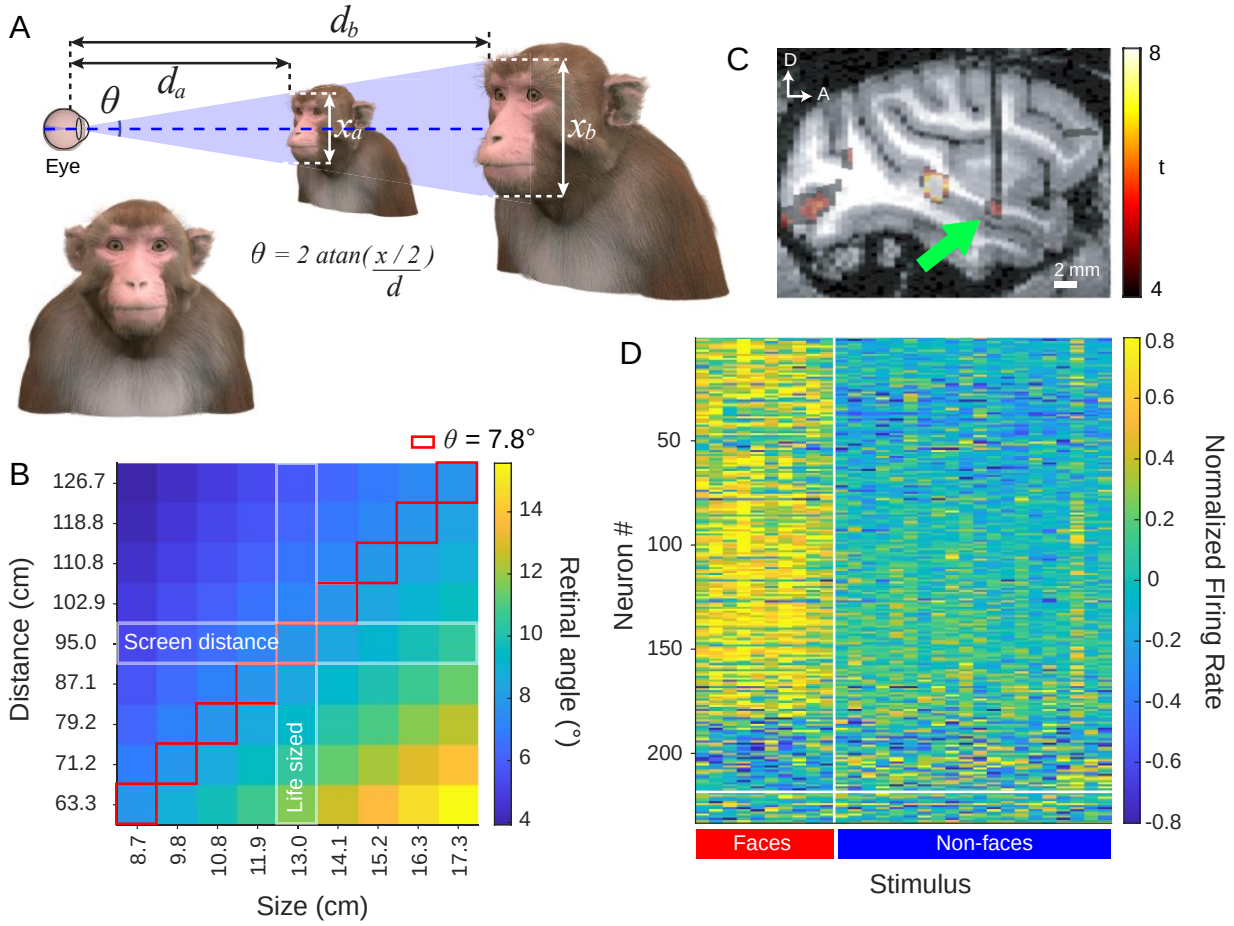

Figure 4.2 Macaque 3D avatar stimulus scaling and retinal angle matching. A. an example of the macaque avatar rendered in 2D and a schematic representing retinal subtense matching in 3D space where d represents the virtual distance between the subject's retina and the stimulus while x represents the size of the presented face and θ is the retinal angle as determined by the formula, which is equal between the two stimuli displayed here. **B.** Heatmap displaying the retinal angle of all size and distance combinations with the red outline indicating the stimuli with equal retinal angle and the white outlines indicate the screen distance and the average size. **C.** An fMRI overlay of the localization and targeting of the AF face patch in one subject, with an arrow indicated the position of the microwire bundle. **D**. Heatmap of the responses of all recorded neurons where each row represents a neuron and each column a stimulus. The heatmap is separated into face selective and non-face selective stimuli by the horizontal white bar and face and non-face stimuli by the vertical bar.

We recorded the activity of a total of 354 neurons from the AF face patch in two adult rhesus macaques (129 from male Monkey SR and 225 from female Monkey Sp, Figure 4.2C). Of these 354 neurons, 114 did not meet our visual response criteria of a response greater than 3spks/s above baseline to at least one of the rendered avatar images and were thus excluded from further analysis. For each neuron, we also calculated the face selectivity index (FSI, see Methods) based

on responses to a different image set, classifying cells as face selective if they had an absolute FSI greater than 0.333 (i.e more than double response to faces compared to nonfaces). Of the remaining population, 90.8% (218/240) neurons were face selective (Figure 4.2D). Of the 240 visually driven neurons, 87 were tested in the first experiment of physical size tuning **(**Figures 4.3-4.5), 69 were tested in the second experiment of preferred size (Figure 4.6-4.7) and 84 were tested in the additional experiment using non-face objects (Figure 4.1,4.8).

To investigate whether the size modulation of AF neural responses is determined more directly by the angular or physical size of the face, we utilized stereoscopic cues to present nine size renderings of the avatar face at nine different distances (Figure 4.1A, B). We placed particular focus on the responses to a subset of size/distance combinations that yielded images with equal retinal angle (Figure 4.1B, red outline).

4.3.1 Neurons sensitive to physical size of the face

In the first experiment, we tested the responses of 87 neurons, 89.7% (78/87) of which were face selective, to a stimulus set of 81 images at 9 sizes and 9 different distances resulting in 9 stimuli with equal retinal angle but distinct size and distance. Surprisingly, neurons were strongly affected by the physical geometry of the stimuli, even when its retinal angle remained unchanged. This effect is illustrated by the responses of an example neuron to a stimulus of fixed identity in Figure 4.3. When the retinal angle was fixed at 7.8° but the physical size and distance of the stimulus were stereoscopically varied, the visual responses were strongly determined by the physical size of the face and head (Figure 4.3A). On the other hand, holding the physical size constant and varying the distance/retinal angle (i.e. left or right column in Figure 4.3B) caused a much smaller size-modulation in the same neuron even as distance and retinal angle varied by a factor of two. Figure 4.3C shows responses to the subset of stimuli sharing a 7.8° retinal angle,

comparing the observed responses to the predictions of the retinal angle model (which would predict no change) and the physical size model (which would predict monotonic change). The

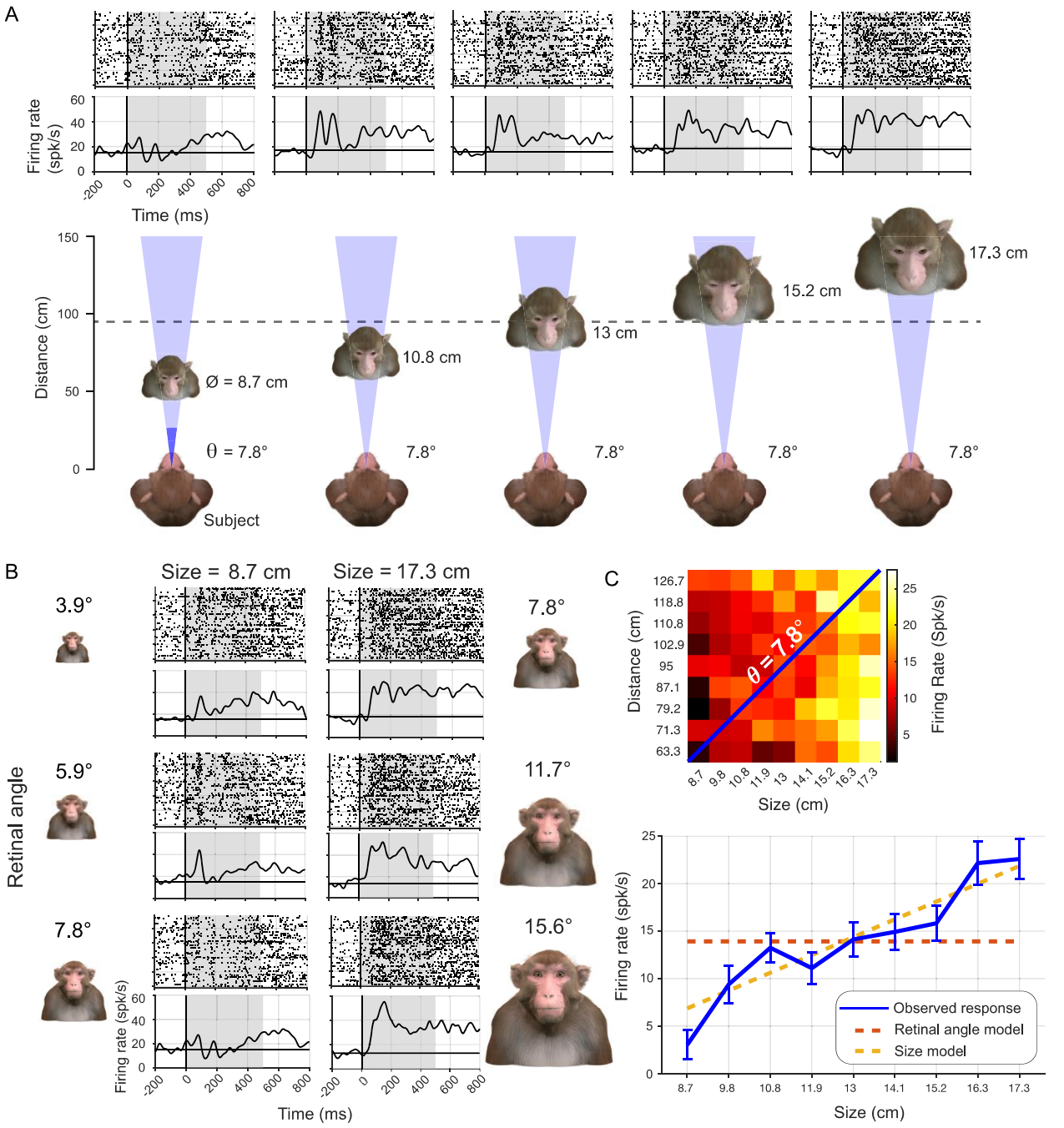

Figure 4.3 Example response of a physical size tuned neuron. A. Responses of an example neuron to five stimuli with the same retinal angle but different physical size-distance combinations displayed in the schematics below. Responses are shown in in spike rasters (top row) and spike density functions (SDF, bottom row). B. Responses of the same example neuron to stimuli with different retinal angle while each column represents a constant physical size of either the smallest (8.7cm) or largest present (17.3cm), demonstrating that physical size played a stronger role than retinal angle in the size tuning of this neuron. C. Average response of this example neuron across all the different physical sizes of equal retinal angle stimuli (blue) compared to the physical size model (yellow) and retinal angle model (orange). The inset heatmap shows the response profile for all size distance combinations.

responses of this example neuron clearly match the expectations of the physical size model.

Across the population, most neurons (81/87) were significantly modulated by physical size, with a partially overlapping subset (60/87) significantly modulated by physical distance (2-Way ANOVA, $P<0.01$). Importantly, 63.3% (56/87) of neurons were sensitive to the specific physical size/distance combinations that rendered the same retinal angle (1-Way ANOVA, $P<0.01$). These findings indicate that the responses of AF neurons are strongly shaped by the real-world geometrical parameters of the face, even when the corresponding visual images have the same retinal angle.

To quantify and compare the relative contribution of physical size versus retinal angle across the AF neural population, we performed a linear regression analysis to determine the factor having the strongest influence on the responses of each neuron. For the response matrix of all size/distance combinations, we compared linear regression models using retinal angle or physical size parameters alone, to a multiple regression model including both factors (for details, see **Methods**). This method provided a means to assess the preference for the retinal angle versus physical size model in capturing the variance of the data under conditions in which the two parameters have considerable covariation (Town, Brimijoin et al. 2017). For this analysis, the model preference could range from a value of -100%, indicating complete dominance retinal angle parameter on neural responses, to +100% indicating complete dominance of the physical size

parameter. When this analysis was applied to the example neuron in Figure 2, the model preference was +62.1%.

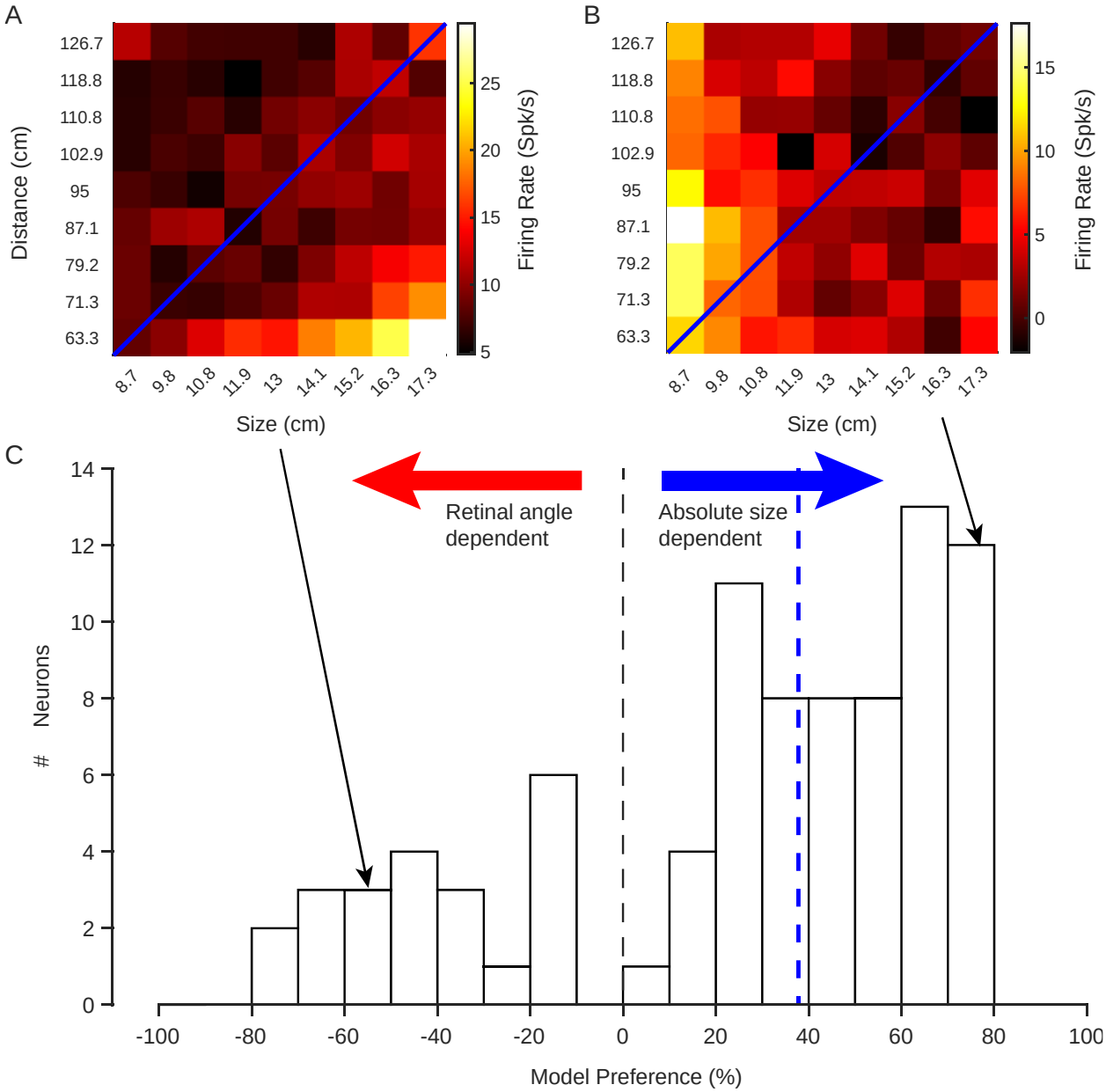

Figure 4.4 Model preference for physical size or retinal angle across the neural population. A, B. Heatmaps of the response of two example neurons that prefer retinal angle (A, model preference =-56.1%) or physical size (B, model preference =+73.4%). The blue diagonal indicates all the stimuli with equal retinal angle C. Histogram mapping the model preference of the whole neural population (see main text and Methods), with arrows indicating the dependence on absolute size or retinal angle and the dashed line displaying the median of the population. Across the population, most neurons responded were more sensitive to absolute size over retinal angle.

Across the population, the majority of neurons registered positive values, indicating that their responses were more strongly determined by physical size than retinal angle, mirroring the ANOVA results above (Figure 4.4). Specifically, 74.7% (65/87) of neurons were driven more strongly by the physical size of the faces, whereas only 25.3% (22/87) of neurons were driven more strongly by angular size of the faces. The median index of 37.8% ($p<0.0001$, student's T-test) for the neural population indicated that, on average, the physical size accounted for more than double the variance of the neural response as compared to the retinal angle model. Though we also manipulated distance, a model preference comparison of physical size and distance demonstrated 74.7% (65/87) of neurons were driven more strongly by physical size than distance, the exact same ratio as the comparison of physical size and retinal angle. These results indicate that the physical size of a face is a stronger determinant for the responses in the AF face patch than the retinal subtense of the corresponding image, indicating these neural responses are rooted in the 3D physical geometry of faces.

4.3.2 Neurons respond most to extreme physical sizes

We next examined the size tuning profiles of these neurons in greater detail to determine if there was a bias to prefer particular sizes across the population. For this, we restricted our analysis to the size/distance combinations having the same retinal size. This analysis unexpectedly revealed that neurons were rarely tuned to the real-world size of the macaque face and head, but instead responded most strongly to the smallest or largest sizes tested (Figure 4.5A). Of the neurons, 58/87 (66.7%) displayed their greatest response to the largest or the smallest sizes while only 5 (5.8%) neurons responded most strongly to the average or real-world size of a face of 13

cm. These results indicate that physical size tuning is a prominent feature if AF neurons, but this tuning is to the relatively large and small sizes of faces rather than to its most common size.

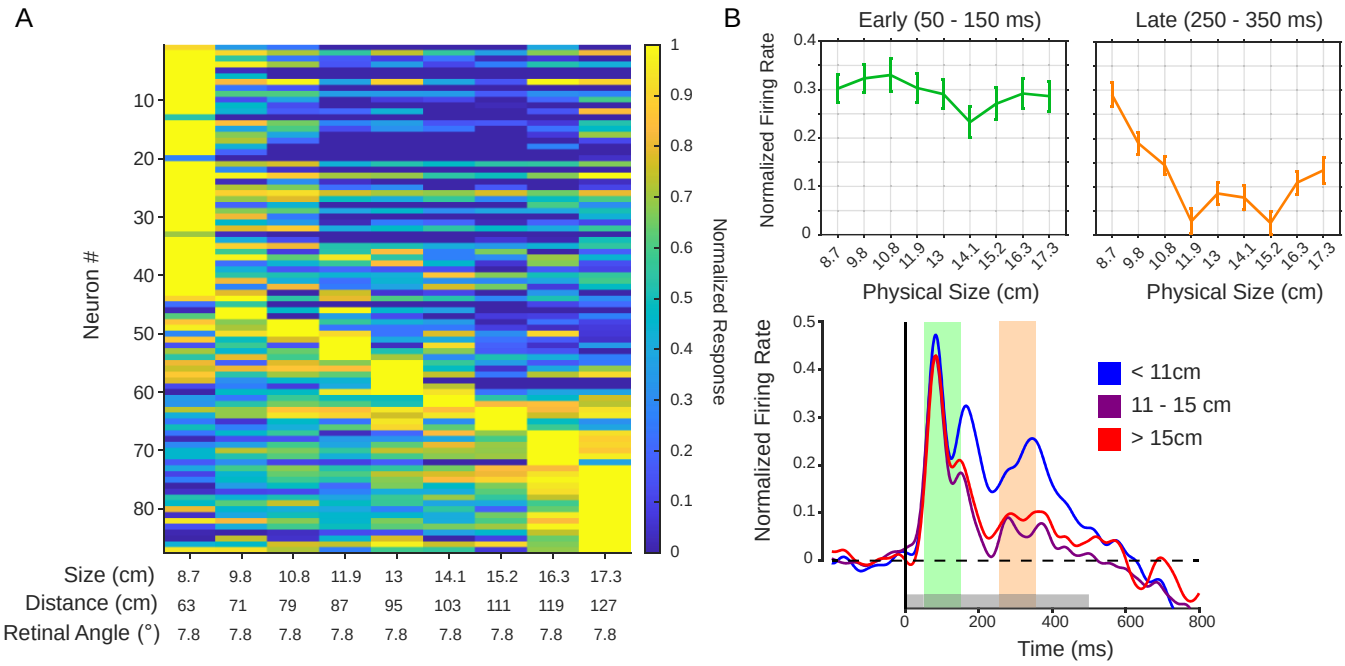

Figure 4.5. Size preference of neural responses. A. Heatmap of the response ration to stimuli of equal retinal angle (7.8 deg) where the strongest response for each neuron is 1 . **B.** the average time course with the smallest sizes in blue, the middle sizes in purple, and the largest in red with average tuning curves for each designated time window. All responses were normalized to the peak response for each neuron. Beyond showing sensitivity to physical size, the results revealed that most neurons preferred the largest or smallest physical sizes of the avatar face particular in the later phase of viewing.

The tuning to extreme sizes emerged gradually following the presentation of the stimulus and was most obvious during the late response period. This trend is visible in the mean neural responses time courses (colored traces in Figure 4.5B), which demonstrate that the stronger response to the largest and smallest physical face sizes emerge only beginning 150 ms following stimulus onset while the earlier response appears largely flat along the same retinal angle. The physical size tuning curves during two epochs (upper panels in Figure 4.5B) demonstrate that only approximately 250 ms following stimulus onset, had the tuning to extreme sizes fully developed.

To investigate whether this preference for extremes extended beyond the sizes we tested initially, we conducted an additional set of experiments in both monkeys in which we substantially

increased the range of physical head sizes. For this experiment, the stimuli were rendered at a 90cm viewing distance, corresponding to the physical position of the display. The subjects viewed

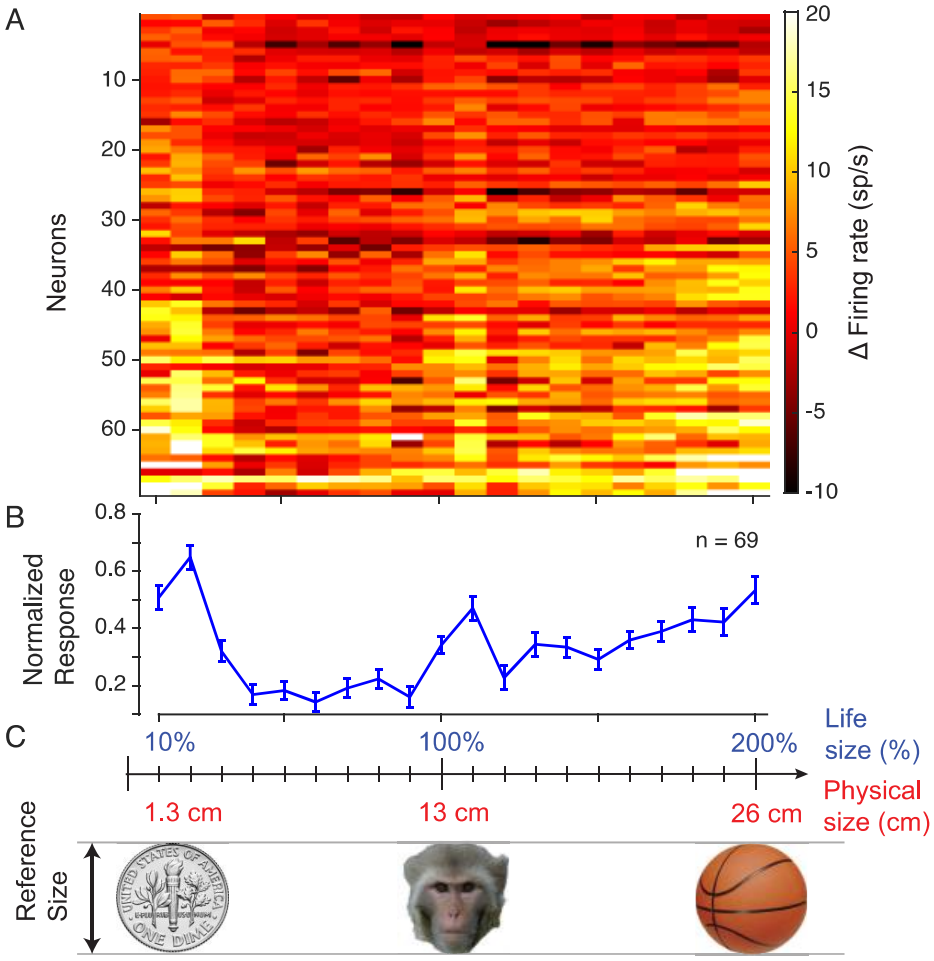

Figure 4.6. Neural response to extreme macaque face sizes. **A.** Heatmap of the raw firing rates of the neural population across all 20 sizes shown. B. Mean normalized firing rate of the neural population across all 20 sizes showing the strongest responses for the extreme sizes. C. Size of 3D avatar face stimuli relative to real-world objects.

avatar face at 20 different sizes ranging from 1.3 cm (one tenth the size of an average macaque face) to 26 cm (twice the size of an average macaque face) at intervals of 1.3cm. We recorded from an additional set of 69 neurons to test responses to these stimuli.

Despite the marked size deviation from a normal macaque face and head, neurons across the population continued to respond most vigorously to the extreme sized stimuli (Figure 4.6A,C). Most individual neurons responded strongest to either the large or small extreme sizes, with some

neurons responded to both extremes and less to intermediate sizes. Across the population, 73.9% (51/69) of neurons exhibited their strongest responses to either very small (<= 2.6 cm) or very large (>=24.7 cm) face stimuli or both. Neurons with higher firing rate were more likely to be tuned to both extremes as 8 of 10 of the highest firing neurons expressed this large and small extreme tuning. These features were also evident in the average population response (Figure 4.6B). In addition to the preference for extreme sizes, a subpopulation of cells also showed an elevated

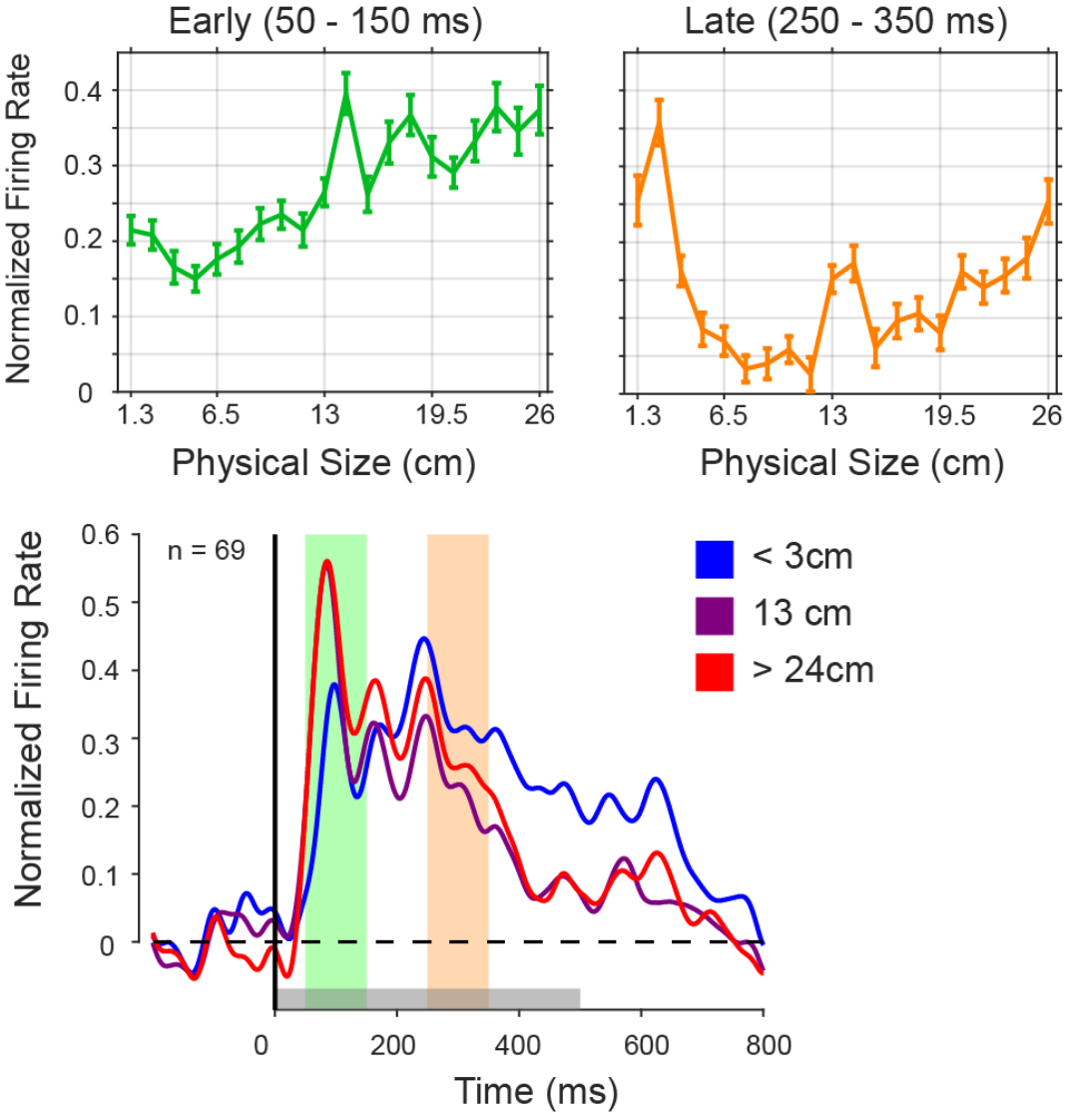

Figure 4.7..Time course for extreme size tuning. the average time course of neurons to the average of the largest 2 sizes in red, the smallest two in blue, and the middle size in purple with average tuning curves across all neurons for different sliding time windows for the time indicated by the color of the shaded windows

responses to the 14.3cm size, possibly reflecting the real-world size of the macaque face. 27.5% (19/69) of neurons showed their maximum response when the largest two and smallest two extreme sizes were excluded. Similar to the tuning within the previous stimulus set, time courses indicated that the tuning to both extremes emerged after the initial presentation (Figure 4.7 bottom panel). The tuning to both extremes also emerged approximately 250ms after the beginning of stimulus presentation, similar to the timing in the previous stimulus set (upper panels in Figure 4.7).

Finally, we asked whether the observed tuning to extreme face sizes would generalize to other rendered 3D objects. To this end, we presented 12 objects of different types at nine physical sizes ranging from 1.3cm to 26 cm all at the virtual distance of the screen (see Methods, Fig S3). These objects included other animals, familiar and unfamiliar fruit, and man-made objects. While the objects were diverse in their real-world sizes (e.g. a fork and a goat), the sizes of all objects in the current experiment were specified with respect to the physical geometry in centimeters, as before. We recorded responses to these stimuli from an additional set of 84 neurons, 77 of which were face selective. As these were face selective neurons, restricted our analysis to the non-monkey-face stimulus that elicited the greatest response. In doing so, we found that the size tuning in some cases aligned with that observed for monkey faces (Figure 4.8A) but in other cases did not, and sometimes favored the opposite extreme (Figure 4.8B). Similar to the tuning for monkey faces, the extreme size objects elicited the strongest responses for many neurons (53.6%, 45/84; Fig 4.8C). Interestingly, the majority of these cells responded to the largest rendition of the favored object, whereas the same cells were more evenly divided between large and small extreme sizes of the macaque faces (Fig 4.8D). Clearly, further investigation is needed to understand the

complex relationship between category selectivity, object-selective IT responses, and the encoding of physical object size.

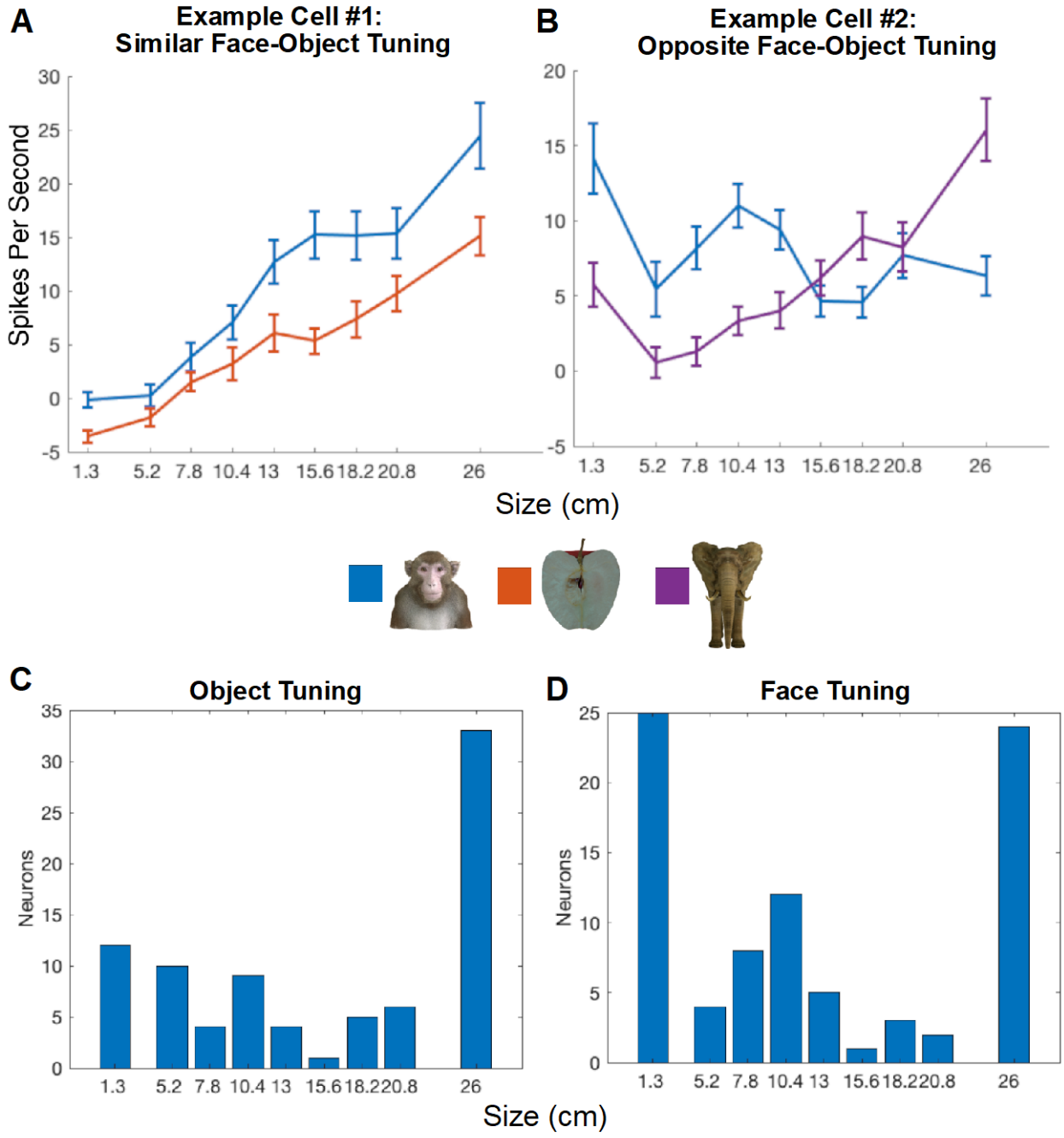

Figure 4.8 Neural responses to 3D object stimuli. A. Tuning plot of a neuron, which shows similar extreme size tuning between face and the most stimulating object, an apple. B Heatmap of a neuron with extreme size tuning the most stimulating object (an elephant) uncorrelated with the size tuning of the face. C, D the size that yields the maximum response for same population of neurons neuron for the most stimulating object and the macaque face stimulus, respectively

4.4 Discussion

The present results demonstrate that object-selective neurons in the inferior temporal cortex are sensitive not only to the identity of objects, but also to their geometrical parameters. Specifically, the response of most AF face patch neurons was strongly influenced by the physical size of a three-dimensional face. Unexpectedly, AF neurons responded to life-sized face stimuli with only moderate responses, reserving their strongest responses to extreme sizes, a feature that emerged in the later phase of stimulus viewing.

4.4.1 Metric information about objects in natural vision

The brain's registration of the physical size and other such 3D features of objects is fundamental to the ability to interact effectively with objects and the environment (Dobbins, Jeo et al. 1998, Murray, Boyaci et al. 2006, Fang, Boyaci et al. 2008, Kravitz, Saleem et al. 2013, Zeng, Fink et al. 2020). However, few studies have attempted to examine how these real-world parameters might affect neural responses in the ventral visual pathway. Traditionally, studies investigating the effect of size tuning among object-selective neurons, including those selective for faces, have reported size modulation relative to the retinal angle of a stimulus (Ito, Tamura et al. 1995, Op De Beeck and Vogels 2000, Freiwald and Tsao 2010). These studies also largely presented stimuli for 200-300 milliseconds. This presentation may also create a bias for tuning to retinal angle as the size tuning in AF largely emerged after the initial presentation beyond even 250ms. Retinal subtense, while undeniably important in early visual areas, is a parameter of the image projection cast through the optics of the eye, which is a joint product of an object's physical size and distance. Various size illusions have demonstrated that the perception of size can be dissociated from the retinal angle of a stimulus in 2D stimuli. Psychophysical examinations of

these size illusions have concluded that macaques like humans show susceptibility to some geometric size illusions (Agrillo, Parrish et al. 2014, Parrish, Brosnan et al. 2015). These size illusions highlight that under even monocular distance cues, retinal angle does not provide a clear reflection of perception.

While for some types of object recognition physical size may be unimportant, many problems of high-level vision are closely linked to the real-world geometry of objects, scenes, and individuals. For example, in establishing the structure of the scene, including its contents and behavioral affordances, the physical parameters of objects and their relative locations are extremely important. In the domain of manual behavior, such metric representation is needed to appropriately guide reaching and grasping movements. This may be particularly important in primates, whose arboreal evolution placed high demands on recognizing the geometric layout of their surroundings in order to interact with their environment. Theoretically, both animate and inanimate objects need to be represented internally with respect to their real-world size, volumetric shape, and position as this information informs perception and enables appropriate physical actions. For example, primates draw upon high-level 3D shape analysis for pre-shaping the hand when manipulating objects, moving through branches, or interacting with one another (Mason, Theverapperuma et al. 2004, Castiello 2005, Leopold, Mitchell et al. 2020). Previous recordings in the STS have suggested a direct interaction of visual object areas with parietal regions supporting visually guided manual behavior (Van Dromme, Premereur et al. 2016, Janssen, Verhoef et al. 2018). Such interaction may be mediated directly or indirectly through connections between the STS and visuomotor parietal cortex (Seltzer and Pandya 1984, Seltzer and Pandya 1994, Webster, Bachevalier et al. 1994) and is further supported by fMRI experiments demonstrating modulation of visual activity by stereoscopic 3D structure, including STS face

patches (Verhoef, Bohon et al. 2015). Furthermore, these regions of the STS largely do not specifically respond to vergence movements as compared to other eyes movements further indicating that these areas encode 3D structure rather than vergence movements (Ward, Bolding et al. 2015).

Even regions as early as V1 exhibit shifts in activity for perceived size and distance as opposed to retinal angle (Dobbins, Jeo et al. 1998, Murray, Boyaci et al. 2006). These responses in V1 are hypothesized to reflect feedback from higher order regions of the cortex of human visual cortex (Fang, Boyaci et al. 2008, Kravitz, Saleem et al. 2013, Zeng, Fink et al. 2020), which be homologous or analogous to the regions examined here. The information in V1 may also influence downstream regions in the STS and elsewhere that are known to respond selectively to 3D stimuli (Janssen, Vogels et al. 1999, Janssen, Vogels et al. 2000), forming a positive feedback loop. Previously observed tuning to 3D stimuli have also employed monocular depth cues such as texture and structure of motion and converges with stereoscopic cues within IT cortex (Liu, Vogels et al. 2004, Mysore, Vogels et al. 2010), consistent with the idea that the brain's analysis of physical geometry in the temporal cortex and elsewhere draws upon multiple, reinforcing 3D signals. Our findings are thus consistent with a wide range of results that suggest the primate STS region receives and responds to information about physical size and 3D shape to facilitate social or manual interactions. Further studies requiring manual behaviors or direct explicit tasks may more clearly illuminate the role of physical size in the responses of these regions.

4.4.2 Role of physical geometry in IT object responses

A few previous studies examining neural responses in the inferior temporal cortex examined the influence real-world geometric properties on IT neurons (Rolls and Baylis 1986, Ashbridge, Perrett et al. 2000, Vaziri, Carlson et al. 2014). One study in particular (Rolls and

Baylis 1986), found that small subset of neurons (a total of four cells) encoded images more directly in the geometry of the world (i.e. in centimeters) than the geometry of the retina (i.e. in degrees). This is likely the same phenomenon observed in the present study. Now equipped with a 3D scalable avatar stimulus, we showed that this sensitivity for physical geometry is abundant in a face-selective region in the STS fundus. The results suggest that the metric sizes of faces and other objects may figure more prominently into the visual system's analysis of a scene than previously believed. For example, since real heads and bodies have a relatively narrow distribution of possible physical sizes, prior experience of natural statistics can help decode the layout of a scene populated by individuals at different physical distances, even under conditions of sensory uncertainty. For social stimuli such as faces and bodies, this geometrical dimension of encoding would likely be superimposed on other dimensions, such as the analysis of identity or actions. This combination would likely then also aid in responding to social conditions and social scenes appropriately. The use of physical size may expand the understanding of IT cortex as a region not only for visual recognition but also for visual interaction (Pitcher and Ungerleider 2021).

A striking result in our study was the maximal responses of many neurons to extreme physical sizes, even for faces that were two times bigger or ten times smaller than a typical face. The sampling of local neural populations within AF indicated an intermixing of neurons responding most strongly to small and large sizes, as well as many neurons responding to both extremes. Moreover, the timing of this extreme tuning appeared largely after the initial 150ms of the stimulus presentation. This consistent finding suggests an encoding principle related to statistical deviations from expected size rather than a precise tuning to specific sizes, commonly known as prediction error (Bell, Summerfield et al. 2016). This finding has parallels with the extreme responses one observes in this and other face patches for faces with high feature

distinctiveness (Leopold, O'Toole et al. 2001, Leopold, Bondar et al. 2006, Chang and Tsao 2017, Koyano, Jones et al. 2021). These observations have led to a theory of norm-based coding where neurons encode identity based on the distinctiveness of face features relative to a prototype average face. While the facial features remained constant in the present study, the size of the animal depicted varied, which may be an important aspect of identity. Thus, whether contributing to the recognition of an individual or other social parameters, the observed tuning to extreme face sizes may be similar or analogous to the tuning to extreme identities, which potentially underpins the norm-based coding of face identity (Koyano, Jones et al. 2021). These encoding mechanisms may also explain why neurons responded to both small and large extreme stimuli, particularly in the expanded stimulus set. Importantly, the emergence of norm-based identity tuning was also late in the response period, similar to the results uncovered here. Another possibility is that the distinctiveness of these sizes may result in increased attention. In this framework, stimuli activate the attentional systems of the brain resulting in greater activity for the extreme sizes, which most capture attention. The later timing of distinct responses may then reflect feedback from frontal brain regions, which modulate the responses of temporal neurons in the later phases of viewing (Kastner and Ungerleider 2000). While the basis of this late modulation for faces is a matter of future studies, one possibility is that it relates the implementation of an expectation, prediction, or attentional signal emerging elsewhere in the brain.

Our results also draw attention to recent studies demonstrating an unexpected broad range of 3D scene statistics that appear to be encoded explicitly in the ventral visual cortex. Specifically, Connor and colleagues reported that some IT neurons apply an allocentric element to their responses, in some cases tuned to gravity-aligned representations of visual stimuli. Other neurons appear tuned to expansive 3D shapes or planar or immersive scene geometries (Vaziri, Carlson et

al. 2014, Vaziri and Connor 2016). These results underscore an emerging view that that object representations in the primate IT cortex may include information about its scene context, including its geometric parameters. At the same time, other recent studies have demonstrated that neurons the ventral visual pathway can sometimes be sensitive to highly specific details and localized details of objects (Ayzenberg and Behrmann 2022), including faces (Waidmann, Koyano et al. 2022) This has led to an emerging view those aspects of temporal cortex processing that involve global shape and scene geometry may derive through direct and indirect input interaction with the parietal cortex (Seltzer and Pandya 1994) (reviewed in (Ayzenberg and Behrmann 2022)). Such communications across dorsal and ventral visual areas may enable the exchange of 3D and global shape information, facilitating the real-time interaction with objects during natural behavior.

The historical focus on 2D image presentation and object recognition has shaped modern conceptions of object representation in the ventral visual pathways. However, assessments of visual responses using stimuli measured in retinal angle may not capture the full capabilities or tunings of neurons in these cortical regions. Previous research has indicated that real-world size may serve as an organizing principal in the visual cortex of humans (Konkle and Oliva 2012). Physical size, while not precisely the same, may also provide an organizational principle in the macaque cortex rooted in the statistics of the natural world rather than those of the retina. Beyond the AF face patch, future work will reveal whether the 3D physical parameters of visual input have a general influence across IT cortex and how physical size may impact the responses of numerous regions of the mammalian brain. Greater attention to metric information may unlock previously unknown functions of IT cortex pertaining to action. Focus on the ventral visual stream has largely been on its role in shape perception or feature decoding (Goodale and Milner 1992), but increasingly it appears that through its interaction with other brain areas, it may be involved in

other aspects of action-oriented vision that require metric information. These results demonstrate that a specialized region of the temporal cortex shows sensitivity to physical size over retinal subtense of faces and objects, marking a new approach to evaluate information processing in this region and clarifying an important aspect of its function. The next chapter combines investigations of these spatial elements with the audiovisual integration described in the previous chapter to examine the role of these effects combined.

5. Audiovisual Spatial Perception

5.1 Introduction

In Chapter 3, we established macaque AF face patch neurons showed audiovisual responses while AM face patch neurons did not. Then in chapter 4, we explored the tuning of AF face patch to 3D physical size suggesting a broader relationship between this region and more spatial visual sections of cortex. In this chapter, I combine elements of both of the previous chapters to examine the spatial properties of audiovisual responses. To conduct these experiments, I developed new experimental set-ups to manipulate spatial properties of both visual and auditory stimuli and further assess how spatial shifts can influence audiovisual responses.

For the auditory and visual systems, each sense processes space in dramatically different ways. As described in chapter 4, the visual system determines space through the geometry of the retina. The STS, including AF face patch, then receives spatial information from dorsal stream regions. In contrast, the auditory system calculates space and distance through binaural differences and intensity cues (McAlpine 2005). At the cortical level, fMRI of the auditory cortex demonstrates that the auditory cortex in macaques does not show a spatial topography of auditory space. Instead, auditory stimulus activating wide ranges of cortex regardless of position (Ortiz-Rios, Azevedo et al. 2017). However, at the single neuron level, neurons in posterior belt regions of the macaque cortex show tuning for precise locations of auditory space (Tian, Reser et al. 2001). This tuning aligns with the dual-stream hypothesis and this information eventually arrives in the parietal cortex to combine with visual information (Lewis and Van Essen 2000).

Unsurprisingly, the visual system often predominates in the perception of spatial features and connects with these posterior auditory regions (Werner-Reiss, Kelly et al. 2003, Kayser,

Petkov et al. 2007). These connections appear to contribute information about visual position and eye position that shapes the spatial tuning of auditory cortex (Bizley and King 2008, Bizley and King 2009). Auditory spatial information combines with visual information at other points in the sensory hierarchy. Neurons in more anterior belt regions also retain a broader spatial tuning, which may directly arrive in the STS given the connections between the regions (Seltzer and Pandya 1989). Similarly, the information in the posterior auditory cortex also arrives to VIP. Across the dorsal visual stream, VIP transforms coordinates from incoming auditory and visual information into a hybrid coordinate system that allows the resolution of space in multiple forms (Mullette-Gillman, Cohen et al. 2005). This region also connects directly with the STS to provide spatial information (Pandya 1984, Seltzer and Pandya 1994) as described in Chapters 2 and 4. However, the spatial properties of the STS have largely been ignored when studying audiovisual integration. The STS has primarily been studied for audiovisual responses in the context of timing and semantic connection as described in Chapter 2. Given the connections with these regions that contain both visual and auditory spatial information, the STS may also play a role in integrating multisensory information across space. One report of neural coherence increase to presentations of audiovisual looming in STS comports with these connections (Maier, Chandrasekaran et al. 2008). But this reports primarily used retinal size cues rather than binocular cues obscuring distance and size. Moreover, studies examining looming often only use synthesized expanding dot stimuli (Maier, Neuhoff et al. 2004, Maier, Chandrasekaran et al. 2008, Cappe, Thut et al. 2009, Cappe, Thelen et al. 2012). Importantly none of these studies has examined how size, distance, or movement in depth affect the response to naturalistic audiovisual stimuli in the STS. AF face patch, then, may combine these spatial properties with social information to assist in the analysis of complicated social scenes.

Here, we interrogate the effects of various spatial factors on single neuron responses in AF face patch in three experiments. In experiment #1, we examine the role of head orientation of the macaque avatar audiovisual stimulus. Given the sensitivity of face neurons in the STS to gaze direction, we hypothesized that AF face patch audiovisual responses would vary across uniformly across gaze azimuth. In experiment #2, we investigated the role of binocular distance and size of the visual stimulus. Given the results of chapter 4 and previous research, we hypothesized that the most extreme sizes, the closest distances, and looming would modulate the audiovisual responses the greatest across neurons. Finally, in experiment #3, we examined the role of relative spatial positioning of auditory and visual components to determine if audiovisual responses in AF face patch were coded in particular reference frame. We hypothesized, as the STS is largely visual, that neurons would code audiovisual responses in eye-centered coordinates. We demonstrate the first indications that spatial information can transform audiovisual responses in single neurons in macaque AF face patch. Each of these features affected neurons in distinct but unexpected ways suggesting that the STS and AF face patch like auditory cortex represents audiovisual space with a distributed code.

5.2 Materials and Methods

5.2.1 Subjects

As in chapter 4, two rhesus macaque monkeys, SP (Monkey 1, Female, 18yrs, 9 Kg) and SR (Monkey 2, Male, 6yrs, 10 kg), were implanted with a chronic microwire electrode bundle in AF face patch housed in a custom MRI-compatible 3D-printed chamber and microdrive. Electrodes were advance towards the target recording depth post-surgically to achieve stable recordings. Monkey SP was implanted in the right hemisphere while Monkey SR was implanted

in the left hemisphere. These subjects participated in all experiments performed in this chapter. All procedures were approved by the Animal Care and Use Committee of the National Institute of Mental Health and were conducted in accordance with the National Academy of Sciences Guide for the Care of Laboratory Animals and the NIH Animal Research Advisory Committee (ARAC) Guidelines. The NIH Animal Care and Use Program is accredited by AAALAC, International.

5.2.2 Experimental Setup and Design

In experiments #1 and #2, we utilized an OLED 3D screen paired with a pair of Tannoy speaker to present audiovisual stimuli the details of which are described further below. In experiment #2 all stimuli were presented in stereoscopic side-by-side 3D, wherein two images of

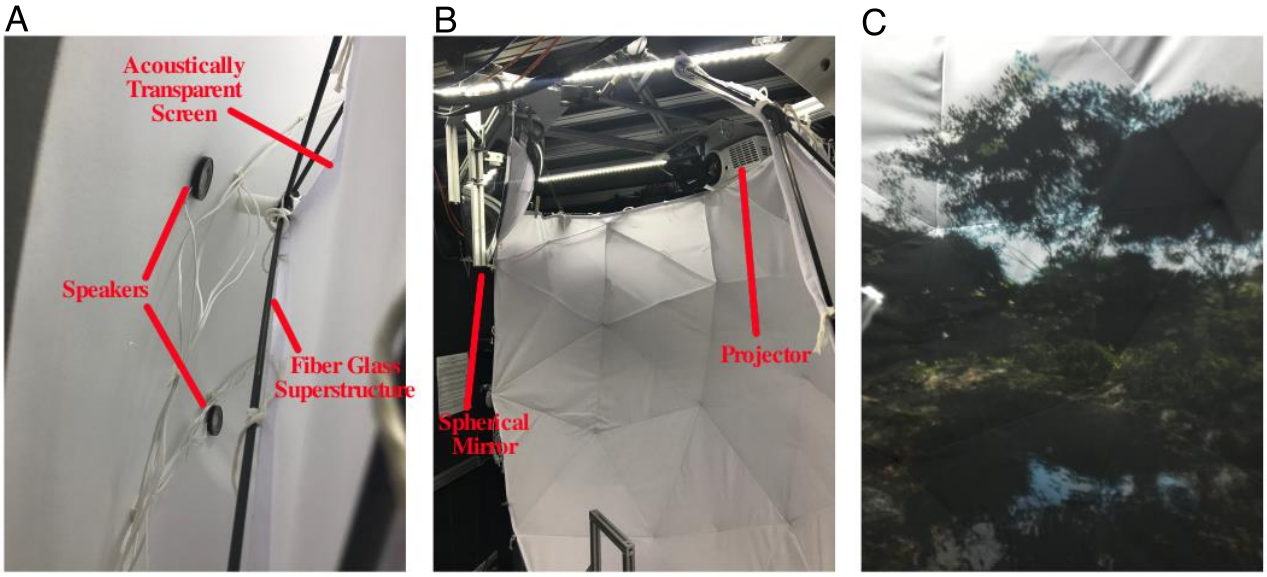

Fig 5.1 Arrangement of Virtual Reality Dome. A. A display of the speakers and superstructure beneath the acoustically transparent screen. **B** A display of the arrangement of the spherical mirror and the projection pattern. **C** Example of a virtual environment projected on to the acoustically transparent screen.

a stimulus were overlapped but given opposite polarization, as in Chapter 4. Briefly, the subject viewed the stimuli with a differently polarized lens positioned in front of each eye with a 3D printed lens holder such that only one image would be visible creating an illusory disparity between the two eyes.

In experiment #3, we employed a 3D virtual reality dome, iDome, with a set of speakers. The speakers are mounted on to the dome and covered with an acoustically transparent screen. The screen was connected to the dome with a super structure suspended above the dome surface and composed of fiber glass rods and 3D-printed carbon-reinforced connectors (Figure 5.1A). Visual stimuli were projected on to the screen using an Epson 1040 projector reflected on to a half-sphere mirror (Figure 5.1B). All visual stimuli had to undergo warping to present upon a curved surface. All visual stimuli were warped beforehand using tgawarp to produce a set of prewarped images or during presentation using the PsychImaging functions of Psych Toolbox-3 following calibration to the surface using the DisplayUndistortionBVL function from the Banks Vision Lab (Figure 5.1C). All acoustic stimuli were presented through each speaker individually amplified with a Behrenger amplifier and controlled with an audio interface. All stimuli had to be synchronized for presentation through vertical blank (VBL) flip interval synchronization to ensure temporal fidelity between the auditory and visual stimulus.

5.2.3 Electrophysiology

All macaque subjects were implanted with a 64-channel microwire bundle that enabled long-term recordings (McMahon, Jones et al. 2014) as in Chapters 3 and 4. Briefly, the subjects underwent a fMRI face localizer contrasting periods of visual presentation of faces with periods of visual presentation of non-face objects or scenes to locate the face patches. Subjects were then surgically implanted with a 3D printed custom microdrive and chamber targeted towards AF face patch. The electrode was advanced until stable recordings and positive yield was achieved. All macaque recordings occurred in a radio shielded room (ETS-Lindgreen) with a RZ2 BioAmp Signal processor (Tucker-Davis Technologies) with a 128-channel capacity collecting a broadband signal of 0.5Hz-20KHz.

5.2.4 Stimuli

In experiments #1 and #2, we also employed the 3D macaque avatar. Briefly, the avatar, described further in Chapters 3 and 4 as well as (Murphy, Faroni et al. 2019), is developed from CT meshes of real macaques. Using these meshes, we created a 3D model of a macaque, which we could manipulate and animate in the software Blender (Blender Foundation). As described in Chapter 3, the movement of the avatar could be then matched to the clips of real macaques vocalizing, thus enabling us to reproduce these vocalizations with more direct experimental control of its visual features. In addition to controlling properties such as identity and contrast, this macaque avatar enables direct control of spatial features including head angle, size, and distance, enabling us to pursue a myriad of examinations. In experiment #1, we changed the azimuth of the head of the avatar to examine how eye contact and direction of vocalization impacted audiovisual modulation. Previous research has indicated head azimuth greatly affects

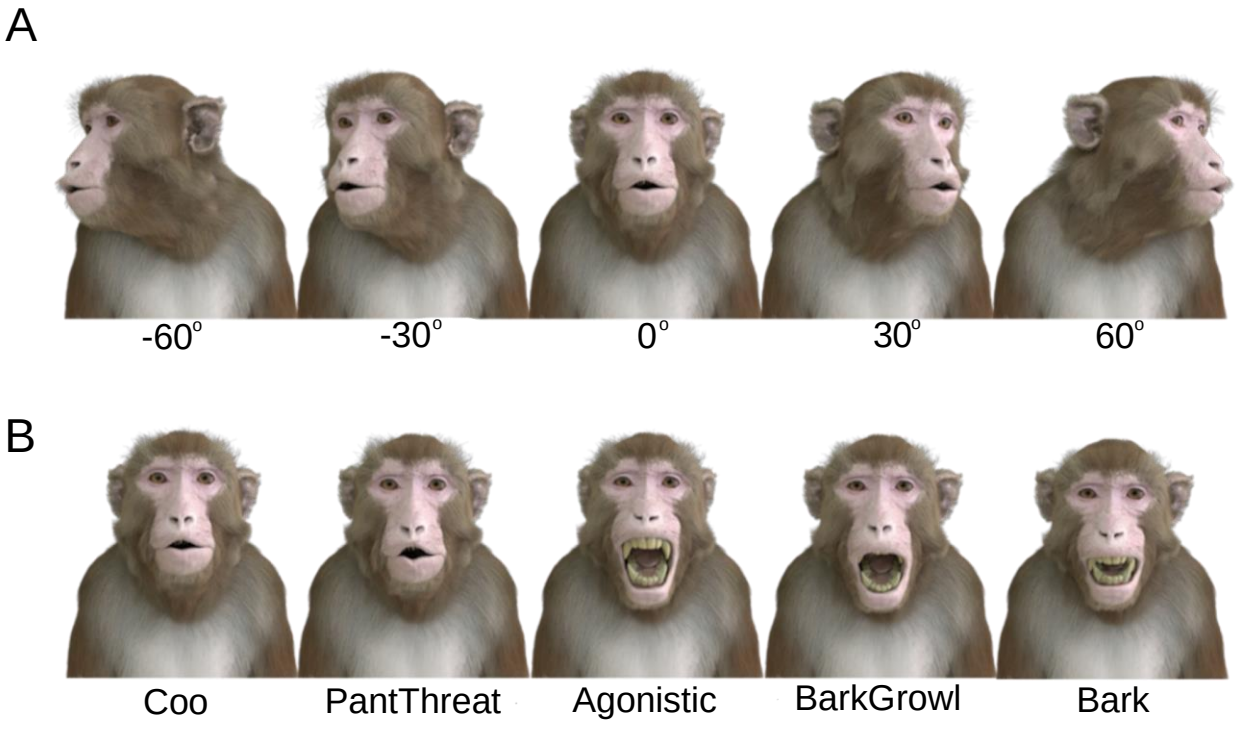

Figure 5.2 Avatar Head Orientation Call Stimuli. A. examples of the azimuths of calls produced by the avatar. **B.** examples of the different call-types used with the avatar for these experiments

responses in different face patches but its importance has never been examined in audiovisual integration (Freiwald and Tsao 2010). We presented five head azimuths ranging from 60 degrees to the right (-60°) to 60 degrees to the left (+60°) at intervals of 30 degrees (-60°, -30°, 0°, +30°, +60°) for five different vocalizations (Figure 5.2A,B). These stimuli enabled us to begin examining spatial elements of a socially important stimulus on audiovisual integration.

In experiment #2, we evaluated the response of neurons to changes of a stimulus in 3D space. All previous avatar stimuli were presented at the average size of a macaque face of 13cm and were flat within the plane of the screen at a distance of 95cm. While stimuli in visual and audiovisual experiments are often presented on screens without volume, these experiments have often neglected the role absolute size and distance in the everyday world. To further examine these spatial properties, we rendered and presented the stimuli in stereoscopic 3D and transformed the absolute size or real size of the avatar as well as the virtual distance to examine the effect of caller size and closeness on neural responses. We animated six calls at three different sizes (8.7 cm, 13 cm, 17.3cm) and three virtual distances (63.7cm, 95cm, 126.3cm) for a total of nine different spatial conditions. These stimuli included three coo calls, a bark growl, an agonistic call, and a threat call. These stimuli were matched such that the smaller stimulus at the closest distance, the middle stimulus at the middle distance, and the largest stimulus at the farthest distance occupied an identical retinal subtense of 7.8° as in Chapter 4. Additionally, we incorporated looming into our visual stimuli, which has previously elicited multisensory responses in the STS and compared it against receding visual stimuli (Maier, Chandrasekaran et al. 2008). The looming stimuli were rendered for every stimulus size and featured the audiovisual or visual stimulus moving from farthest distance to the nearest distance while receding stimuli featured the stimuli moving the opposite direction from the nearest distance to

the farther. All presented videos contained a 500ms still frame before all movies to disentangle transient visual response produced from flashes on a screen from more naturalistic conditions to examine audiovisual integration.

Finally, in experiment #3, we presented a set of three of the original audiovisual macaque calls of a single identity to minimize differences. Each stimulus was of a different call type to determine if differences in call type could cause changes in the audiovisual spatial effects. All stimuli were presented at various locations in the virtual reality dome. To trigger each stimulus, subject had to fixate at one of three positions at -30°, 0°, and +30° in horizontal azimuth (Figure 5.3A). Once fixation was achieved, the auditory stimulus could play directly on this position, to 30° azimuth shifted to the left or to the right, or elevation shifted 45° above or below for a total

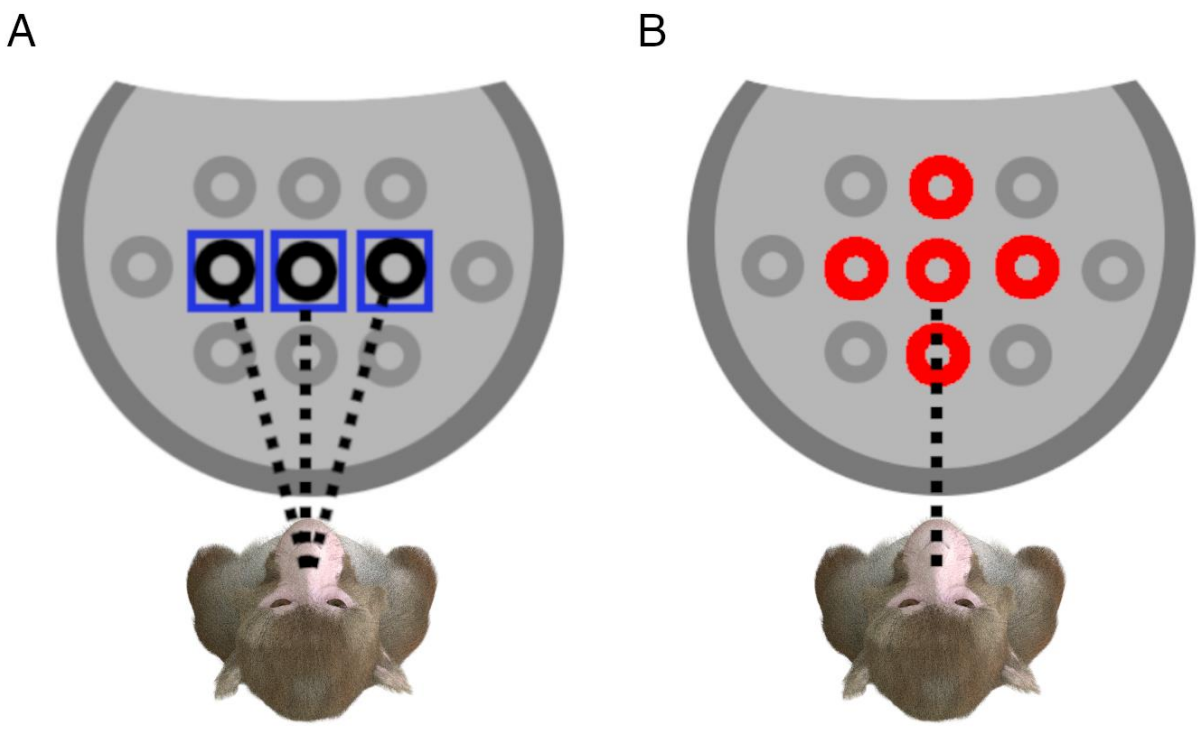

Figure 5.3 Schematic of Audiovisual Presentation of Dome Experiments. A. a schematic of the visual positions for stimuli. Each blue square represents a location for the visual stimulus. The black and grey circles represent speaker positions. **B.** a schematic of the positions of the auditory stimuli for the center visual positions. The positions of the auditory stimuli are represented by the red speakers.

of five auditory positions at each fixation position (Figure 5.3B). All movie stimuli were presented in audio only, visual only, and audiovisual formats.

5.2.5 Data Analysis

All recordings were sorted into spikes offline using the automated wave_clus spike sorting algorithm through the NIH high-performance computing cluster (Quiroga, Nadasdy et al. 2004). For face neurons, neurons were concatenated across days by matching waveform and selectivity based on responses to fingerprinting stimulus. For audiovisual stimuli, we calculated the responses for a window from 500ms (the end of the still frame and the beginning of motion) to100ms after the conclusion of the movie stimulus and a baseline window between -300 and 0ms before the onset of the still frame. For experiments #1 evaluating the role of head orientation, we tested significance by using a three-way ANOVA comparing the audiovisual, visual, head azimuth conditions and the audiovisual index described in similar to Chapter 3. We also used the audiovisual index described in chapter 3 to examine the magnitude of responses. In experiments #2 and #3, we also conducted regression analysis for each cell across a variety of factors. To prevent overfitting, we also conducted LASSO regularization to select variables. LASSO regularization imposes a penalty on increased number of factors and ensures only the factors that significantly impact the spike rate remain. With LASSO regression we could narrow down the spatial features that had the clearest and strongest effect on neural responses as well as determine the neurons affected by spatial audiovisual factors, which then underwent further analysis. In experiment #2, factors included physical size, distance, and movement as well as the interactions of these features with the addition of auditory stimuli. Size and distance were encoded as continuous variables normalized such that the average size and the middle distance were one. Looming and receding were encoded as discrete variables (either a 0 or 1) to separate

from conditions at a fixed distance. In experiment #3, features included eye position, sound location, and their various interactions with auditory and visual stimuli. In this regression, we encoded each sound azimuth, elevation, and eye position as an independent discrete variable. For example, the variable of azimuth did not range from -60 - +60^{o} but each as an independent condition in the factor matrix (e.g. sound azimuth +60^{o} is encoded as [0,0,0,0,1] whereas sound azimuth +30^{o} is encoded as [0,0,0,1,0]).

5.3 Results and Brief Discussion

5.3.1 Experiment #1 Results: Effects of Head Azimuth

The head orientation of a visual stimulus can dramatically affect the responses of single neurons in the face patches and STS. Combined with motion sensitivity in this region, these results suggest that these head orientation may also affect audiovisual responses in AF face patch. We presented audiovisual stimuli of macaque avatar making one of 5 different calls of a variety of affects at 5 different head orientation. Subjects completed 25-40 repetitions of all

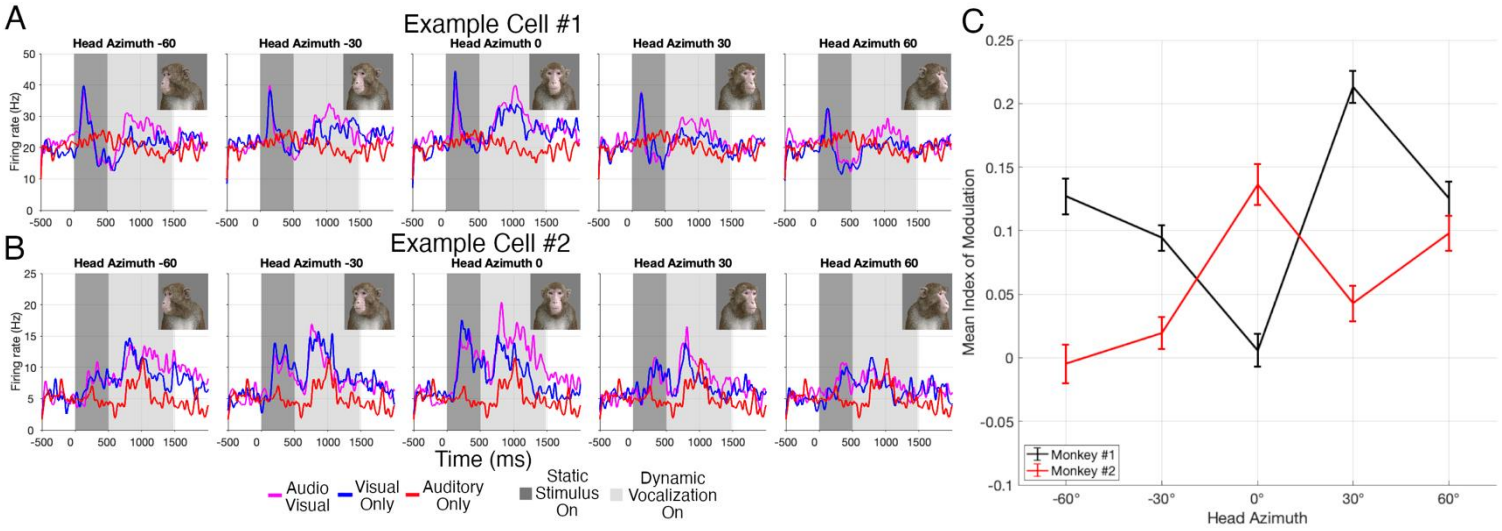

Figure 5.4 Effect of Head Orientation on Neural Responses. A. an example neuron with audiovisual interactions tuned for the more extreme head azimuth stimuli, taken from Monkey #1. **B.** an example neuron with audiovisual interactions tuned for the 0^{o} azimuth taken from Monkey #2. **C.** the mean index of modulation across all neurons for each monkey showing the overall preferred azimuth for each subject

stimuli. Shifting head azimuth of the stimulus similarly did affect the multisensory responses of AF neurons but inconsistently and without a specific trend among the neural population. Neurons demonstrated audiovisual responses to specific head angles of the avatar over others with particular importance of the 0° condition, which involved direct eye contact. Many single neurons demonstrate stronger audiovisual modulation to the more extreme head azimuths rather than direct eye contact with those extremes showing a larger difference between audiovisual and visual only conditions (Figure 5.4A). But others demonstrated stronger modulation for direct eye contact, with some only being modulated by the 0° condition while showing no change from the visual response for others (Figure 5.4B).

To summarize these responses, we conducted an ANOVA for each neuron with audiovisual compared to visual condition, and head azimuth as factors. Across the population, 118/133 (87.2%, $p<=0.05$) showed a main effect for head azimuth and 74/133 (55.6%, $p<=0.05$) a significant main effect of audiovisual. Unsurprisingly, 68/133 (51.38%) showed a significant main effect for both these factors. Interestingly, 25/133 (18.8%) showed an interaction effect for both of these features. However, we also examined the magnitude of responses again using the index of modulation for the different orientations and to look for groupings of these neurons with different tunings. Examined with the index, these azimuths produced differing effects across animals as the population of neurons in one animal experienced the most enhancement for the more extreme head azimuths, while the neurons in the other experienced the most enhancement for direct eye contact (Figure 5.4C). These effects appear largely consistent across both animals as the index at the direct eye contact position is significantly different from the other azimuths in both animals (t-test of means, $p<0.05$). Combined with the ANOVA results, this analysis suggest

that populations of neurons within AF face patch can show changes in audiovisual responses dependent on the orientation of the head being viewed.

5.3.2 Experiment #1 Discussion

These results indicate the head orientation of a caller can change the way sound can influence a visual response. This observation and the widespread influence of head azimuth on AF neurons align with previous assessments of face selective neurons in the STS, which finds the head azimuth and gaze direction can greatly shift the visual response of neurons (Jellema, Maassen et al. 2004, Freiwald and Tsao 2010). However, the macaques in our study showed opposite audiovisual tuning to the central position, in contrast with our initial hypothesis. At the neural level, few studies have examined audiovisual integration while systematically varying the direction of faces during vocalization. Recordings in VLPFC have previously found visual and auditory responsive neurons preferred faces oriented at 0^o or 30^o, indicating that these responses may encode direct social communication, although these neurons were not assessed for combined audiovisual calls (Romanski and Diehl 2011). As described in Chapter 3, the STS and the VLPFC show extensive interconnection suggesting that these connections may further process audiovisual signals in a feedforward or feedback manner (Seltzer and Pandya 1989). However, the relative variation between neurons and subjects that preferred audiovisual calls with particular head azimuth indicate that the STS serves may serve a role in segmenting broader social scenes rather than direct vocal communication.

AF face patch may combine observed head orientation with audiovisual signals to decode social interactions in larger group dynamics essential to the natural behavior of macaques. Previous work has argued that the sensitivity to head direction also reflects a sensitivity to the

attention and actions of a conspecific (Jellema, Baker et al. 2000). This encoding of the direction of action may extend to audiovisual calls as well. Similarly, parts of the STS also play a causative role in gaze following (Roy, Shepherd et al. 2014, Chong, Ramezanpour et al. 2023), which further reflects their role in broader, more complicated scenes. It remains unclear if these gaze-following regions but given the interconnections of the STS this information likely also reaches AF face patch (Seltzer and Pandya 1989). The audiovisual responses in AF face patch may combine with these gaze-following functions and social processing to decode the target of audiovisual calls between members of a colony further reflecting the ethology of macaque social structure. The variance between the two subjects could also reflect the effect of dominance in macaque behaviors. Eye contact is an aggressive posture in macaque social structure so more submissive animals could be more sensitive to direct eye contact for survival (Dewaal and Yoshihara 1983, Harrod, Coe et al. 2020). Importantly, the density gray matter of the STS can vary with both the social group size and dominance status (Noonan, Sallet et al. 2014). The effect of dominance may then shift neural responses in the STS depending on the social condition of the subject, which could explain variance between preferred head orientation. Again, the focal nature of recording may have affected the audiovisual modulation discovered and any conclusions regarding the effect of dominance on these responses remains unclear. Importantly, though each call had a different affect, we did not examine the interaction of call affect and head orientation. Aggressive or affiliative calls could have different affects and have different ethological relevance depending on the head azimuth. Overall, the effect of head orientation on audiovisual integrations requires further examination across the whole of AF face patch and with a variety of macaques ranging the social hierarchy and more attention to call varieties.

5.3.3 Experiment #2 Results: Effects of Size and Distance

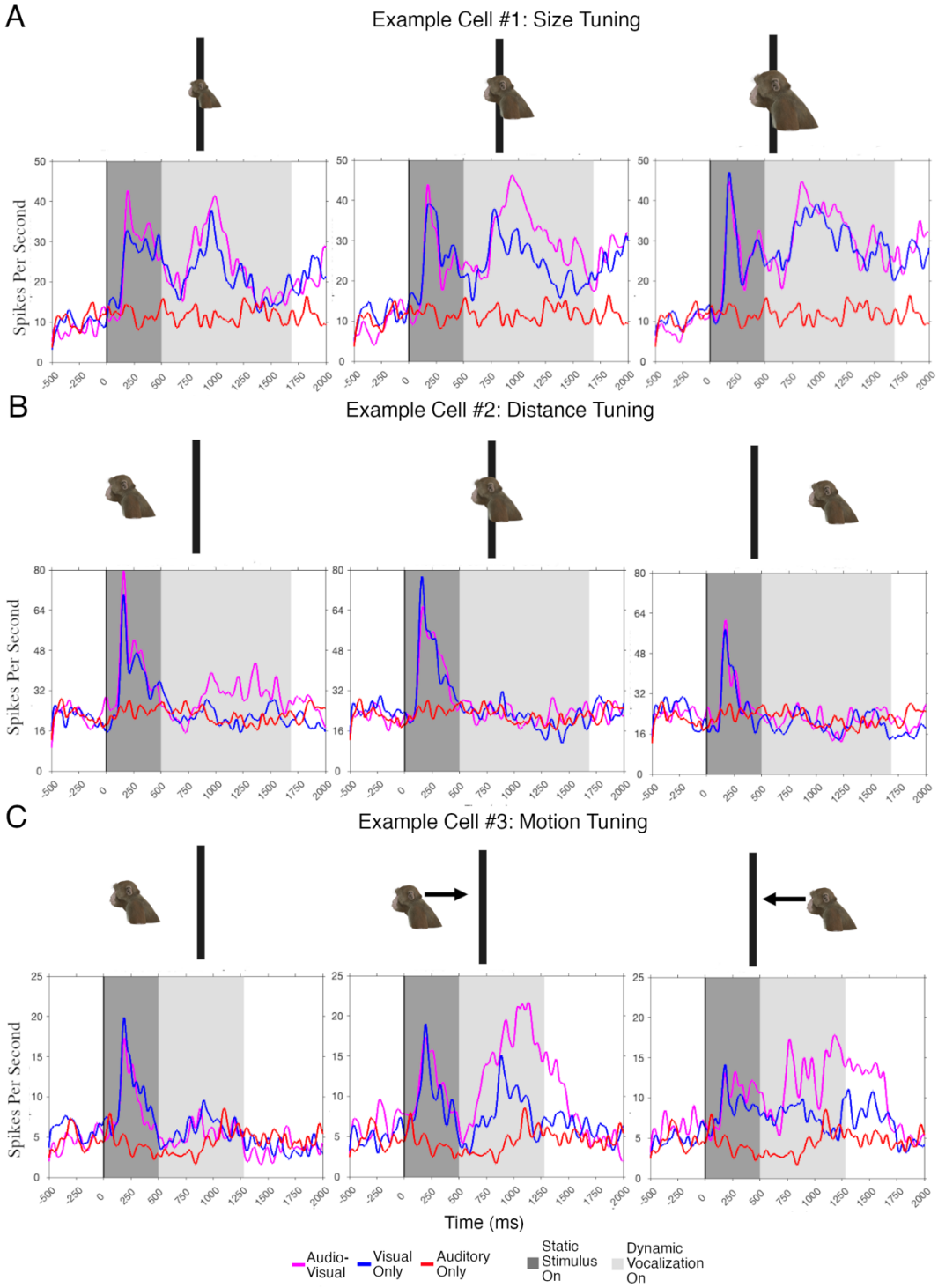

Figure 5.5 Effects of 3D Manipulations on Audiovisual Integration. Example neuron responses with schematics of the stimuli above individual SDFs where the black bar indicates the position of the screen and the arrows represent motion **A.** responses of an example neuron, which shows different responses to AV stimuli depending on the size with a preference for the middle size. **B**. responses of an example neuron, which shows an AV preference for the nearest distance. **C.** responses of an example neuron, which shows AV preference for looming and receding. All responses are chosen for the strongest single stimulus example

We then turned to examine the effect of size, distance, and motion in depth on audiovisual responses in AF. These factors have received very little attention in studies of audiovisual integration. We presented audiovisual call stimuli at a variety of sizes and distances, as well as looming and receding. These stimuli were only manipulated in the visual domain, while auditory stimuli were presented from the same location. Subjects viewed 20 repetitions of each stimulus. We discovered these factors similarly modulated the audiovisual responses of AF neurons as it did for the visual responses.

Some AF neurons demonstrated no audiovisual interaction at one size but exhibit audiovisual enhancement when the stimulus was another size across different call types (example in Figure 5.5A). These changes could represent a tuning dependent on the size of a conspecific relative to the self or as another method of distinguishing identity. Similarly, changes in distance could also shift audiovisual responses in single neurons with modulation only appearing at certain distances as compared to others (Figure 5.5B) in contrast to the relative lack of distance modulation in visual responses. These responses may indicate AF neurons might play a role in audiovisual communication with a conspecific as nearness and farness can denote whether a conspecific is interacting with the subject or not. Looming and receding could also elicit unique responses for both audiovisual and visual stimuli, indicating these AF neurons are modulated by 3D spatial movement (Figure 5.5C). This modulation in indicated binocular cues and size alone did not drive these responses. Importantly, the stimuli did not contain any auditory motion; audiovisual modulation was only present when the stimulus contained visual motion. Altogether each of these results indicate that 3D visual factors not only visual responses (Chapter 4) but also influence audiovisual responses.

To summarize these responses and gauge the effect across the neural population, we conducted regression analysis, using LASSO regularization, to examine the cells with significant effects for size, distance, movement in depth. and the combination of all these elements with audiovisual integration. We recorded 140 neurons, of which, 8 showed no significant effects to any stimulus, visual or auditory, and were eliminated from further analysis. Across the population, 80/132 (60.6%) of neurons exhibited some form of audiovisual integration or auditory response. Of these neurons, 63/80 (78.8%) displayed an audiovisual interaction with size, distance, motion in depth, or a combination of these features, showing 3D visual features

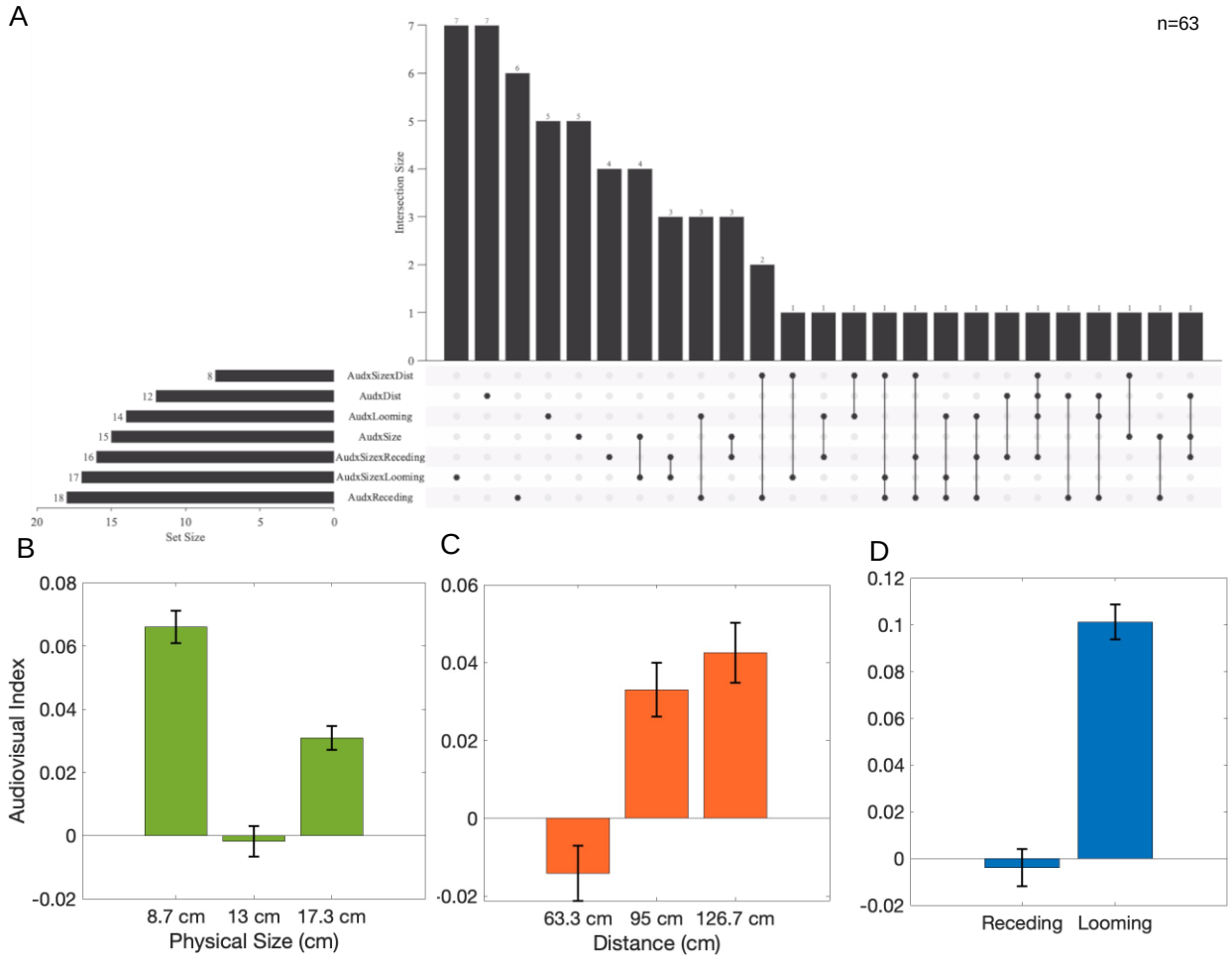

Figure 5.6 Effects of 3D Manipulations on Audiovisual Modulation. A UpSet plots of interactions of audiovisual 3d spatial factors on the neural population. These upset plots display the number of neurons that show an interaction for each labeled effects and the intersection of effects in columns. **B, C, D** the average audiovisual index for each neuron across all stimuli for each size (B), distance (C), or 3D motion condition (D)

can affect the audiovisual interactions of much of this neural population. The neural population did not show a particular tuning for any feature over another. To visualize these differences, we plotted the LASSO data in an upSet plot (Figure 5.6). This plot presents the effects the number of neurons with combinations of significant effects. Each column of the bar graph depicts the number of neurons with the combination of significant effects represented with the black dots below. Dots with connections drawn between the indicate the neurons indicated in the column show significant interactions for each of the effects in the rows with dots. Columns with more dots filled indicate more significant effects. In these plots, neurons with effects for size, distance, or motion also largely showed tuning for specific interactions between these features rather than multiple different interactions seen in the main effects. Interestingly, fewer neurons showed interactions between size and distance for audiovisual responses than for any other interactions. Audiovisual looming and receding also both showed effects on neurons both with and without interactions with size, in contrast with the common finding that looming predominates over receding (Figure 5.6A). AF neurons commonly also displayed auditory interactions between motion in distance, and size. These interactions may reflect a sensitivity to motion for particular sizes as a method of distinguishing identity or danger.

We then further examined the trends across the population for both size, distance, and motion in depth for these 63 neurons. We calculated the audiovisual index described in chapter 3 across all stimuli for size, distance, and 3D motion for each neuron. We then took the average across all audiovisual spatial neurons. Across the population, neurons appeared to show the weakest audiovisual amplification to the average size of the face similar to the weaker tuning to the average face size in the visual domain (Chapter 4). However, only the smallest size showed a significantly larger response than the middle size (Paired t-test of means, $p<0.01$). Interestingly,

the population also showed a preference for further distance, in contrast to our hypothesis, though the mean of the audiovisual index did not differ significantly between distances. Finally, though similar numbers of neurons showed interactions with auditory for both looming and receding, the audiovisual index was significantly larger for looming than for receding (paired t-test of means, $p<0.001$). These differences suggest that audiovisual receding cause both audiovisual enhancement and suppression, while audiovisual looming largely contribute to audiovisual enhancement. Overall, these results indicate that 3D visual features can also affect the audiovisual responses of AF face patch neurons.

5.3.4 Experiment #2 Discussion

We determined that a variety of 3D factors including size, distance, and motion in depth could modulate audiovisual responses in AF face patch. The effects of size reflect the changes produced by changing size in the visual domain demonstrated in chapter 4. Few other studies have examined the effect of changes in 3D space of audiovisual signals in single neurons in the macaque. One previous examination of the STS with audiovisual movies of real macaques found that the size of a caller could influence audiovisual responses (Perrodin, Kayser et al. 2014). Again, this study did not use 3D stimuli and determined the size by weight of the calling macaque caller but, using real differences in macaque size, provided evidence that changes in size could reflect changes in audiovisual responses in the STS. Similarly, our results also reflect that physical size of a calling animal can influence neurons in AF face patch but with a more systematic method of examining size. In line with our hypothesis, both the largest and smallest size exhibited greater audiovisual enhancement across the population with seemingly less modulation for the middle size, like the response to physical size in visual domain. However, only the smaller size significantly different from the average size. As with visual signals related

to physical size, this modulation by size may reflect a method of distinguishing identity examining deviations from normal size during calls. The affect of the call might also vary with physical size of the caller based on social relevance. But there remains a relative dearth of other studies examining the role of size in audiovisual processing. Future work will require a wider range of sizes to confirm the similarity with the visual domain and more stimuli of aggressive or agonistic affects to examine the role of social relevance.

The role of distance in audiovisual integration has received more attention primarily in through investigations of human behavior. These studies have primarily shown that stimuli with smaller retinal angle at greater distances (~2m and beyond) show the greatest audiovisual augmentation but only up to the limits of when the speed of light and sound would be significantly separated in time (van der Stoep, Serino et al. 2016, Van der Stoep, Van der Stigchel et al. 2016). Interestingly, in contrast with our hypothesis, neurons in AF face patch trended towards preferring further distances but did not show significantly different responses. However, our stimuli may not have explored enough of the range of distances required to see this change as all stimuli were presented from 62-128cm. Behaviorally, humans can distinguish spatial dislocations of auditory and visual stimuli across distance indicating that calculations of both depth in both domains is integrated into a single percept (Van der Stoep, Nijboer et al. 2014). In visual studies, early visual areas show tuning to the distance of images (Dobbins, Jeo et al. 1998), which may extend to facilitating audiovisual interactions and in auditory studies posterior, higher-order auditory regions including the superior temporal gyrus play a role in coding auditory space (Rauschecker and Tian 2000, Tian, Reser et al. 2001). However, it remains unclear whether these responses converge in the STS. One examination of the timing of visual and auditory signals to the STS using current source density (CSD) suggested that latency

of signals largely overlapped until the exceed the distance speed and sound are relatively synced in time (Schroeder and Foxe 2002) in line with reports of human behavior. However, within the range of distances (~1-20m), single neuron audiovisual responses may still vary to distinguish the urgency of communication, particularly in face sensitive AF. Overall, few studies have examined the effect of distance on audiovisual responses in single neurons and it requires further examination of its role across audiovisual regions.

Most other studies examining the effect of size or distance have primarily examined looming, which utilized dot stimuli or used changes in retinal angle rather than using binocular cues to examine changes in distance and physical size (Maier, Neuhoff et al. 2004, Cappe, Thut et al. 2009, Cappe, Thelen et al. 2012). Our results also indicated that motion in depth modulated audiovisual responses compared to calls at static distances but differed from previous reports in two substantial ways. First, these motion responses arose without any change to the auditory stimulus in contrast with other assessments of looming or receding in the macaque (Maier, Chandrasekaran et al. 2008). The continued audiovisual interactions during these solely visual motion signals indicates that the changes in spike rate seen with combined audiovisual looming may not stem from the semantic combination but from individual features of both stimuli. Second, the neural population did not prefer looming over receding while previous reports indicate audiovisual looming has greater behavioral significance than audiovisual receding (Cappe, Thelen et al. 2012, Conrad, Kleiner et al. 2013). However, the population did show greater audiovisual enhancement to looming, where receding likely caused both audiovisual enhancement and suppression. Another study of local field potential (LFP) during looming stimuli found looming but not receding increased coherence in of LFP signals and therefore signals arriving in this region (Maier, Chandrasekaran et al. 2008). However, LFP does not

necessarily correlate with single neuron spiking responses (Herreras 2016) and these results may not necessarily contradict as individual neurons may still show audiovisual modulation to both looming and receding. The increased coherence may also explain the enhancement across the population during looming as opposed to receding. Overall, motion in depth likely requires further examination for its role in this region AF face patch, as described in Chapter 3, has sensitivity to facial motion (Fisher and Freiwald 2015). This motion sensitivity may extend to looming and receding signals in audiovisual conditions as well. Interestingly, STS neural responses can depend heavily on the movement leading up to final face or body position including movements towards and away (Jellema and Perrett 2003). This history dependence may also influence these audiovisual responses in these motion conditions as movement in depth is a salient cue to understand the context of a call. Altogether, these 3D manipulations each impacted audiovisual responses in previously unreported ways indicating these factors require further study across AF and STS on the impact of these factors both in vision and audiovisual integration.

5.3.5 Experiment #3 Results: Spatial Separation and Eye Position

Finally, we examined the role of shifting a variety of spatial conditions including eye position of the subject and the location of auditory stimulus relative to the visual stimulus expecting that these neurons would encode audiovisual responses in eye-centered response frames. Through these manipulation in a 3D virtual reality dome, we attempted to disentangle the coding of audiovisual stimuli in different reference frames. In this experiment, we determined that various spatial factors including sound azimuth, eye position, and sound elevation could all effect single neuron responses in AF face patch indicating that neurons encode in a variety of

reference frames. Single neurons could show weaker auditory responses and audiovisual responses to particular parts of space than another (Figure 5.7A). This result indicates that certain neurons retain their tuning to head-centered space as in much of the auditory system. Similarly, eye position could shift audiovisual integration with the greatest audiovisual augmentation occurring at one position of the eye but not at another (Figure 5.7B). These responses reflect the

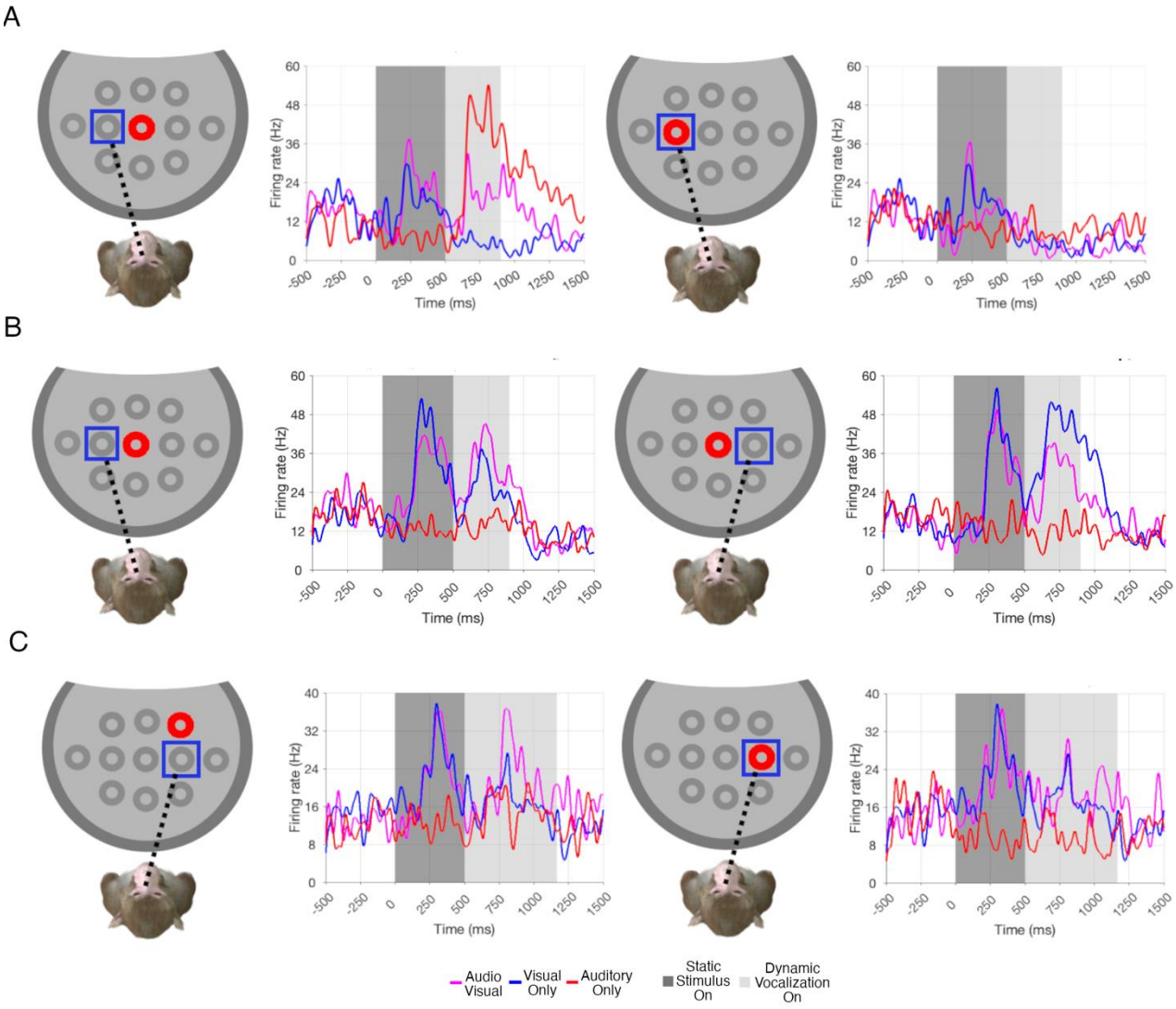

Figure 5.7 Effects of Audiovisual Spatial Variation on Single Neurons. Single neuron examples tuned to various changes in eye position, visual azimuth, and auditory stimulus position with schematics where the blue box represents the visual stimulus and eye position and the speaker schematic in red represents the auditory position. These schematics indicate the location of the stimuli that elicited the responses shown in the accompanying SDFs. **A.** a single neuron example where variation of auditory azimuth leads to substantial decrease in auditory and audiovisual response between the two SDFs. **B.** another single neuron example where a change in eye position shifts in audiovisual responses. **C.** a final single neuron example where a change in auditory elevation.

transformation of auditory signals into eye-centered space for some neurons in this region. However, other neurons presented with more complicated and less consistent patterns that vary according to multiple factors simultaneously (Figure 5.7C), reflecting more of the hybrid coordinate systems found in the IPS. These responses indicate that neurons contain a variety of tuning for audiovisual space in AF face patches.

We analyzed these complex factors again with regression analysis using LASSO regularization to identify the variables significantly affecting the spike rate. The LASSO further indicates that the recorded neural responses were influenced by a variety of spatial manipulations but that the population does not favor any of these factors over another. We recorded a total of 68 neurons, of which 11 showed no significant effects to any stimulus and were removed from further analysis. Of the remaining 57 neurons, 63.2% (36/57) continued to display an audiovisual effect, of which 61.1% (22/36) show an interaction between audiovisual interaction and spatial position (Figure 5.8A). We classified these spatial audiovisual neurons as having head-centered reference frame if they showed an interaction between visual presentation and one of the auditory spatial positions, eye-centered an interaction between auditory presentation and one of the eye positions, and hybrid an interaction with both. Though we initially hypothesized that neurons primarily encode sounds in an eye centric frame, we found, by this measure, 40.9% (9/22) of neurons showed a head-centric audiovisual frame, equal to the percentage of neurons showing eye-centric frames, and 18.2% (4/22) of neurons showed a hybrid reference frame (Figure 5.8B). However, we also examined the response to eye position in visual domain alone and found 52.6% (30/57) showed a significant effect for eye positions on visual responses alone. Interestingly, 77.8% (7/9) of the neurons with a head-centered frame in the audiovisual also

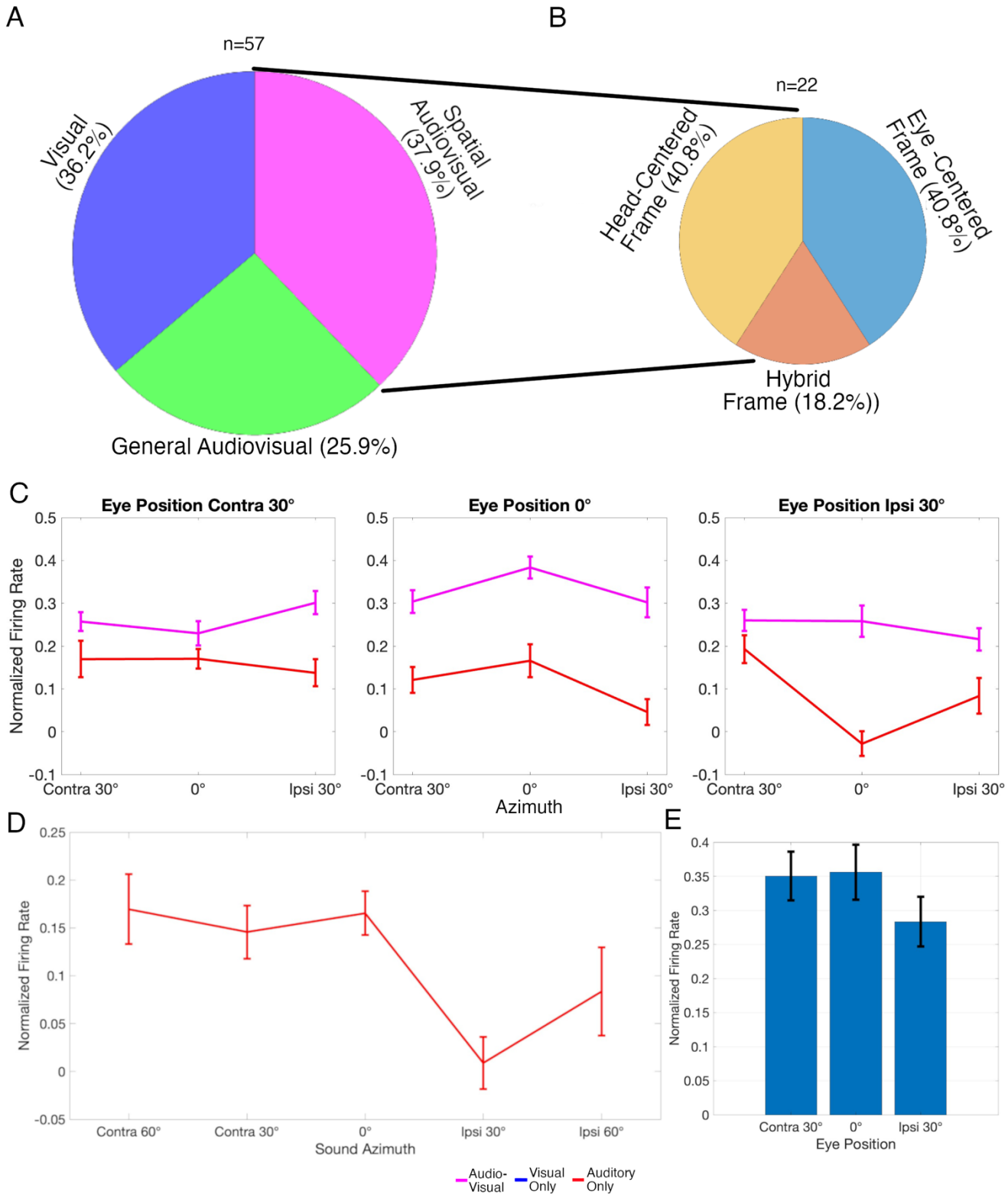

Figure 5.8 Audiovisual Spatial Tuning. A B, pie charts of the percentage of neurons showcasing certain audiovisual effects. **A** shows the percentage of neurons with audiovisual interaction and spatial audiovisual interactions. **B** shows the percentages of audiovisual spatial neurons in different reference frams **C** A set of plots at each eye position for the variations of audiovisual and auditory response across azimuth **D** A plot of the normalized auditory responses for all the neurons showing a significant effect for an audiovisual spatial manipulation across both subjects and all eye positions for the azimuths. **E** a bar graph of the average normalized firing rate at each gaze direction for visual conditions only.

can separate between audiovisual and visual conditions.

To examine if the effects targeted specific parts of space, we then normalized the responses across all these neurons across both subjects to calculate average responses. We first examined the variation of audiovisual and auditory across eye positions for all auditory spatial neurons (Figure 5.8C). The responses indicated that there was little variation across responses in audiovisual conditions but in auditory responses alone the 30^{o} position on the ipsilateral side for auditory responses showed significantly smaller responses in both animals ($p<0.01$, student's t-test) (Figure 5.8C). This trend appeared in a head-centric form as the response at this azimuth remained weaker regardless of eye positions (Figure 5.8D) indicating that auditory presentation ipsilateral space can show weaker auditory responses in AF face patch but not when a visual stimulus was present. Though individual neural responses varied across the population, the mean responses did not vary strongly across eye positions (Figure 5.8E). This lack of variation in visual responses may erase the dearth of responses for auditory stimulus for the audiovisual condition. Altogether, these responses indicate that audiovisual interaction in AF face patch neurons depends on spatial factors but do not have a clear pattern of spatial tuning across the neural population.

5.3.6 Experiment #3 Discussions

Using the 3D virtual reality dome, we revealed that the relative spatial position of auditory and visual stimuli could have a large impact on the responses of neurons in AF face patch. These responses may match responses from the superior colliculus where some misalignment of auditory and visual stimuli can degrade audiovisual response while other arrangements of misalignment can increase the response (Stein and Stanford 2008, Lee and Groh

2012). Unfortunately, where superior colliculus neurons can map out space with receptive fields, these responses in AF face patch did not show a tuning over the population at large. AF neurons instead were tuned to various combinations of eye positions, sound azimuth, and sound elevation in both eye and head centered reference frames. These tunings may stem from connections with VIP, which shows more consistent spatial tuning and hybrid reference frames (Mullette-Gillman, Cohen et al. 2005, Schlack, Sterbing-D'Angelo et al. 2005, Avillac, Ben Hamed et al. 2007). AF face patch may then receive the information from VIP but may be more selective in combining them. Another explanation may be that VIP does not influence audiovisual responses in AF face patch and instead connections with auditory cortex and IT cortex primarily determine the audiovisual responses in this region. The audiovisual variations across space may then be part of the first steps of aligning these features spatially. However, the human STS appears to respond more strongly spatially coincident audiovisual stimuli in fMRI (Meienbrock, Naumer et al. 2007, Plank, Rosengarth et al. 2012). This result may not correspond precisely to the response of individual neurons or may vary between the STS of macaques and humans but it suggests that correspondence between auditory and visual stimuli in space remains important in the STS. Importantly, the effects of shifting eye position remain unexplored in visual-only responses in AF face patch and the STS. These overlapping considerations indicate that more work is required to understand the exact effect of spatial position of sounds and images on AF face patch and across the entire STS.

5.4 Discussion and Future Directions

These experiments have begun to examine previously unexplored spatial principles in the STS of macaques and the purposes these AF audiovisual responses may serve in the macaque

brain. All these results indicate that spatial elements likely modulate the audiovisual responses of AF neurons. First, these results indicate position of the face could influence this modulation with faces pointed different directions causing shifts in audiovisual enhancement. This audiovisual modulation may enable AF neurons to understand the direction of calls and if they are targeted towards other conspecifics. Second, these results indicate that 3D spatial factors can also influence the audiovisual modulation of AF neurons. Multisensory looming signals have previously shown effects on the STS, but sizes and distances also influenced the AF responses to audiovisual stimuli. These results indicate that AF then might play a role in broader understanding of the spatial aspects of social communication. Finally, these results indicate that the position of auditory stimulus relative to the visual stimulus can shift, sometimes drastically, the audiovisual modulation of these regions.

Overall, these results combined indicate that spatial elements of audiovisual stimuli can dramatically affect single neuron responses in this region but do not appear to contain a specific tuning for a particular spatial manipulation. Rather, neurons show a wide range of manipulations can increase or decrease the spiking response to audiovisual stimuli. The overall population of neurons then may reflect these changes in rate or distributed codes across a wide number of neurons rather than with specific tuning like the tuning for space in the auditory cortex (Werner-Reiss and Groh 2008, Ortiz-Rios, Azevedo et al. 2017). These manipulations could also indicate that while space may play an important role the region itself does not encode audiovisual space. Finally, another possible explanation may be that these differences stem from the focal nature of the recording with the microwire bundle (McMahon, Bondar et al. 2014). The whole of AF face patch may have a clearer preference than the specific positions sampled by this recording technique. Future work should more widely sample both AF face patch and the STS to determine

if neurons show specific neural tuning. Similarly, using these paradigms in intraparietal regions or in prefrontal regions may unveil regions with more consistent tunings of audiovisual space that receive information from the STS.

However, while these shifts in activity demonstrate a spatial sensitivity in audiovisual responses within these neurons, these experiments remain simple passive viewing tasks within a restricted setting that required no distinguishing between positions in space or no competing stimuli. In previous studies adding multiple stimuli could shift audiovisual responses with little spatial sensitivity to more spatially tuned (Spence 2013, Fleming, Noyce et al. 2020). In AF face patch, stimulus competition could more clearly align the tuning of audiovisual spatial responses. Importantly, the STS may play a role in decoding social interactions between various members of the same group. This function may then require multiple macaques and signals targeted toward another conspecific to fully express audiovisual spatial tuning. Future work should evaluate more competitive stimulus environments to determine the principles of audiovisual modulation in space more clearly. Overall, this chapter has evaluated the role of AF face patch in combining audiovisual signals in space. The next chapter will take a comparative approach to understanding social regions in the temporal lobe of mammals.

6. Comparative Physiology of Carnivores and Primates

6.1 Introduction

The previous chapters established audiovisual responses, spatial responses, and the combined effect of these features in AF face patch. These results furthered the understanding of how the temporal lobe of the primate combines information. In contrast, this chapter explores a comparative approach to understanding the balance of specialization and combining of information in the temporal lobe. Comparative neuroscience can help explore the common bases for neural functions and the development of areas. More thoroughly, looking across species outlines the evolutionary purposes of regions and potentially shared homology. While the face patch system has mainly been studied in macaques and humans, many other vertebrate species have shown the ability to recognize conspecific and human faces (Leopold and Rhodes 2010). Behaviorally, many species depend on the ability to visually recognize faces or gain and convey information from them. Dogs, for example, must recognize the faces of their owners as compared to other human faces (Adachi, Kuwahata et al. 2007). Similarly, domestic sheep must distinguish faces of various handlers, other sheep, and sheep-herding dogs (Tate, Fischer et al. 2006). Similarly, various other species such as killer whales and bottlenose dolphins show signs of self-recognition when looking into a mirror (Delfour and Marten 2001, Marino, Connor et al. 2007). Beyond this recognition, many species express emotion through changes in facial features like dogs and elephants and can follow the gaze of conspecifics like goats (Langbauer 2000, Kaminski, Riedel et al. 2005). These wide array species with emotive facial movements and recognition signals that these behaviors stem from a common ancestor. However, these common behaviors do not necessarily suggest that all these species have developed the cortical apparatus present in primates and requires deeper study.

In terms of neural specialization, while many mammalian species outside of primates do not appear to have regions of consistent face-selective cells such as the face patches, some have face cells in the cortex. Outside of primates, the most thoroughly studied mammalian species for single neural specialization are sheep. Neurons in temporal cortex of sheep actually show face selectivity (Kendrick and Baldwin 1987). More importantly, many of these cells can detect differences between individuals and can distinguish specific features including the length of horns. These features largely resemble the features of various face neurons across the primate face patch system including viewpoint selectivity and tuning to distinctiveness (Freiwald and Tsao 2010, Koyano, Jones et al. 2021). Other species have remained largely unstudied for conspecific recognition in single neurons, but fMRI and neuroimaging have suggested that dogs, have patches that respond to human faces over non-face objects. These regions also lie in the temporal lobe signaling that this region, outside of primates, can also specialize to commonly viewed and important visual stimuli including faces (Cuaya, Hernandez-Perez et al. 2016). The presence of face cells and face processing regions in these other animals suggests that the temporal lobe may contain a homologous bauplan for specialization.

Despite this work, the evolution of face processing remains mysterious and developing further models of face processing could elucidate more features of how the face processing system is assembled in humans and primates. The lack of many species with clear face-processing systems may suggest that face neurons in sheep is merely convergent evolution. However, other domestic animals may occupy a similar niche and provide further examples of species with this kind of specialized neuron. One example may be the domestic ferret. The domestic ferret lives in colonies with other ferrets and has consistent interactions with humans (Larrat and Summa 2021), which may, like in dogs and sheep, lead to conspecific and human

face selectivity in the cortex. Alternatively, ferrets have a variety of visual displays to communicate emotional state to conspecifics such as the weasel war dance (Fisher 2006), though it remains unclear if ferrets can recognize one another. These behaviors may require neural circuitry that can at least detect the position of a body or face. Ferrets may not have face or body selectivity in the same pattern these other species might but may instead have face-selective or body-selective neurons that facilitate more social functions, which could also provide further insight into the evolution of selective regions in the temporal lobe. Importantly, developing the ferret as a model for social vision could enable further study of face selective regions in a more flexible species. With a ferret model, studies could examine face-selective neurons in more naturalistic contexts, such as free-moving and active sensing that could radically change the tuning and functions of neurons in this region. In this chapter, we record from ferret temporal lobe while the subjects view various objects including ferret faces and bodies.

6.2 Materials and Methods

6.2.1 Subjects

Subjects included 6 pigmented ferrets which were all female, ranging from 1-6 years old and 650-950g in weight. Ferrets were all housed in groups of 2 or more in enriched housing. Before tasks and recording, ferrets were water-restricted and obtained at least 60ml/kg of water through task performance or through supplementation. Water consumption and weight of ferrets were closely monitored to ensure the health and safety of the subjects. All procedures were approved by the local ethical review committees at University College London and the Royal Veterinary College and conducted under an approved UK Home office license in accordance with all regulation of the Animals Act of 1986.

6.2.2 Experimental Design and Setup

Ferrets performed a simple viewing task to assess single cell selectivity, similar to those used in macaques. Stimuli were presented using a gamma-corrected monitor presented 40cm in front of the subject. Ferrets conducted both passive and active viewing tasks. For passive viewing, the ferret was trained to maintain its position and to place its mouth on a spout, as detected by an infrared beam, to begin stimulus presentation, receiving a reward for maintaining its position at the start and the end of a trial. Stimuli were presented at 15° visual angle for 200 milliseconds in trials of three stimulus presentations. In the active task, the subject again had to maintain position at the spout to initiate a trial. Following the presentation of stimuli, a checkerboard stimulus appeared signaling the subject to move to another spout to receive a reward, requiring the ferret to actively detect the change from naturalistic stimuli to the checkerboard. Stimuli were presented with a custom GUI designed in MATLAB. Timing and rewards were controlled by a TDT RX6 Multifunction Processor and PTB-3 in MATLAB.

6.2.3 Electrophysiology

All ferrets were implanted with a WARP 32 microdrive, which contains 32 independently movable tungsten electrodes. After a period of recording of 3-4 weeks, each electrode was advanced approximately 200 microns to sample new cells. In passive recording experiments, ferrets were implanted primarily over auditory regions but with some extending electrodes into nearby visual areas including those bordering area 20b. For longer term experiments and active recording. Ferrets F2201 and F2203 were implanted over areas 20a, 20b, as well as nearby the nearby secondary visual areas such as area 21 and auditory regions such as PEG and MEG. With

this wide array of recordings, we could thoroughly assess selectivity in temporal areas of ferrets. Recording was processed using a W2100 wireless recording system using an IFB-C control board (Multi-Channel Systems, Harvard BioScience) and collected as a broadband signal of 0.5Hz-10000Hz

6.2.4 Stimuli

The comparisons of these regions required direct interrogation of ferret area 20 and other visual areas with social and naturalistic stimuli. To examine the role these regions played in processing social stimuli, we captured pictures of familiar and unfamiliar ferrets faces and bodies within the ferret colony as well as pictures of unfamiliar human faces. Neural responses to these stimuli will be compared to neural response to other non-face stimuli including a variety of scenes and unfamiliar objects. These stimuli will be presented to ferrets in the passive-viewing task in a randomly interleaved order.

6.2.5 Data Analysis

All recordings will be sorted into spikes using the wave_clus automatic spike sorting algorithm. Following sorting into spikes, we will assess neurons from the passive-viewing experiment for selectivity to faces. First, we will isolate the window of stimulus presentation, compute the average spike rate as compared to a baseline window, and conduct an ANOVA comparing each of the categories to assess if neurons exhibit a significant category modulation. Then, we will compute a selectivity index (SI), similar to those used in chapters 3 and 4 and other studies (Tsao, Freiwald et al. 2006), with the following equation:

SI = (Response_Target – Response_NonTarget)/(Response_Target + Response_ NonTarget)

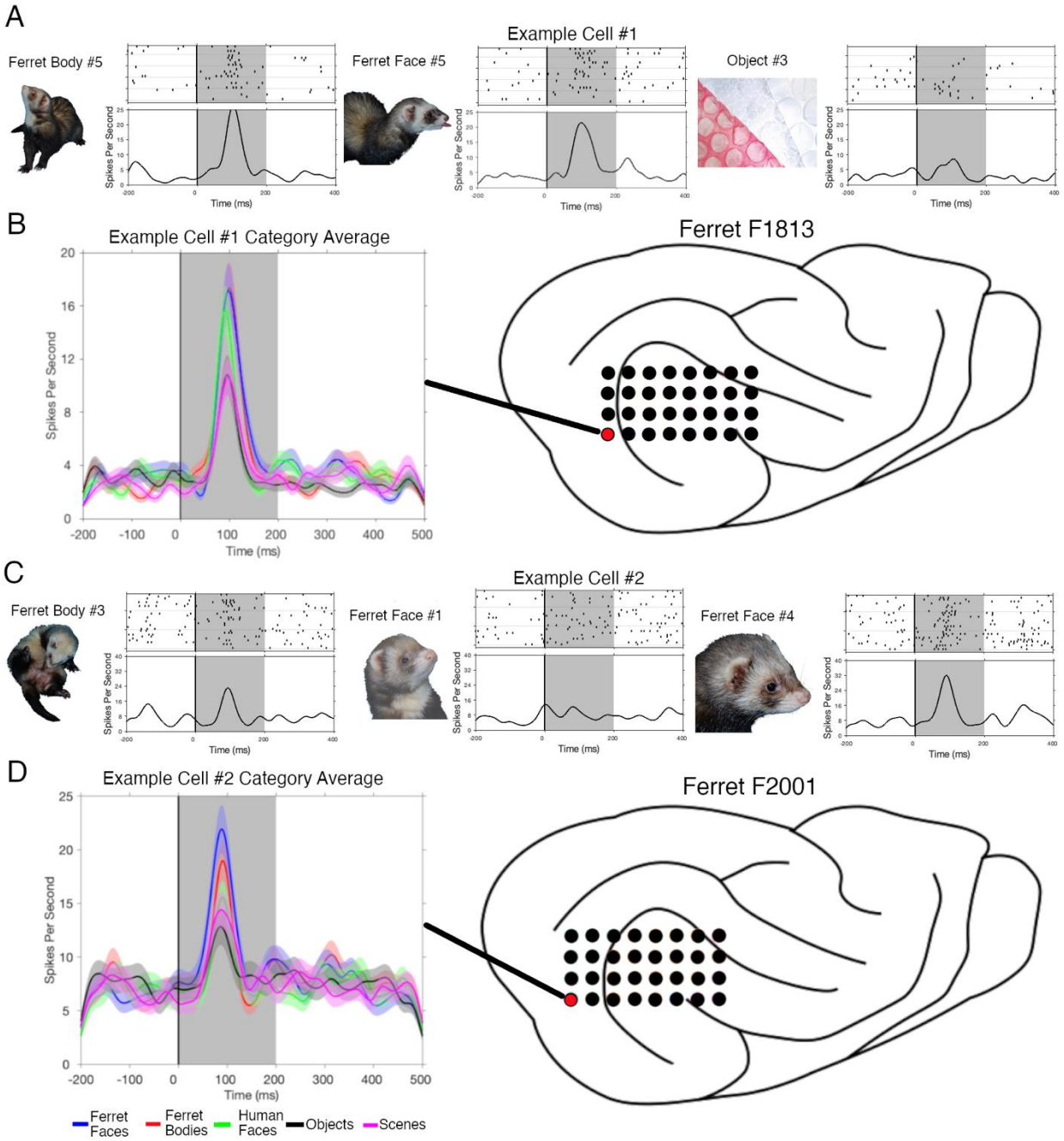

Figure 6.1 Examples of Tuning in Ferret Cortex. A, C. Raster plots and the SDFs for the response of an example neuron to faces and bodies as compared to a non-face object (**A**) and for the responses of an example neuron to different faces and bodies (**C**). **B, D.** the mean SDF for all categories for each example face-selective neuron and a schematic of the electrode placement in the cortex. Diagrams in **B, D** adapted with permission from (Poole 2023)

Where the target response is the mean response for ferret faces, ferret bodies, or human faces individually and the non-target response is the mean response to scenes and objects.

Neurons that exhibit an SI greater than 0.333 or less than -0.333 are classed as ferret face, ferret body selective, or human face selective.

6.3 Results

6.3.1 Face and Body Tuning in Passive Recordings

We recorded from 4 ferrets with electrodes placed over the auditory cortex and in a variety of location in the visual regions bordering auditory cortex. To distinguish visually responsive neurons, we used a T-test comparing the response to baseline for all stimuli ($p<=0.01$) resulting in 15 visually responsive neurons. Among these neurons, many displayed tunings for ferret faces, human faces, and ferret bodies or non-face objects (Figure 6.1A). Some selective neurons did not display major differences between human faces and ferret faces indicating that these neurons are tuned to faces as a category rather than ferrets or conspecifics (Figure 6.1B). Conversely, some displayed differences within category showing greater responses for specific images as compared to others of the same category (Figure 6.1C). Similarly, neurons showed variance across categories of preferred tunings (Figure 6.1D). Importantly, images of ferret faces and bodies presented included familiar ferrets while the human faces were all completely unfamiliar.

Of these visually active neurons, 80% (12/15) showed selectivity for human faces, ferret faces, or ferret bodies with a SI greater than 0.333 or less than -0.333. This tuning often stemmed from the neurons in the visual cortex bordering along the major auditory regions (Figure 6.2A-D). Among these 12 neurons, 10 showed ferret body selectivity, 8 ferret face selective, and 5 human face selective but with considerable overlap (Figure 6.2E). Some appeared to be in regions of auditory cortex near these bordering visual cortices, but the electrodes likely deflected to or penetrated the more visual regions lying beneath in the Suprasylvian Sulcus. Other face

selective neurons rested in both sections of auditory cortex and in more dorsal regions of the visual cortex. While these neurons indicate that selectivity may be more widespread, these responses all require further examination across the cortex.

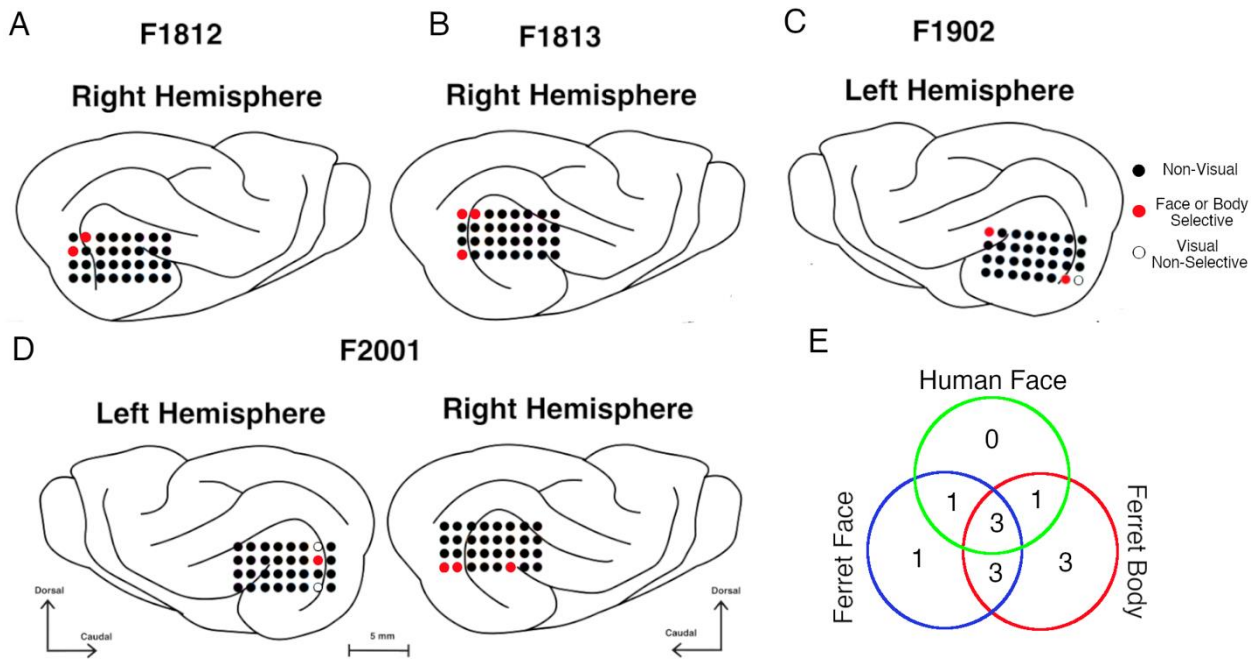

Figure 6.2 Summary of selective responses across ferrets. A-D maps of all the locations of visual responsive neurons in every recorded ferret. Each circle represents an electrode location while the color indicates the visual responsiveness and selectivity at each site. **E** a Venn diagram of the selectivity of individual neurons and the overlap among all the neurons (n=12). **A-D** diagrams repurposed and adapted from (Poole 2023)

6.3.2 Behavioral Responses

To expand from passive viewing into more active viewing tasks, we trained to ferrets to wait for a checkerboard stimulus following a series of images. The ferrets, after training, maintained position at a forward-facing spout until the presentation of the target stimulus (Figure 6.3A), after which the ferret had to turn to another spout to receive a reward (Figure 6.3B). Subjects were trained until they reached at least a 70% completion rate after staring a trial. In these completed trials the time from start of a trial to making the correct response varied more than the length of a trial according to a Brown-Forsyth test of variation ($p<0.0001$) (Figure

6.3C). The variation of responses indicates that the ferret responded when it perceived the target stimulus rather than moving at a consistent time. With this active task we can explore the responses of area 20 neurons in implanted animals to examine neural responses when the ferret must attend to the visual stimulus.

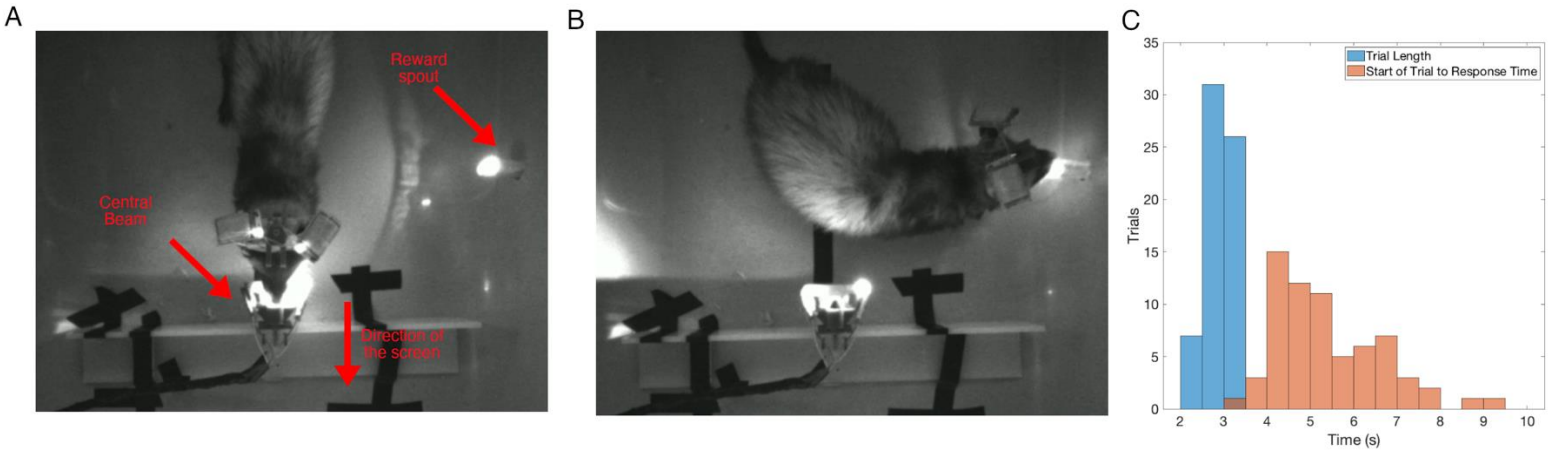

Figure 6.3 Behavioral Responses of Ferrets. A the first phase of the task where the subject held position at a central cone to initiate the trial and viewed stimuli before the appearance of the target. **B** the completion of the task when the ferret moved to the reward spout. **C** a histogram comparing the time of a trial (blue) to the time from the initiation of the trial (orange).

6.4 Discussion and Future Direction

6.4.1 Tuning for Naturalistic Stimuli in Ferret Cortex

These results remain preliminary but suggest that a population of neurons within visual regions of the ferret could be selective for faces or bodies. Some cells appeared in visual regions bordering auditory cortex including area 20b and area PS, ventral regions of the ferret visual system. Other selective neurons also resided in more early visual areas and dorsal visual areas, such as area 21 and area PMLS, which responds selectively to motion and corresponds more with similar dorsal visual areas of the primate cortex (Dunn-Weiss, Nummela et al. 2019). This result may not be entirely surprising as the face patches of marmosets rest in earlier visual areas, V2 and V4 (Hung, Yen et al. 2015). However, few primates show face or body selective neurons

in any dorsal visual regions. These variations require further investigation and direct targeting of these visual areas for recordings to accurately assess the number and location of these selective neurons. Interestingly, though these results are preliminary, neurons in ferret cortex appeared more tuned to bodies than faces alone. Macaques also have body patches in addition in the face patches in the temporal lobe (Popivanov, Jastorff et al. 2012, Popivanov, Jastorff et al. 2015, Popivanov, Schyns et al. 2016). These patches operate in parallel but sit alongside the face patches (Premereur, Taubert et al. 2016, Hesse and Tsao 2020). Body selectivity in ferrets may provide another course of study to further understand the origins of social vision but the relationship between body processing across species remains mysterious. These selective responses require further study across to species to provide a clearer understanding of the relationship of these visual processing networks and the origins of social specialization in primates.

These selective responses also indicate some level of homology or analogy between higher level visual areas in the ferret and primates. Importantly, the ferret has lower acuity vision than primates or sheep (Veilleux and Kirk 2014) and likely uses vision as a method of tracking motion in activities such as hunting like in other lower acuity mammals such as mice (Hoy, Yavorska et al. 2016). While the selectivity of neurons is often assessed with static images, biological motion may further stimulate neurons in ferret visual regions as it does in the primate STS (Jellema, Maassen et al. 2004, Russ, Kaneko et al. 2016). We also did not strictly track the eye position of the ferrets during this experiment though they were required to position their head and body towards the screen. Experiments in macaques often require strict fixation for examining face responsive neurons but some studies have found that strict fixation may not be necessary to delineate face responses for dynamic stimuli (Russ and Leopold 2015). It also

remains unclear if ferrets use eye movements independently from movement of the head to foveate stimuli. Studies that require fixation or that incorporate biological motion may further show face or body selectivity in the ferret temporal cortex. Conversely, macaque temporal cortex has been assessed with more categories of image (McMahon, Bondar et al. 2014) and the selective neurons found here in the ferret may also respond to other categories. Further examination with more categories may be necessary to confirm this selectivity. However, these initial results indicate that visually selective neurons in exist in the ferret visual cortex.

6.4.2 Visual Specialization During Free Movement

These results also indicate that ferrets can also be trained to conduct visual-form dependent tasks, which has not before been demonstrated. Previous work has shown ferrets can perform higher order motion-related tasks as well and that dorsal visual areas respond to complex motion features (Dunn-Weiss, Nummela et al. 2019). Both these results indicate that the ferret can both detect and respond to some higher-order visual signals. Again, given the ferret's relatively low visual acuity, it remains unlikely that ferrets can perform all the tasks that primates or other more visual species can, but these tasks do indicate that some exploration of higher-order visual behaviors remains possible. The ferret may serve as another model system for analyzing both the ventral and the dorsal stream of vision without the limitations of macaques

While higher-order regions in the visual cortex are more thoroughly studied in the primate, these visual studies have rarely been able to allow the subject to move and explore its environment (Leopold and Park 2020). Free movement has enabled the discovery of different properties of visual regions, such as studies of the mouse in early visual cortex that have shown that locomotion can sharpen the tuning of individual neurons (Niell and Stryker 2010). Free movement has largely only been explored in macaques through free eye movements. While these

eye movements triggered changes in the responses of earlier visual regions, they largely did not affect neurons in the temporal lobe, which instead responded mainly to stimulus movement or target features (Sheinberg and Logothetis 2001, Russ and Leopold 2015, Russ, Kaneko et al. 2016). However, these free-viewing paradigms still lack the locomotion of natural movement that shapes natural vision. Developing a ferret model of higher order vision may enable further study of these regions while the ferret can move and explore its environment. Though ferret visual tasks require further development and deeper exploration, these results indicate that ferret could provide numerous advantages in understanding the effects of locomotion and active sensing on high-level vision.

6.4.3 Broader Implications and Future Directions

More broadly, selective neurons in the temporal cortex of ferrets indicates that some mammalian species outside of primates may have a proto-organization in more ventral temporal areas for form-based. These results, if further verified and more thoroughly explored, may establish a link between association areas rarely examined across mammalian species. By establishing selectivity in the temporal lobe of ferrets, we can better understand the mechanisms the mammalian brain uses to process social stimuli and, more particularly, the origins of the face or body patch system in the macaque. These responses may also stem from convergent evolution rather than a common ancestor. The existence of face and body selective neurons may then depend on other conditions such as domestication or living in large social groups. Further investigation of other species of domesticated or social animal may also provide more insight into the nature of these single neuron functions. Similarly, these cells in the ferret need further assessment to compare primates and even sheep. In macaques, selective neurons have tuning not only to faces or bodies but also specific features and face distinctiveness (Popivanov, Schyns et

al. 2016, Chang and Tsao 2017, Koyano, Jones et al. 2021, Waidmann, Koyano et al. 2022) while sheep face cells can show tuning for the size of horns and familiarity of the face (Kendrick 2006, Tate, Fischer et al. 2006). Ferret face neurons require further investigation to determine if they have tuning to any features of faces or bodies. These neurons in ferrets could also express audiovisual integration similar to the integration expressed by AF face patch in Chapter 3. They also sit at the border of visual and auditory regions with connections between them (Dell, Innocenti et al. 2019). Audiovisual responses would provide another parallel between these regions but may also indicate that these responses in ferrets may subserve more social functions. Overall, these results suggest that ferrets contain selective neurons but require more investigation to understand the similarities and differences of the visual system between ferrets and other species. In the next and final chapter, we examine the results of all previous chapters and elaborate on future directions we plan to explore.

7. General Discussion

7.1 Conclusions

Combining macaque and ferret electrophysiology this research has explored the many ways the temporal lobe combines seemingly separate information to begin weaving together a complete percept. First, we provide the first evidence that neurons in a visually selective region of the macaque cortex, AF face patch, are modulated by the addition an of auditory stimulus. Additionally, audiovisual modulations were only observed when the visual stimulus was a face but persisted when the vocalization was replaced with matched noise. These responses suggest this region of the macaque brain plays a larger role in social and audiovisual communication. The STS has previously shown multisensory responses (Bruce, Desimone et al. 1981, Baylis, Rolls et al. 1987), but the discovery of visually selective regions overshadowed the potential role of face patches in multisensory processing. These results, however, fit into a newer framework of the visual system that theorizes that the STS is part of a third pathway that is specialized for social vision in both macaques and humans (Pitcher and Ungerleider 2021). As partially described in Chapter 3, the STS also connects with other audiovisual social regions including the VLPFC and the amygdala (Aggleton, Burton et al. 1980, Seltzer and Pandya 1989), which also have face selective neurons (Sugihara, Diltz et al. 2006, Gothard, Battaglia et al. 2007, Romanski and Diehl 2011). AF face patch may provide part of the information needed for these regions to also connect face information with auditory information. The STS then may more broadly enable social communication rather than merely social vision and AF face patch may serve as a starting point for connecting auditory and visual information specifically for faces and conspecifics.

We also determined that AF face patch neurons showed distinct tuning to the physical size in 3D space of faces and objects. This tuning did not depend on retinal angle, which has

been the traditional measure of size in visual experiments. The tuning most prominently preferred faces of the largest and smallest sizes even when theses sizes extending into the dramatically unrealistic. This preference emerged in the part of the neural response indicating that longer presentations may reveal more neural tuning to real-world physical properties. These responses demonstrate that AF face patch neurons receives and responds to 3D spatial information about faces. More broadly, these results also reflect the combination of spatial and shape information in the STS and temporal lobe. This combination may again implicate AF face patch in a third visual pathway that sits between the tradition pathways and receives information from both. Previous work has demonstrated the sensitivity of AF face patch and the STS to biological motion (Fisher and Freiwald 2015, Russ and Leopold 2015) further indicating features more commonly associated with the dorsal stream can reach AF face patch and the STS. Physical size information in the STS may then contribute to understanding how to interact socially with a conspecific. Physical size could be very important in the ethology of macaques during aggressive behavior. However, neurons across IT have sensitivity to 3D structure (Yamane, Carlson et al. 2008, Vaziri, Carlson et al. 2014) but it remains unclear if other face patches deeper in the IT cortex also have tuning to physical size. Further study of the more caudal face patches and other regions of the STS is required to determine if this feature is excusive to the STS or if the tuning for physical size extend to parts of IT cortex. Interestingly, previous work has suggested that the dorsal stream can also encode a courser version of shape information (Freud, Ganel et al. 2017). The responses in AF face patch may similarly reflect a course encoding of 3D physical features to perform actions. Alternatively, the connections between these regions may enable continuous interaction to compute the 3D size and shape during action (Janssen, Verhoef et al. 2018). The relationship of the dorsal and ventral stream within the STS remains unclear but these results

provide a first step to understanding how spatial and form information combine in the temporal lobe.

We also examined AF face patch for sensitivity to spatial factors of audiovisual integration, combining elements of both chapters 3 and 4. The STS has been examined for its role in temporal binding of audiovisual stimuli (Barraclough, Xiao et al. 2005) but has rarely been examined for spatial changes in auditory and visual stimulus components. Indeed, our neural recordings indicate that spatial factors of both visual and auditory influence the response of AF neurons. First, we demonstrated that the head azimuth of a visually presented face stimulus could shape the audiovisual modulation. Visually, STS neurons show different responses to different head orientations (Desimone, Albright et al. 1984). However, despite audiovisual work in association regions, it remains unclear how tuning in one modality affects the response to multisensory stimuli. Second, we found that the absolute size and distance of the visual stimulus could also shift the audiovisual modulation of AF face patch neurons but without a tuning for a specific size or distance independent of the visual tuning. Previous assessments of distance in audiovisual responses have indicated that smaller stimuli at longer distances increase audiovisual behavioral facilitation (van der Stoep, Serino et al. 2016, Van der Stoep, Van der Stigchel et al. 2016) suggesting that these neurons may produce more consistent audiovisual modulation at similarly smaller sizes and further distances. Finally, we found that the binding of visual and auditory information can vary dramatically across different spatial manipulations such as changing sound azimuth and eye position. These tasks changed the eye position without changing head position separating reference frame for the eye from reference frames centered on the head. Visually, neurons in these regions are assumed to be invariant to eye position changes but we find changes here in both the visual and the audiovisual domain. Importantly, the STS is

connected to parietal regions that shows hybrid coordinate reference frames for audiovisual stimuli (Seltzer and Pandya 1984, Mullette-Gillman, Cohen et al. 2005). In AF face patch, these experiments indicate that AF neurons have a variety of reference frames, but many depended on a conjunction of eye positions and head position. However, in other regions, motion and tasks shifted the reference frame of encoding over time (Mullette-Gillman, Cohen et al. 2009, Caruso, Pages et al. 2021). It is possible that a spatial task beyond passive fixation may further unveil reference frame tuning in AF face patch. Altogether, these results indicate that the temporal combines spatial and audiovisual information in complex patterns that still require further investigation.

Finally, we compared the temporal lobe of macaques to that of ferrets by searching for face and body selective neurons in ferret visual regions. We discovered initial evidence of face or body selective neurons in regions of visual cortex bordering the auditory cortex similar to the macaque STS. Some of these neurons also showed sensitivity to human faces as well indicating that they were responsive to faces as a category rather than low level features. If confirmed, these responses would represent the first selective neurons discovered in the ferret. These results are preliminary but may indicate a previously unknown comparative link between primates and a carnivore species. However, it remains unlikely that ferrets exhibit selectivity in the same ways as macaques. On the cortical level, ferrets likely do not have specific regions of cortex with concentrated populations of the neurons. A previous search for higher order feature selectivity in ferret auditory regions did not find a patch like the voice patch in macaques (Landemard, Bimbard et al. 2021), though single neurons may still respond to ferret calls or voices. This evidence further suggests that ferrets will likely not have concentrated patches of selective neurons. Similarly, the face and body selective neurons discovered may serve different functions

that selective neurons in macaques. Neurons in the macaque can be sensitive to expression and gaze direction (Jellema, Baker et al. 2000, Jellema, Maassen et al. 2004, Taubert, Japee et al. 2020) rather than solely to the identity of a face. Ferret selective neurons may embody more of these functions to determine gaze direction or body posture rather than the same precise social functions, but the tuning of these neurons and their role in behaviors requires further investigation. These responses also indicate other carnivore species such as the cat, may also have face selective neurons that remain unexplored. Incorporating more new species in studies of visual selectivity would further our understanding of the evolution of these selective neurons and how specialized regions such as the face patches Altogether, these chapters reveal many previously unknown features of the visual and auditory processing in the temporal lobe and open new and exciting questions that require deeper study.

7.2 Broader Implications

7.2.1 Theory and Organization of Audiovisual Processing

As emphasized in chapter 3 and 5, these studies provide new facets to the overall theory of audiovisual integration. These results indicate that specialized visual areas can be recruited in the processing of audiovisual stimuli. However, in contrast to more traditional models (Stein and Stanford 2008), AF face patch does not fit neatly into the integration proposed for secondary association areas or primary sensory areas. This audiovisual function overlaps with tuning for high-level visual features such as identity, expression (Taubert, Japee et al. 2020, Koyano, Jones et al. 2021), and size (Chapter 4) but how these functions interact with one another remains mysterious. Despite these audiovisual responses, the neurons in the region largely responded less to audiovisual integration with the expanding disc stimulus. These overlapping functions indicate that AF face patch is involved in a variety of functions of social interaction depending on the

stimulus presented. The STS more broadly plays a role in audiovisual integration during presentation of stimuli that both include and do not include faces (Barraclough, Xiao et al. 2005, Dahl, Logothetis et al. 2009, Dahl, Logothetis et al. 2010). It appears to combine information based on visually specific features and matched temporal dynamics with semantically aligned actions (Barraclough, Xiao et al. 2005) similar to the results found in AF face patch. These pieces of evidence suggest that networks of audiovisual integration can and must recruit multiple regions depending on the stimulus features involved and the attention needed.

Similarly, neurons in the voice patch of macaques have a specialization for macaque voices but respond to multisensory stimuli. These voice patch neurons show specific tuning to auditory features but not visual features, an inverse of the results in AF face patch (Perrodin, Kayser et al. 2014). These inverted functions between these two areas suggest that the responses these higher-order areas can contribute to audiovisual integration based on the demands of the environment. From these results, we can suggest a new model of audiovisual integration based on the recruitment of specialized areas, either through feed-forward or feedback signals, depending on the attentional or environmental demands of the tasks required. In this model, structures such as the superior colliculus are recruited for orienting or shifts of gaze towards audiovisual stimuli whereas areas such as AF face patch are recruited for identifying the identity of conspecific making a vocalization. To further evaluate this model, future work needs to assess task specific effects rather than passive integration in AF face patch to understand if attention to specific features or competition between multiple audiovisual stimuli can affect AF neurons.

7.2.2 Face and Body Processing Across Species

This thesis also demonstrates that face neurons may exist in the ferret providing new insights about the organization of face processing in mammals. Importantly, while the face

patches exist in macaque, marmoset, and humans, the relationship of these regions between species remains mysterious (Tsao, Moeller et al. 2008, Pinsk, Arcaro et al. 2009, Hung, Yen et al. 2015, Hesse and Tsao 2020) even before considering the ferret. Humans and macaques show perhaps the most homology. Human face patches include occipital face area (OFA), fusiform face area (FFA), STS face area (STS-FA), and Anterior STS face area (aSTS-FA) (Tsao, Moeller et al. 2008, Pinsk, Arcaro et al. 2009). The STS-FA and aSTS face area appear most aligned with AF face patch as they show preference for moving face (Tsao, Moeller et al. 2008, Zhang, Japee et al. 2020). These results highlight the human STS face patches as a part of the third visual pathway in humans like AF face patch in macaques. However, humans appear to have combined certain face patches or lack some patches that appear in macaque. Therefore, the exact relationship on the neural level remains an area for further study. Results from the marmoset further complicate these comparisons and present alternate models for the organization of the face patches. Marmosets, like macaques, appear to have six face patches in the occipitotemporal pathway but two face patches appear in visual areas as early as V2 and V4 while the later patches appear across the temporal lobes (Hung, Yen et al. 2015). These early face patches indicate that face perception may have origins in earlier parts of cortex in other species.

Functionally, much investigation of face processing across species has emphasized the decoding of identity (Hesse and Tsao 2020). Some models of the face patch system in macaques have found evidence that tuning becomes more invariant to variations in head angle and expression as neural responses progress from posterior to anterior (Freiwald and Tsao 2010). However, while this system undoubtedly supports these functions, information about these non-identity features remains important and the visual system must use these features to understand social interactions. The correspondence of face motion processing in the STS across macaques

and humans indicates that the combined divide between the STS face patches as a site of behavioral recognition. The face patch system may then have stemmed from social systems designed to visually understand expression or head angle. Many mammalian species behaviorally understand these features even when incapable of visual face identity recognition (Leopold and Rhodes 2010). Ferret face processing may also stem from the social specialization proposed in the third visual pathway hypothesis. The ferret may use these functions primarily for detection of other ferrets or of ferret mental state in numerous behaviors.

Similarly, our results also indicated that neurons in the ferret may also have tuning for bodies, which, for many smaller mammals, may have more behavioral relevance than the face alone. Across primate species, body selective areas exist in parallel to face processing systems (Pinsk, Arcaro et al. 2009, Popivanov, Jastorff et al. 2012), but unlike face patches, the correspondence between them remains more unexplored. In evolutionary terms, body responsive cells may have arisen before face cells to detect conspecifics. As visual functions became more dominant in social interactions these systems may have separated into distinct systems. But some studies have indicated, controversially, that body responses may be found in face patch regions. Microstimulation in both systems has indicated they remain separate (Premereur, Taubert et al. 2016) but in both human and macaques, some face patches respond more strongly to faces when combined with bodies including AF face patch (Fisher and Freiwald 2015, Taubert, Ritchie et al. 2022). Some body patches also respond more strongly to dynamic bodies suggesting that these patches too may also participate in social functions (Pitcher, Ianni et al. 2019, Raman, Bognar et al. 2023). Alternatively, social specialization rather than being a product of these face and body tuning regions may instead be their origin. Future examination of ferret visual regions with moving faces and bodies could potentially demonstrate social specialization in ferret neurons,

which would indicate that face tuning may have arisen from social interaction. Combined with previous work, the results of this thesis highlight new paths of study to understand the evolution of the face and body processing systems more clearly and that each of these model systems requires more investigation of their comparative relationships to understand results in different species.

7.2.3 Single Neurons and the Population Doctrine of Neural Processing

This thesis broadly focused on single neuron tuning to features to evaluate the how the region combines different forms of sensory information. This course of study contrasts with new attention to overall populations responses, which has now become more popular in neurophysiology (Saxena and Cunningham 2019, Ebitz and Hayden 2021). However, the population doctrine requires assumptions on the nature of the state space the neuron inhabits. The goals of these studies were primarily to examine the tuning of responses to these features rather than the patterns of distinguishing stimuli among all the neurons. Within the broader context of the organization of the brain, these single neuron responses must be situated in the larger functions of brain areas to perceive the world. Population doctrine hold that single neuron responses are projections that shape the overall manifold created by the population in state space (Ebitz and Hayden 2021). The results of this thesis may represent previously unknown subspaces occupied by these neurons that can be gated by task and behavioral relevance.

Population tuning principles have previously been applied to the face patches particularly in identity tuning. Interestingly, single AF neurons, even within the same voxel, can have high correlation with visual responses in different regions where one neuron may show broad correlation with early visual regions in IT cortex and IT cortex while another may show high correlation with visual motion regions (Park, Russ et al. 2017, Park, Koyano et al. 2022). These

responses may indicate that AF face patch has multiple different neural populations some shared with other face patches. This is borne out in studies on identity tuning in the macaque face patches. This work has suggested that the various neural responses to faces enter a face space built upon the principal components of the population that can decode the identity of different faces, though the shape and nature of this tuning remains controversial (Chang and Tsao 2017, Koyano, Jones et al. 2021). The tuning to extreme sizes resembles the v-shaped tuning to extreme face identity (Leopold, Bondar et al. 2006, Koyano, Jones et al. 2021) as described in chapter 4. This tuning is shared across face patches, which remains an area for future study, However, the size tuning at the single neuron level in AF face patch then may reflect population tuning to these extreme identities in face space. Size tuning could also reflect another subspace that face patch neurons occupy separate from facial identity that becomes more utilized during certain actions.

Similarly, population responses in AF face patch may reflect a disparate tuning in audiovisual modulation that can coalesce during different task demands. In audiovisual integration, previous work has indicated that when stimulus competition is introduced previously generalized audiovisual behavioral improvements suddenly becomes more dependent on space (Spence 2013, Fleming, Noyce et al. 2020). Within the population doctrine, these results may reflect a change in behavioral state that in turn changes the neural state and requires the brain to draw on the neural population in different ways depending on the task. This framework may explain the seemingly disparate tunings to spatial components in chapter 5, which may align more clearly if the subject performed a task or had to distinguish multiple stimuli. Future work should reapproach these studies through the lens of analyzing population responses and applying different behavioral states now that these results have highlighted new functions and the single

neuron level. While this thesis has shown new forms of tuning at the single neuron level, only through a broader understanding of the population can we truly understand how these regions combine sensory information.

7.3 Current and Future Work

More work is required to fully understand audiovisual spatial integration as well as the comparative aspects between the ferret and the macaque within the temporal lobe. While the preliminary results are encouraging, our investigations of the ferret temporal lobe will continue to assess the number of selective neurons in the ferret temporal lobe. We have implanted two ferrets more directly over area 20 and trained them to respond to a target stimulus as described in Chapter 6. We will be recording and continuing to examine the selectivity of area 20 neurons as well as other visual regions with a visual task. Beyond these tasks, we will also incorporate new stimuli including moving ferrets as well as audiovisual stimuli. By establishing the audiovisual activity within ferret area 20, we can further cement the comparative physiology of this region and the macaque STS.

Moving forward, I also intend to take further advantage of the flexibility allowed in ferret behavior to further understand audiovisual spatial tuning. As opposed to macaques, ferrets can freely roam an environment and their relative size enables us to explore changes more thoroughly in absolute distance as opposed to virtual distance. Importantly, then, ferrets enable a clearer assessment of modulation by distance in the auditory system as well as the visual. We will prepare and train ferrets to traverse a large speaker field for tasks differentiating between competing stimuli to assess audiovisual space in the lateral temporal lobe of ferrets. Recording from area 20 and nearby auditory areas during these free-moving tasks a will elucidate the role of active sensing in completing these tasks of resolving audiovisual competition. We will conduct extensive tracking

of ferret body position and develop new paradigms to track ferret gaze direction during these tasks. Through tracking position and body posture, we can start to delineate the different reference frames used by different populations of neurons in space and by tracking gaze, we can develop a greater understanding of ferret visual behaviors and the neural activity underlying these responses. These approaches will allow us to begin understanding how the brain segment scenes naturalistic scenes using multisensory integration and active sensing.

Through these experiments, we will continue to establish the comparative connections between the ferret and the macaque as well as the features that differentiate these regions between species. Despite the elaboration of the temporal lobe in the primate, the temporal lobe of other species can still enable a greater understanding of the organization and function of the region. These experiments can also elucidate the connection of audiovisual space in the temporal lobe between the ferret and the macaque. Overall, these new approaches and experiments, will continue to investigate how the temporal lobe combines information and how it contributes to real-world perception.

8. References

Adachi, I., H. Kuwahata and K. Fujita (2007). "Dogs recall their owner's face upon hearing the owner's voice." Anim Cogn **10**(1): 17-21.

Aggleton, J. P., M. J. Burton and R. E. Passingham (1980). "Cortical and subcortical afferents to the amygdala of the rhesus monkey (Macaca mulatta)." Brain Res **190**(2): 347-368.

Agrillo, C., A. E. Parrish and M. J. Beran (2014). "Do rhesus monkeys (Macaca mulatta) perceive the Zollner illusion?" Psychon Bull Rev **21**(4): 986-994.

Alais, D. and D. Burr (2004). "The ventriloquist effect results from near-optimal bimodal integration." Curr Biol **14**(3): 257-262.

Allman, B. L., R. E. Bittencourt-Navarrete, L. P. Keniston, A. E. Medina, M. Y. Wang and M. A. Meredith (2008). "Do cross-modal projections always result in multisensory integration?" Cereb Cortex **18**(9): 2066-2076.

Andersen, R. A. and H. Cui (2009). "Intention, action planning, and decision making in parietal-frontal circuits." Neuron **63**(5): 568-583.

Aparicio, P. L., E. B. Issa and J. J. DiCarlo (2016). "Neurophysiological Organization of the Middle Face Patch in Macaque Inferior Temporal Cortex." J Neurosci **36**(50): 12729-12745.

Ashbridge, E., D. I. Perrett, M. W. Oram and T. Jellema (2000). "Effect of image orientation and size on object recognition: responses of single units in the macaque monkey temporal cortex." Cogn Neuropsychol **17**(1): 13-34.

Avillac, M., S. Ben Hamed and J. R. Duhamel (2007). "Multisensory integration in the ventral intraparietal area of the macaque monkey." J Neurosci **27**(8): 1922-1932.

Ayzenberg, V. and M. Behrmann (2022). "Does the brain's ventral visual pathway compute object shape?" Trends Cogn Sci **26**(12): 1119-1132.

Barraclough, N. E. and D. I. Perrett (2011). "From single cells to social perception." Philos Trans R Soc Lond B Biol Sci **366**(1571): 1739-1752.

Barraclough, N. E., D. Xiao, C. I. Baker, M. W. Oram and D. I. Perrett (2005). "Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions." J Cogn Neurosci **17**(3): 377-391.

Baylis, G. C., E. T. Rolls and C. M. Leonard (1987). "Functional subdivisions of the temporal lobe neocortex." J Neurosci **7**(2): 330-342.

Beauchamp, M. S., B. D. Argall, J. Bodurka, J. H. Duyn and A. Martin (2004). "Unraveling multisensory integration: patchy organization within human STS multisensory cortex." Nat Neurosci **7**(11): 1190-1192.

Beauchamp, M. S., K. E. Lee, B. D. Argall and A. Martin (2004). "Integration of auditory and visual information about objects in superior temporal sulcus." Neuron **41**(5): 809-823.

Bell, A. H., N. J. Malecek, E. L. Morin, F. Hadj-Bouziane, R. B. Tootell and L. G. Ungerleider (2011). "Relationship between functional magnetic resonance imaging-identified regions and neuronal category selectivity." J Neurosci **31**(34): 12229-12240.

Bell, A. H., C. Summerfield, E. L. Morin, N. J. Malecek and L. G. Ungerleider (2016). "Encoding of Stimulus Probability in Macaque Inferior Temporal Cortex." Curr Biol **26**(17): 2280-2290.

Benevento, L. A., J. Fallon, B. J. Davis and M. Rezak (1977). "Auditory--visual interaction in single cells in the cortex of the superior temporal sulcus and the orbital frontal cortex of the macaque monkey." Exp Neurol **57**(3): 849-872.
Bizley, J. K. and A. J. King (2008). "Visual-auditory spatial processing in auditory cortical neurons." Brain Res **1242**: 24-36.
Bizley, J. K. and A. J. King (2009). "Visual influences on ferret auditory cortex." Hear Res **258**(1-2): 55-63.
Bizley, J. K., R. K. Maddox and A. K. C. Lee (2016). "Defining Auditory-Visual Objects: Behavioral Tests and Physiological Mechanisms." Trends Neurosci **39**(2): 74-85.
Bizley, J. K., F. R. Nodal, V. M. Bajo, I. Nelken and A. J. King (2007). "Physiological and anatomical evidence for multisensory interactions in auditory cortex." Cereb Cortex **17**(9): 2172-2189.
Bizley, J. K., B. G. Shinn-Cunningham and A. K. Lee (2012). "Nothing is irrelevant in a noisy world: sensory illusions reveal obligatory within-and across-modality integration." J Neurosci **32**(39): 13402-13410.
Blank, H., A. Anwander and K. von Kriegstein (2011). "Direct structural connections between voice- and face-recognition areas." J Neurosci **31**(36): 12906-12915.
Bondar, I. V., D. A. Leopold, B. J. Richmond, J. D. Victor and N. K. Logothetis (2009). "Long-term stability of visual pattern selective responses of monkey temporal lobe neurons." PLoS One **4**(12): e8222.
Bruce, C., R. Desimone and C. G. Gross (1981). "Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque." J Neurophysiol **46**(2): 369-384.
Cantone, G., J. Xiao and J. B. Levitt (2006). "Retinotopic organization of ferret suprasylvian cortex." Vis Neurosci **23**(1): 61-77.
Cappe, C., E. M. Rouiller and P. Barone (2009). "Multisensory anatomical pathways." Hear Res **258**(1-2): 28-36.
Cappe, C., A. Thelen, V. Romei, G. Thut and M. M. Murray (2012). "Looming signals reveal synergistic principles of multisensory integration." J Neurosci **32**(4): 1171-1182.
Cappe, C., G. Thut, V. Romei and M. M. Murray (2009). "Selective integration of auditory-visual looming cues by humans." Neuropsychologia **47**(4): 1045-1052.
Caruso, V. C., D. S. Pages, M. A. Sommer and J. M. Groh (2021). "Compensating for a shifting world: evolving reference frames of visual and auditory signals across three multimodal brain areas." J Neurophysiol **126**(1): 82-94.
Castiello, U. (2005). "The neuroscience of grasping." Nat Rev Neurosci **6**(9): 726-736.
Chang, L. and D. Y. Tsao (2017). "The Code for Facial Identity in the Primate Brain." Cell **169**(6): 1013-1028 e1014.
Cherry, E. C. (1953). "Some Experiments on the Recognition of Speech, with One and with 2 Ears." Journal of the Acoustical Society of America **25**(5): 975-979.
Chong, I., H. Ramezanpour and P. Thier (2023). "Causal manipulation of gaze-following in the macaque temporal cortex." Prog Neurobiol **226**: 102466.
Colin, C., M. Radeau and P. Deltenre (2005). "Top-down and bottom-up modulation of audiovisual integration in speech." European Journal of Cognitive Psychology **17**(4): 541-560.
Conrad, V., M. Kleiner, A. Bartels, J. Hartcher O'Brien, H. H. Bulthoff and U. Noppeney (2013). "Naturalistic stimulus structure determines the integration of audiovisual looming signals in binocular rivalry." PLoS One **8**(8): e70710.

Conway, B. R. (2018). "The Organization and Operation of Inferior Temporal Cortex." Annu Rev Vis Sci **4**: 381-402.
Cox, R. W. (1996). "AFNI: software for analysis and visualization of functional magnetic resonance neuroimages." Comput Biomed Res **29**(3): 162-173.
Cuaya, L. V., R. Hernandez-Perez and L. Concha (2016). "Our Faces in the Dog's Brain: Functional Imaging Reveals Temporal Cortex Activation during Perception of Human Faces." PLoS One **11**(3): e0149431.
Dahl, C. D., N. K. Logothetis and C. Kayser (2009). "Spatial organization of multisensory responses in temporal association cortex." J Neurosci **29**(38): 11924-11932.
Dahl, C. D., N. K. Logothetis and C. Kayser (2010). "Modulation of visual responses in the superior temporal sulcus by audio-visual congruency." Front Integr Neurosci **4**: 10.
Delfour, F. and K. Marten (2001). "Mirror image processing in three marine mammal species: killer whales (Orcinus orca), false killer whales (Pseudorca crassidens) and California sea lions (Zalophus californianus)." Behav Processes **53**(3): 181-190.
Dell, L. A., G. M. Innocenti, C. C. Hilgetag and P. R. Manger (2019). "Cortical and thalamic connectivity of temporal visual cortical areas 20a and 20b of the domestic ferret (Mustela putorius furo)." J Comp Neurol **527**(8): 1333-1347.
Desimone, R., T. D. Albright, C. G. Gross and C. Bruce (1984). "Stimulus-Selective Properties of Inferior Temporal Neurons in the Macaque." Journal of Neuroscience **4**(8): 2051-2062.
Desimone, R., T. D. Albright, C. G. Gross and C. Bruce (1984). "Stimulus-selective properties of inferior temporal neurons in the macaque." J Neurosci **4**(8): 2051-2062.
Dewaal, F. B. M. and D. Yoshihara (1983). "Reconciliation and Redirected Affection in Rhesus-Monkeys." Behaviour **85**: 224-241.
DiCarlo, J. J. and D. D. Cox (2007). "Untangling invariant object recognition." Trends Cogn Sci **11**(8): 333-341.
DiCarlo, J. J., D. Zoccolan and N. C. Rust (2012). "How Does the Brain Solve Visual Object Recognition?" Neuron **73**(3): 415-434.
Diehl, M. M. and L. M. Romanski (2014). "Responses of prefrontal multisensory neurons to mismatching faces and vocalizations." J Neurosci **34**(34): 11233-11243.
Dittrich, W. (1990). "Representation of Faces in Longtailed Macaques (Macaca fascicularis)." Ethology **85**(4): 265-278.
Dobbins, A. C., R. M. Jeo, J. Fiser and J. M. Allman (1998). "Distance modulation of neural activity in the visual cortex." Science **281**(5376): 552-555.
Dunn-Weiss, E., S. U. Nummela, A. A. Lempel, J. M. Law, J. Ledley, P. Salvino and K. J. Nielsen (2019). "Visual Motion and Form Integration in the Behaving Ferret." eNeuro **6**(4).
Eastman, K. M. and A. C. Huk (2012). "PLDAPS: A Hardware Architecture and Software Toolbox for Neurophysiology Requiring Complex Visual Stimuli and Online Behavioral Control." Front Neuroinform **6**: 1.
Ebitz, R. B. and B. Y. Hayden (2021). "The population doctrine in cognitive neuroscience." Neuron **109**(19): 3055-3068.
Falchier, A., S. Clavagnier, P. Barone and H. Kennedy (2002). "Anatomical evidence of Multimodal integration in primate striate cortex." Journal of Neuroscience **22**(13): 5749-5759.
Fang, F., H. Boyaci, D. Kersten and S. O. Murray (2008). "Attention-dependent representation of a size illusion in human V1." Curr Biol **18**(21): 1707-1712.

Felleman, D. J. and D. C. Van Essen (1991). "Distributed Hierarchical Processing in the Primate Cerebral Cortex." Cerebral Cortex **1**(1): 1-47.
Fisher, C. and W. A. Freiwald (2015). "Contrasting specializations for facial motion within the macaque face-processing system." Curr Biol **25**(2): 261-266.
Fisher, C. and W. A. Freiwald (2015). "Whole-agent selectivity within the macaque face-processing system." Proc Natl Acad Sci U S A **112**(47): 14717-14722.
Fisher, P. G. (2006). "[Image: see text] FERRET BEHAVIOR." Exotic Pet Behavior: 163-205.
Fishman, M. C. and P. Michael (1973). "Integration of auditory information in the cat's visual cortex." Vision Res **13**(8): 1415-1419.
Flanagan, J. R., J. P. Bittner and R. S. Johansson (2008). "Experience can change distinct size-weight priors engaged in lifting objects and judging their weights." Curr Biol **18**(22): 1742-1747.
Fleming, J. T., A. L. Noyce and B. G. Shinn-Cunningham (2020). "Audio-visual spatial alignment improves integration in the presence of a competing audio-visual stimulus." Neuropsychologia **146**: 107530.
Freiwald, W. A. (2020). "The neural mechanisms of face processing: cells, areas, networks, and models." Curr Opin Neurobiol **60**: 184-191.
Freiwald, W. A. and D. Y. Tsao (2010). "Functional compartmentalization and viewpoint generalization within the macaque face-processing system." Science **330**(6005): 845-851.
Freud, E., T. Ganel, I. Shelef, M. D. Hammer, G. Avidan and M. Behrmann (2017). "Three-Dimensional Representations of Objects in Dorsal Cortex are Dissociable from Those in Ventral Cortex." Cereb Cortex **27**(1): 422-434.
Ghazanfar, A. A., C. Chandrasekaran and N. K. Logothetis (2008). "Interactions between the superior temporal sulcus and auditory cortex mediate dynamic face/voice integration in rhesus monkeys." J Neurosci **28**(17): 4457-4469.
Ghazanfar, A. A., J. X. Maier, K. L. Hoffman and N. K. Logothetis (2005). "Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex." J Neurosci **25**(20): 5004-5012.
Ghazanfar, A. A. and L. R. Santos (2004). "Primate brains in the wild: the sensory bases for social interactions." Nat Rev Neurosci **5**(8): 603-616.
Ghazanfar, A. A. and C. E. Schroeder (2006). "Is neocortex essentially multisensory?" Trends Cogn Sci **10**(6): 278-285.
Gielen, S. C., R. A. Schmidt and P. J. Van den Heuvel (1983). "On the nature of intersensory facilitation of reaction time." Percept Psychophys **34**(2): 161-168.
Gleiss, S. and C. Kayser (2013). "Eccentricity dependent auditory enhancement of visual stimulus detection but not discrimination." Front Integr Neurosci **7**: 52.
Goodale, M. A. and A. D. Milner (1992). "Separate visual pathways for perception and action." Trends Neurosci **15**(1): 20-25.
Gothard, K. M., F. P. Battaglia, C. A. Erickson, K. M. Spitler and D. G. Amaral (2007). "Neural responses to facial expression and face identity in the monkey amygdala." J Neurophysiol **97**(2): 1671-1683.
Gothard, K. M., K. N. Brooks and M. A. Peterson (2009). "Multiple perceptual strategies used by macaque monkeys for face recognition." Anim Cogn **12**(1): 155-167.

Gouzoules, S., H. Gouzoules and P. Marler (1984). "Rhesus monkey (Macaca mulatta) screams: representational signalling in the recruitment of agonistic aid." Animal Behaviour **32**(1): 182-193.
Grimaldi, P., K. S. Saleem and D. Tsao (2016). "Anatomical Connections of the Functionally Defined "Face Patches" in the Macaque Monkey." Neuron **90**(6): 1325-1342.
Groh, J. M., A. S. Trause, A. M. Underhill, K. R. Clark and S. Inati (2001). "Eye position influences auditory responses in primate inferior colliculus." Neuron **29**(2): 509-518.
Hackett, T. A., I. Stepniewska and J. H. Kaas (1998). "Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys." J Comp Neurol **394**(4): 475-495.
Harrod, E. G., C. L. Coe and P. M. Niedenthal (2020). "Social Structure Predicts Eye Contact Tolerance in Nonhuman Primates: Evidence from a Crowd-Sourcing Approach." Sci Rep **10**(1): 6971.
Hauser, M. D. (1991). "Sources of Acoustic Variation in Rhesus Macaque (Macaca mulatta) Vocalizations." Ethology **89**(1): 29-46.
Hauser, M. D. and P. Marler (1993). "Food-associated calls in rhesus macaques (Macaca mulatta): I. Socioecological factors." Behavioral Ecology **4**(3): 194-205.
Herreras, O. (2016). "Local Field Potentials: Myths and Misunderstandings." Front Neural Circuits **10**: 101.
Hesse, J. K. and D. Y. Tsao (2020). "The macaque face patch system: a turtle's underbelly for the brain." Nat Rev Neurosci **21**(12): 695-716.
Hikosaka, K., E. Iwai, H. Saito and K. Tanaka (1988). "Polysensory properties of neurons in the anterior bank of the caudal superior temporal sulcus of the macaque monkey." J Neurophysiol **60**(5): 1615-1637.
Hinkle, D. A. and C. E. Connor (2001). "Disparity tuning in macaque area V4." Neuroreport **12**(2): 365-369.
Hinkle, D. A. and C. E. Connor (2002). "Three-dimensional orientation tuning in macaque area V4." Nat Neurosci **5**(7): 665-670.
Homman-Ludiye, J., P. R. Manger and J. A. Bourne (2010). "Immunohistochemical parcellation of the ferret (Mustela putorius) visual cortex reveals substantial homology with the cat (Felis catus)." J Comp Neurol **518**(21): 4439-4462.
Hoy, J. L., I. Yavorska, M. Wehr and C. M. Niell (2016). "Vision Drives Accurate Approach Behavior during Prey Capture in Laboratory Mice." Curr Biol **26**(22): 3046-3052.
Hubel, D. H. and T. N. Wiesel (1959). "Receptive fields of single neurones in the cat's striate cortex." J Physiol **148**(3): 574-591.
Hubel, D. H. and T. N. Wiesel (1962). "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex." J Physiol **160**(1): 106-154.
Hubel, D. H. and T. N. Wiesel (1965). "Receptive Fields and Functional Archichitecture in 2 Nonstriate Visual Areas (18 and 19) of Cat." Journal of Neurophysiology **28**(2): 229-+.
Hung, C. C., C. C. Yen, J. L. Ciuchta, D. Papoti, N. A. Bock, D. A. Leopold and A. C. Silva (2015). "Functional mapping of face-selective regions in the extrastriate visual cortex of the marmoset." J Neurosci **35**(3): 1160-1172.
Hung, C. P., G. Kreiman, T. Poggio and J. J. DiCarlo (2005). "Fast readout of object identity from macaque inferior temporal cortex." Science **310**(5749): 863-866.

Hutchinson, E. B., S. C. Schwerin, K. L. Radomski, N. Sadeghi, J. Jenkins, M. E. Komlosh, M. O. Irfanoglu, S. L. Juliano and C. Pierpaoli (2017). "Population based MRI and DTI templates of the adult ferret brain and tools for voxelwise analysis." Neuroimage **152**: 575-589.
Ito, M., H. Tamura, I. Fujita and K. Tanaka (1995). "Size and position invariance of neuronal responses in monkey inferotemporal cortex." J Neurophysiol **73**(1): 218-226.
Janssen, P., B. E. Verhoef and E. Premereur (2018). "Functional interactions between the macaque dorsal and ventral visual pathways during three-dimensional object vision." Cortex **98**: 218-227.
Janssen, P., R. Vogels and G. A. Orban (1999). "Macaque inferior temporal neurons are selective for disparity-defined three-dimensional shapes." Proc Natl Acad Sci U S A **96**(14): 8217-8222.
Janssen, P., R. Vogels and G. A. Orban (2000). "Selectivity for 3D shape that reveals distinct areas within macaque inferior temporal cortex." Science **288**(5473): 2054-2056.
Janssen, P., R. Vogels and G. A. Orban (2000). "Three-dimensional shape coding in inferior temporal cortex." Neuron **27**(2): 385-397.
Jastorff, J., I. D. Popivanov, R. Vogels, W. Vanduffel and G. A. Orban (2012). "Integration of shape and motion cues in biological motion processing in the monkey STS." Neuroimage **60**(2): 911-921.
Jellema, T., C. I. Baker, B. Wicker and D. I. Perrett (2000). "Neural representation for the perception of the intentionality of actions." Brain Cogn **44**(2): 280-302.
Jellema, T., G. Maassen and D. I. Perrett (2004). "Single cell integration of animate form, motion and location in the superior temporal cortex of the macaque monkey." Cereb Cortex **14**(7): 781-790.
Jellema, T. and D. I. Perrett (2003). "Perceptual history influences neural responses to face and body postures." J Cogn Neurosci **15**(7): 961-971.
Kaas, J. H. (2011). "The Evolution of Auditory Cortex: The Core Areas." Auditory Cortex: 407-427.
Kaas, J. H. and T. A. Hackett (2000). "Subdivisions of auditory cortex and processing streams in primates." Proc Natl Acad Sci U S A **97**(22): 11793-11799.
Kaminski, J., J. Riedel, J. Call and M. Tomasello (2005). "Domestic goats,
, follow gaze direction and use social cues in an object choice task." Animal Behaviour **69**: 11-18.
Kanwisher, N., J. McDermott and M. M. Chun (1997). "The fusiform face area: a module in human extrastriate cortex specialized for face perception." J Neurosci **17**(11): 4302-4311.
Kastner, S. and L. G. Ungerleider (2000). "Mechanisms of visual attention in the human cortex." Annu Rev Neurosci **23**: 315-341.
Kayser, C., N. K. Logothetis and S. Panzeri (2010). "Visual enhancement of the information representation in auditory cortex." Curr Biol **20**(1): 19-24.
Kayser, C., C. I. Petkov, M. Augath and N. K. Logothetis (2007). "Functional imaging reveals visual modulation of specific fields in auditory cortex." J Neurosci **27**(8): 1824-1835.
Kendrick, K. M. (2006). "Brain asymmetries for face recognition and emotion control in sheep." Cortex **42**(1): 96-98.
Kendrick, K. M. and B. A. Baldwin (1987). "Cells in Temporal Cortex of Conscious Sheep Can Respond Preferentially to the Sight of Faces." Science **236**(4800): 448-450.

Khandhadia, A. P., A. P. Murphy, K. W. Koyano, E. M. Esch and D. A. Leopold (2023). "Encoding of 3D physical dimensions by face-selective cortical neurons." Proc Natl Acad Sci U S A **120**(9): e2214996120.
Khandhadia, A. P., A. P. Murphy, L. M. Romanski, J. K. Bizley and D. A. Leopold (2021). "Audiovisual integration in macaque face patch neurons." Curr Biol **31**(9): 1826-1835 e1823.
Konkle, T. and A. Oliva (2011). "Canonical visual size for real-world objects." J Exp Psychol Hum Percept Perform **37**(1): 23-37.
Konkle, T. and A. Oliva (2012). "A real-world size organization of object responses in occipitotemporal cortex." Neuron **74**(6): 1114-1124.
Kording, K. P., U. Beierholm, W. J. Ma, S. Quartz, J. B. Tenenbaum and L. Shams (2007). "Causal inference in multisensory perception." PLoS One **2**(9): e943.
Koyano, K. W., A. P. Jones, D. B. T. McMahon, E. N. Waidmann, B. E. Russ and D. A. Leopold (2021). "Dynamic Suppression of Average Facial Structure Shapes Neural Tuning in Three Macaque Face Patches." Curr Biol **31**(1): 1-12 e15.
Kravitz, D. J., K. S. Saleem, C. I. Baker, L. G. Ungerleider and M. Mishkin (2013). "The ventral visual pathway: an expanded neural framework for the processing of object quality." Trends Cogn Sci **17**(1): 26-49.
Lafer-Sousa, R. and B. R. Conway (2013). "Parallel, multi-stage processing of colors, faces and shapes in macaque inferior temporal cortex." Nat Neurosci **16**(12): 1870-1878.
Lakatos, P., G. Karmos, A. D. Mehta, I. Ulbert and C. E. Schroeder (2008). "Entrainment of neuronal oscillations as a mechanism of attentional selection." Science **320**(5872): 110-113.
Landemard, A., C. Bimbard, C. Demene, S. Shamma, S. Norman-Haignere and Y. Boubenec (2021). "Distinct higher-order representations of natural sounds in human and ferret auditory cortex." Elife **10**.
Landi, S. M. and W. A. Freiwald (2017). "Two areas for familiar face recognition in the primate brain." Science **357**(6351): 591-595.
Langbauer, W. R. (2000). "Elephant communication." Zoo Biology **19**(5): 425-445.
Larrat, S. and N. Summa (2021). "Ferret Behavior Medicine." Vet Clin North Am Exot Anim Pract **24**(1): 37-51.
Lee, J. and J. M. Groh (2012). "Auditory signals evolve from hybrid- to eye-centered coordinates in the primate superior colliculus." J Neurophysiol **108**(1): 227-242.
Lempel, A. A. and K. J. Nielsen (2019). "Ferrets as a Model for Higher-Level Visual Motion Processing." Curr Biol **29**(2): 179-191 e175.
Leopold, D. A., I. V. Bondar and M. A. Giese (2006). "Norm-based face encoding by single neurons in the monkey inferotemporal cortex." Nature **442**(7102): 572-575.
Leopold, D. A., J. F. Mitchell and W. A. Freiwald (2020). Chapter 25 - Evolved Mechanisms of High-Level Visual Perception in Primates. Evolutionary Neuroscience (Second Edition). J. H. Kaas. London, Academic Press**:** 589-625.
Leopold, D. A., A. J. O'Toole, T. Vetter and V. Blanz (2001). "Prototype-referenced shape encoding revealed by high-level aftereffects." Nat Neurosci **4**(1): 89-94.
Leopold, D. A. and S. H. Park (2020). "Studying the visual brain in its natural rhythm." Neuroimage **216**: 116790.
Leopold, D. A. and G. Rhodes (2010). "A comparative view of face perception." J Comp Psychol **124**(3): 233-251.

Lewis, J. W. and D. C. Van Essen (2000). "Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey." J Comp Neurol **428**(1): 112-137.
Liu, Y., R. Vogels and G. A. Orban (2004). "Convergence of depth from texture and depth from disparity in macaque inferior temporal cortex." J Neurosci **24**(15): 3795-3800.
MacDonald, J. and H. McGurk (1978). "Visual influences on speech perception processes." Percept Psychophys **24**(3): 253-257.
Maddox, R. K., H. Atilgan, J. K. Bizley and A. K. Lee (2015). "Auditory selective attention is enhanced by a task-irrelevant temporally coherent visual stimulus in human listeners." Elife **4**.
Maier, J. X., C. Chandrasekaran and A. A. Ghazanfar (2008). "Integration of bimodal looming signals through neuronal coherence in the temporal lobe." Curr Biol **18**(13): 963-968.
Maier, J. X., J. G. Neuhoff, N. K. Logothetis and A. A. Ghazanfar (2004). "Multisensory integration of looming signals by rhesus monkeys." Neuron **43**(2): 177-181.
Manger, P. R., D. Kiper, I. Masiello, L. Murillo, L. Tettoni, Z. Hunyadi and G. M. Innocenti (2002). "The representation of the visual field in three extrastriate areas of the ferret (Mustela putorius) and the relationship of retinotopy and field boundaries to callosal connectivity." Cereb Cortex **12**(4): 423-437.
Manger, P. R., H. Nakamura, S. Valentiniene and G. M. Innocenti (2004). "Visual areas in the lateral temporal cortex of the ferret (Mustela putorius)." Cereb Cortex **14**(6): 676-689.
Marino, L., R. C. Connor, R. E. Fordyce, L. M. Herman, P. R. Hof, L. Lefebvre, D. Lusseau, B. McCowan, E. A. Nimchinsky, A. A. Pack, L. Rendell, J. S. Reidenberg, D. Reiss, M. D. Uhen, E. Van der Gucht and H. Whitehead (2007). "Cetaceans have complex brains for complex cognition." PLoS Biol **5**(5): e139.
Mason, C. R., L. S. Theverapperuma, C. M. Hendrix and T. J. Ebner (2004). "Monkey Hand Postural Synergies During Reach-to-Grasp in the Absence of Vision of the Hand and Object." Journal of Neurophysiology **91**(6): 2826-2837.
McAlpine, D. (2005). "Creating a sense of auditory space." J Physiol **566**(Pt 1): 21-28.
McMahon, D. B., I. V. Bondar, O. A. Afuwape, D. C. Ide and D. A. Leopold (2014). "One month in the life of a neuron: longitudinal single-unit electrophysiology in the monkey visual system." J Neurophysiol **112**(7): 1748-1762.
McMahon, D. B., A. P. Jones, I. V. Bondar and D. A. Leopold (2014). "Face-selective neurons maintain consistent visual responses across months." Proc Natl Acad Sci U S A **111**(22): 8251-8256.
McMahon, D. B., B. E. Russ, H. D. Elnaiem, A. I. Kurnikova and D. A. Leopold (2015). "Single-unit activity during natural vision: diversity, consistency, and spatial sensitivity among AF face patch neurons." J Neurosci **35**(14): 5537-5548.
Meienbrock, A., M. J. Naumer, O. Doehrmann, W. Singer and L. Muckli (2007). "Retinotopic effects during spatial audio-visual integration." Neuropsychologia **45**(3): 531-539.
Meredith, M. A., B. L. Allman, L. P. Keniston and H. R. Clemo (2009). "Auditory influences on non-auditory cortices." Hear Res **258**(1-2): 64-71.
Mishkin, M. and L. G. Ungerleider (1982). "Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys." Behav Brain Res **6**(1): 57-77.
Mishkin, M., L. G. Ungerleider and K. A. Macko (1983). "Object Vision and Spatial Vision - 2 Cortical Pathways." Trends in Neurosciences **6**(10): 414-417.

Miyashita, Y. (2019). "Perirhinal circuits for memory processing." Nat Rev Neurosci **20**(10): 577-592.
Moeller, S., W. A. Freiwald and D. Y. Tsao (2008). "Patches with links: a unified system for processing faces in the macaque temporal lobe." Science **320**(5881): 1355-1359.
Molholm, S., W. Ritter, M. M. Murray, D. C. Javitt, C. E. Schroeder and J. J. Foxe (2002). "Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study." Brain Res Cogn Brain Res **14**(1): 115-128.
Morrell, F. (1972). "Visual system's view of acoustic space." Nature **238**(5358): 44-46.
Mullette-Gillman, O. A., Y. E. Cohen and J. M. Groh (2005). "Eye-centered, head-centered, and complex coding of visual and auditory targets in the intraparietal sulcus." J Neurophysiol **94**(4): 2331-2352.
Mullette-Gillman, O. A., Y. E. Cohen and J. M. Groh (2009). "Motor-related signals in the intraparietal cortex encode locations in a hybrid, rather than eye-centered reference frame." Cereb Cortex **19**(8): 1761-1775.
Murphy, A. P. and D. A. Leopold (2019). "A parameterized digital 3D model of the Rhesus macaque face for investigating the visual processing of social cues." J Neurosci Methods **324**: 108309.
Murphy, R., A. Faroni, J. Wong and A. Reid (2019). "Protocol for a phase I trial of a novel synthetic polymer nerve conduit 'Polynerve' in participants with sensory digital nerve injury (UMANC)." F1000Res **8**: 959.
Murray, S. O., H. Boyaci and D. Kersten (2006). "The representation of perceived angular size in human primary visual cortex." Nat Neurosci **9**(3): 429-434.
Mysore, S. G., R. Vogels, S. E. Raiguel, J. T. Todd and G. A. Orban (2010). "The selectivity of neurons in the macaque fundus of the superior temporal area for three-dimensional structure from motion." J Neurosci **30**(46): 15491-15508.
Niell, C. M. and M. P. Stryker (2010). "Modulation of visual responses by behavioral state in mouse visual cortex." Neuron **65**(4): 472-479.
Nityananda, V. and J. C. A. Read (2017). "Stereopsis in animals: evolution, function and mechanisms." J Exp Biol **220**(Pt 14): 2502-2512.
Noonan, M. P., J. Sallet, R. B. Mars, F. X. Neubert, J. X. O'Reilly, J. L. Andersson, A. S. Mitchell, A. H. Bell, K. L. Miller and M. F. Rushworth (2014). "A neural circuit covarying with social hierarchy in macaques." PLoS Biol **12**(9): e1001940.
Op De Beeck, H. and R. Vogels (2000). "Spatial sensitivity of macaque inferior temporal neurons." J Comp Neurol **426**(4): 505-518.
Ortiz-Rios, M., F. A. C. Azevedo, P. Kusmierek, D. Z. Balla, M. H. Munk, G. A. Keliris, N. K. Logothetis and J. P. Rauschecker (2017). "Widespread and Opponent fMRI Signals Represent Sound Location in Macaque Auditory Cortex." Neuron **93**(4): 971-983 e974.
Pandya, B. S. a. D. N. (1984). "Further Observations on Parieto-Temporal Connection."
Pandya, D. N. and B. Seltzer (1982). "Association areas of the cerebral cortex." Trends in Neurosciences **5**: 386-390.
Park, S. H., K. W. Koyano, B. E. Russ, E. N. Waidmann, D. B. T. McMahon and D. A. Leopold (2022). "Parallel functional subnetworks embedded in the macaque face patch system." Sci Adv **8**(10): eabm2054.

Park, S. H., B. E. Russ, D. B. T. McMahon, K. W. Koyano, R. A. Berman and D. A. Leopold (2017). "Functional Subpopulations of Neurons in a Macaque Face Patch Revealed by Single-Unit fMRI Mapping." Neuron **95**(4): 971-981 e975.
Parrish, A. E., S. F. Brosnan and M. J. Beran (2015). "Do you see what I see? A comparative investigation of the Delboeuf illusion in humans (Homo sapiens), rhesus monkeys (Macaca mulatta), and capuchin monkeys (Cebus apella)." J Exp Psychol Anim Learn Cogn **41**(4): 395-405.
Pasupathy, A. and C. E. Connor (2002). "Population coding of shape in area V4." Nat Neurosci **5**(12): 1332-1338.
Perrett, D. I., J. K. Hietanen, M. W. Oram and P. J. Benson (1992). "Organization and functions of cells responsive to faces in the temporal cortex." Philos Trans R Soc Lond B Biol Sci **335**(1273): 23-30.
Perrett, D. I., E. T. Rolls and W. Caan (1982). "Visual neurones responsive to faces in the monkey temporal cortex." Exp Brain Res **47**(3): 329-342.
Perrodin, C., C. Kayser, N. K. Logothetis and C. I. Petkov (2011). "Voice cells in the primate temporal lobe." Curr Biol **21**(16): 1408-1415.
Perrodin, C., C. Kayser, N. K. Logothetis and C. I. Petkov (2014). "Auditory and visual modulation of temporal lobe neurons in voice-sensitive and association cortices." J Neurosci **34**(7): 2524-2537.
Perrodin, C., C. Kayser, N. K. Logothetis and C. I. Petkov (2015). "Natural asynchronies in audiovisual communication signals regulate neuronal multisensory interactions in voice-sensitive cortex." Proc Natl Acad Sci U S A **112**(1): 273-278.
Petkov, C. I., C. Kayser, T. Steudel, K. Whittingstall, M. Augath and N. K. Logothetis (2008). "A voice region in the monkey brain." Nat Neurosci **11**(3): 367-374.
Pinsk, M. A., M. Arcaro, K. S. Weiner, J. F. Kalkus, S. J. Inati, C. G. Gross and S. Kastner (2009). "Neural representations of faces and body parts in macaque and human cortex: a comparative FMRI study." J Neurophysiol **101**(5): 2581-2600.
Pitcher, D., G. Ianni and L. G. Ungerleider (2019). "A functional dissociation of face-, body- and scene-selective brain areas based on their response to moving and static stimuli." Sci Rep **9**(1): 8242.
Pitcher, D. and L. G. Ungerleider (2021). "Evidence for a Third Visual Pathway Specialized for Social Perception." Trends Cogn Sci **25**(2): 100-110.
Plank, T., K. Rosengarth, W. Song, W. Ellermeier and M. W. Greenlee (2012). "Neural correlates of audio-visual object recognition: Effects of implicit spatial congruency." Human Brain Mapping **33**(4): 797-811.
Poole, K., C (2023). How does the brain extract acoustic patterns? A behavioural and neural study. PhD Doctoral, University College London.
Popivanov, I. D., J. Jastorff, W. Vanduffel and R. Vogels (2012). "Stimulus representations in body-selective regions of the macaque cortex assessed with event-related fMRI." Neuroimage **63**(2): 723-741.
Popivanov, I. D., J. Jastorff, W. Vanduffel and R. Vogels (2015). "Tolerance of macaque middle STS body patch neurons to shape-preserving stimulus transformations." J Cogn Neurosci **27**(5): 1001-1016.
Popivanov, I. D., P. G. Schyns and R. Vogels (2016). "Stimulus features coded by single neurons of a macaque body category selective patch." Proc Natl Acad Sci U S A **113**(17): E2450-2459.

Premereur, E. and P. Janssen (2020). "Effective Connectivity Reveals an Interconnected Inferotemporal Network for Three-Dimensional Structure Processing." J Neurosci **40**(44): 8501-8512.
Premereur, E., J. Taubert, P. Janssen, R. Vogels and W. Vanduffel (2016). "Effective Connectivity Reveals Largely Independent Parallel Networks of Face and Body Patches." Curr Biol **26**(24): 3269-3279.
Quiroga, R. Q., Z. Nadasdy and Y. Ben-Shaul (2004). "Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering." Neural Comput **16**(8): 1661-1687.
Raman, R., A. Bognar, G. G. Nejad, N. Taubert, M. Giese and R. Vogels (2023). "Bodies in motion: Unraveling the distinct roles of motion and shape in dynamic body responses in the temporal cortex." Cell Rep **42**(12): 113438.
Ratan Murty, N. A., S. Teng, D. Beeler, A. Mynick, A. Oliva and N. Kanwisher (2020). "Visual experience is not necessary for the development of face-selectivity in the lateral fusiform gyrus." Proc Natl Acad Sci U S A **117**(37): 23011-23020.
Rauschecker, J. P. and S. K. Scott (2009). "Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing." Nat Neurosci **12**(6): 718-724.
Rauschecker, J. P. and B. Tian (2000). "Mechanisms and streams for processing of "what" and "where" in auditory cortex." Proc Natl Acad Sci U S A **97**(22): 11800-11806.
Rockland, K. S. and H. Ojima (2003). "Multisensory convergence in calcarine visual areas in macaque monkey." Int J Psychophysiol **50**(1-2): 19-26.
Rohlfing, T., C. D. Kroenke, E. V. Sullivan, M. F. Dubach, D. M. Bowden, K. A. Grant and A. Pfefferbaum (2012). "The INIA19 template and NeuroMaps atlas for primate brain image parcellation and spatial normalization." Frontiers in Neuroinformatics **6**.
Rolls, E. T. and G. C. Baylis (1986). "Size and contrast have only small effects on the responses to faces of neurons in the cortex of the superior temporal sulcus of the monkey." Exp Brain Res **65**(1): 38-48.
Romanski, L. M., B. B. Averbeck and M. Diltz (2005). "Neural representation of vocalizations in the primate ventrolateral prefrontal cortex." J Neurophysiol **93**(2): 734-747.
Romanski, L. M., J. F. Bates and P. S. Goldman-Rakic (1999). "Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey." J Comp Neurol **403**(2): 141-157.
Romanski, L. M. and M. M. Diehl (2011). "Neurons responsive to face-view in the primate ventrolateral prefrontal cortex." Neuroscience **189**: 223-235.
Romanski, L. M. and J. Hwang (2012). "Timing of audiovisual inputs to the prefrontal cortex and multisensory integration." Neuroscience **214**: 36-48.
Romanski, L. M., B. Tian, J. Fritz, M. Mishkin, P. S. Goldman-Rakic and J. P. Rauschecker (1999). "Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex." Nat Neurosci **2**(12): 1131-1136.
Roy, A., S. V. Shepherd and M. L. Platt (2014). "Reversible inactivation of pSTS suppresses social gaze following in the macaque (Macaca mulatta)." Soc Cogn Affect Neurosci **9**(2): 209-217.
Russ, B. E., T. Kaneko, K. S. Saleem, R. A. Berman and D. A. Leopold (2016). "Distinct fMRI Responses to Self-Induced versus Stimulus Motion during Free Viewing in the Macaque." J Neurosci **36**(37): 9580-9589.
Russ, B. E. and D. A. Leopold (2015). "Functional MRI mapping of dynamic visual features during natural viewing in the macaque." Neuroimage **109**: 84-94.

Rust, N. C. and J. J. Dicarlo (2010). "Selectivity and tolerance ("invariance") both increase as visual information propagates from cortical area V4 to IT." J Neurosci **30**(39): 12978-12995.
Sary, G., R. Vogels and G. A. Orban (1993). "Cue-invariant shape selectivity of macaque inferior temporal neurons." Science **260**(5110): 995-997.
Saxena, S. and J. P. Cunningham (2019). "Towards the neural population doctrine." Curr Opin Neurobiol **55**: 103-111.
Schlack, A., S. J. Sterbing-D'Angelo, K. Hartung, K. P. Hoffmann and F. Bremmer (2005). "Multisensory space representations in the macaque ventral intraparietal area." J Neurosci **25**(18): 4616-4625.
Schroeder, C. E. and J. J. Foxe (2002). "The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex." Brain Res Cogn Brain Res **14**(1): 187-198.
Schroeder, C. E., P. Lakatos, Y. Kajikawa, S. Partan and A. Puce (2008). "Neuronal oscillations and visual amplification of speech." Trends Cogn Sci **12**(3): 106-113.
Schroeder, C. E., D. A. Wilson, T. Radman, H. Scharfman and P. Lakatos (2010). "Dynamics of Active Sensing and perceptual selection." Curr Opin Neurobiol **20**(2): 172-176.
Seltzer, B. and D. N. Pandya (1978). "Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey." Brain Res **149**(1): 1-24.
Seltzer, B. and D. N. Pandya (1984). "Further observations on parieto-temporal connections in the rhesus monkey." Exp Brain Res **55**(2): 301-312.
Seltzer, B. and D. N. Pandya (1984). "Further observations on parieto-temporal connections in the rhesus monkey." Experimental Brain Research **55**(2): 301-312.
Seltzer, B. and D. N. Pandya (1989). "Frontal lobe connections of the superior temporal sulcus in the rhesus monkey." J Comp Neurol **281**(1): 97-113.
Seltzer, B. and D. N. Pandya (1989). "Intrinsic connections and architectonics of the superior temporal sulcus in the rhesus monkey." J Comp Neurol **290**(4): 451-471.
Seltzer, B. and D. N. Pandya (1994). "Parietal, temporal, and occipita projections to cortex of the superior temporal sulcus in the rhesus monkey: A retrograde tracer study." Journal of Comparative Neurology **343**(3): 445-463.
Seltzer, B. and D. N. Pandya (1994). "Parietal, temporal, and occipital projections to cortex of the superior temporal sulcus in the rhesus monkey: a retrograde tracer study." J Comp Neurol **343**(3): 445-463.
Shams, L., Y. Kamitani and S. Shimojo (2000). "Illusions. What you see is what you hear." Nature **408**(6814): 788.
Sheinberg, D. L. and N. K. Logothetis (2001). "Noticing familiar objects in real world scenes: the role of temporal cortical neurons in natural vision." J Neurosci **21**(4): 1340-1350.
Spence, C. (2013). "Just how important is spatial coincidence to multisensory integration? Evaluating the spatial rule." Ann N Y Acad Sci **1296**: 31-49.
Srinath, R., A. Emonds, Q. Wang, A. A. Lempel, E. Dunn-Weiss, C. E. Connor and K. J. Nielsen (2021). "Early Emergence of Solid Shape Coding in Natural and Deep Network Vision." Curr Biol **31**(1): 51-65 e55.
Stein, B. E., M. A. Meredith and M. T. Wallace (1993). "The visually responsive neuron and beyond: multisensory integration in cat and monkey." Prog Brain Res **95**: 79-90.
Stein, B. E. and T. R. Stanford (2008). "Multisensory integration: current issues from the perspective of the single neuron." Nat Rev Neurosci **9**(4): 255-266.

Sugihara, T., M. D. Diltz, B. B. Averbeck and L. M. Romanski (2006). "Integration of auditory and visual communication information in the primate ventrolateral prefrontal cortex." J Neurosci **26**(43): 11138-11147.
Tate, A. J., H. Fischer, A. E. Leigh and K. M. Kendrick (2006). "Behavioural and neurophysiological evidence for face identity and face emotion processing in animals." Philosophical Transactions of the Royal Society B-Biological Sciences **361**(1476): 2155-2172.
Taubert, J., S. Japee, A. P. Murphy, C. T. Tardiff, E. A. Koele, S. Kumar, D. A. Leopold and L. G. Ungerleider (2020). "Parallel Processing of Facial Expression and Head Orientation in the Macaque Brain." J Neurosci **40**(42): 8119-8131.
Taubert, J., J. B. Ritchie, L. G. Ungerleider and C. I. Baker (2022). "One object, two networks? Assessing the relationship between the face and body-selective regions in the primate visual system." Brain Struct Funct **227**(4): 1423-1438.
Taubert, J., G. Van Belle, R. Vogels and B. Rossion (2018). "The impact of stimulus size and orientation on individual face coding in monkey face-selective cortex." Sci Rep **8**(1): 10339.
Tian, B., D. Reser, A. Durham, A. Kustov and J. P. Rauschecker (2001). "Functional specialization in rhesus monkey auditory cortex." Science **292**(5515): 290-293.
Town, S. M., W. O. Brimijoin and J. K. Bizley (2017). "Egocentric and allocentric representations in auditory cortex." PLoS Biol **15**(6): e2001878.
Tsao, D. Y., W. A. Freiwald, T. A. Knutsen, J. B. Mandeville and R. B. Tootell (2003). "Faces and objects in macaque cerebral cortex." Nat Neurosci **6**(9): 989-995.
Tsao, D. Y., W. A. Freiwald, R. B. Tootell and M. S. Livingstone (2006). "A cortical region consisting entirely of face-selective cells." Science **311**(5761): 670-674.
Tsao, D. Y., S. Moeller and W. A. Freiwald (2008). "Comparing face patch systems in macaques and humans." Proc Natl Acad Sci U S A **105**(49): 19514-19519.
Tsao, D. Y., N. Schweers, S. Moeller and W. A. Freiwald (2008). "Patches of face-selective cortex in the macaque frontal lobe." Nat Neurosci **11**(8): 877-879.
Ungerleider, L. G., T. W. Galkin, R. Desimone and R. Gattass (2008). "Cortical connections of area V4 in the macaque." Cerebral Cortex **18**(3): 477-499.
Updyke, B. V. (1986). "Retinotopic Organization within the Cat Posterior Suprasylvian Sulcus and Gyrus." Journal of Comparative Neurology **246**(2): 265-280.
van Atteveldt, N., M. M. Murray, G. Thut and C. E. Schroeder (2014). "Multisensory integration: flexible use of general operations." Neuron **81**(6): 1240-1253.
Van der Stoep, N., T. C. W. Nijboer and S. Van der Stigchel (2014). "Exogenous orienting of crossmodal attention in 3-D space: Support for a depth-aware crossmodal attentional system." Psychonomic Bulletin & Review **21**(3): 708-714.
van der Stoep, N., A. Serino, A. Farne, M. Di Luca and C. Spence (2016). "Depth: the Forgotten Dimension in Multisensory Research." Multisensory Research **29**(6-7): 493-524.
Van der Stoep, N., S. Van der Stigchel, T. C. Nijboer and M. J. Van der Smagt (2016). "Audiovisual integration in near and far space: effects of changes in distance and stimulus effectiveness." Exp Brain Res **234**(5): 1175-1188.
Van Dromme, I. C., E. Premereur, B. E. Verhoef, W. Vanduffel and P. Janssen (2016). "Posterior Parietal Cortex Drives Inferotemporal Activations During Three-Dimensional Object Vision." PLoS Biol **14**(4): e1002445.

Vaziri, S., E. T. Carlson, Z. Wang and C. E. Connor (2014). "A channel for 3D environmental shape in anterior inferotemporal cortex." Neuron **84**(1): 55-62.
Vaziri, S. and C. E. Connor (2016). "Representation of Gravity-Aligned Scene Structure in Ventral Pathway Visual Cortex." Curr Biol **26**(6): 766-774.
Veilleux, C. C. and E. C. Kirk (2014). "Visual acuity in mammals: effects of eye size and ecology." Brain Behav Evol **83**(1): 43-53.
Verhoef, B. E., K. S. Bohon and B. R. Conway (2015). "Functional architecture for disparity in macaque inferior temporal cortex and its relationship to the architecture for faces, color, scenes, and visual field." J Neurosci **35**(17): 6952-6968.
Verhoef, B. E., R. Vogels and P. Janssen (2012). "Inferotemporal cortex subserves three-dimensional structure categorization." Neuron **73**(1): 171-182.
von Kriegstein, K., A. Kleinschmidt, P. Sterzer and A. L. Giraud (2005). "Interaction of face and voice areas during speaker recognition." J Cogn Neurosci **17**(3): 367-376.
Waidmann, E. N., K. W. Koyano, J. J. Hong, B. E. Russ and D. A. Leopold (2022). "Local features drive identity responses in macaque anterior face patches." Nat Commun **13**(1): 5592.
Wallis, G. and E. T. Rolls (1997). "Invariant face and object recognition in the visual system." Prog Neurobiol **51**(2): 167-194.
Ward, M. K., M. S. Bolding, K. P. Schultz and P. D. Gamlin (2015). "Mapping the macaque superior temporal sulcus: functional delineation of vergence and version eye-movement-related activity." J Neurosci **35**(19): 7428-7442.
Webster, M. J., J. Bachevalier and L. G. Ungerleider (1994). "Connections of inferior temporal areas TEO and TE with parietal and frontal cortex in macaque monkeys." Cereb Cortex **4**(5): 470-483.
Webster, M. J., J. Bachevalier and L. G. Ungerleider (1994). "Connections of inferior temporal areas TEO and TE with parietal and frontal cortex in macaque monkeys." Cerebral cortex **4**(5): 470-483.
Werner-Reiss, U. and J. M. Groh (2008). "A rate code for sound azimuth in monkey auditory cortex: implications for human neuroimaging studies." J Neurosci **28**(14): 3747-3758.
Werner-Reiss, U., K. A. Kelly, A. S. Trause, A. M. Underhill and J. M. Groh (2003). "Eye position affects activity in primary auditory cortex of primates." Curr Biol **13**(7): 554-562.
Wheatstone, C. (1962). "On some remarkable and hitherto unobserved phenomena of binocular vision." Optom Wkly **53**: 2311-2315.
Yamane, Y., E. T. Carlson, K. C. Bowman, Z. Wang and C. E. Connor (2008). "A neural code for three-dimensional object shape in macaque inferotemporal cortex." Nat Neurosci **11**(11): 1352-1360.
Zeng, H., G. R. Fink and R. Weidner (2020). "Visual Size Processing in Early Visual Cortex Follows Lateral Occipital Cortex Involvement." J Neurosci **40**(22): 4410-4417.
Zhang, H., S. Japee, A. Stacy, M. Flessert and L. G. Ungerleider (2020). "Anterior superior temporal sulcus is specialized for non-rigid facial motion in both monkeys and humans." Neuroimage **218**: 116878.
Zhivago, K. A. and S. P. Arun (2016). "Selective IT neurons are selective along many dimensions." J Neurophysiol **115**(3): 1512-1520.
Zhu, L. L. and M. S. Beauchamp (2017). "Mouth and Voice: A Relationship between Visual and Auditory Preference in the Human Superior Temporal Sulcus." J Neurosci **37**(10): 2697-2708.

Zion Golumbic, E., G. B. Cogan, C. E. Schroeder and D. Poeppel (2013). "Visual input enhances selective speech envelope tracking in auditory cortex at a "cocktail party"." J Neurosci **33**(4): 1417-1426.
Zion Golumbic, E. M., N. Ding, S. Bickel, P. Lakatos, C. A. Schevon, G. M. McKhann, R. R. Goodman, R. Emerson, A. D. Mehta, J. Z. Simon, D. Poeppel and C. E. Schroeder (2013). "Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party"." Neuron **77**(5): 980-991.