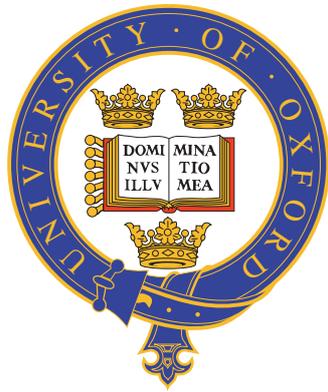


# Computational and physiological investigations of information processing in auditory cortex



**Aleksandar Ivanov**

Supervisors: Dr Kerry Walker,  
Dr Nicol Harper,  
Dr Benjamin Willmore,  
Prof Andrew King

Department of Physiology, Anatomy and Genetics  
St John's college, University of Oxford

This thesis is submitted for the degree of  
*Doctor of Philosophy*  
Hilary term, 2021



To my grandparents

# Computational and physiological investigations of information processing in auditory cortex

Aleksandar Ivanov, MSc

St John's college, University of Oxford

A thesis submitted for the degree of  
*Doctor of Philosophy*

Hilary Term, 2021

Auditory cortex has been shown to play a critical role in a plethora of processes, including the extraction of complex acoustic features, auditory scene analysis, decision-making, and learning. However, there are still many aspects of auditory cortex's structure and function that we do not understand. In this thesis, I explored three separate but related questions regarding cortical tuning properties, inhibition and their interplay in adaptation in the auditory cortex of the ferret. Firstly, I investigated the spatial distribution of preferred frequencies ("tonotopy") amongst cortical neurons in the ferret using 2-photon calcium imaging and electrophysiology, and compared it to the mouse. I found that cortical neurons in mice and ferrets have equivalent local heterogeneity in their tuning, while retaining a globally tonotopic map. Much of the observed heterogeneity in the tonotopic maps could be explained by the existence of neurons with complex receptive fields. Secondly, I developed electrophysiological and imaging approaches in order to study the properties of excitatory and inhibitory neurons, allowing me to label and record from excitatory and subclasses of inhibitory neurons. I did not find significant differences in the proportions of tuned neurons across neuron classes, or in their tuning characteristics. I also found that the co-tuning within and between classes are similar. Lastly, I explored the role of auditory cortex in adaptation to reverberation: What is the optimal transformation from reverberant to anechoic sounds in a statistical model, and can this model explain the robust representation of natural sounds in auditory cortex? The model recapitulated known properties of auditory cortical neurons, in addition to making novel predictions which I confirmed using electrophysiological recordings. These predictions were: (1) the inhibitory component of neuronal receptive fields shifts in time in proportion to the amount of reverberation, and (2) the degree of this inhibitory shift is frequency dependent. Taken together, these findings shed new light onto the organization and function of auditory cortex, providing exciting avenues for future research.

## Acknowledgements

This thesis would not have been possible without the help and support of people, who are too numerous to list. However, I will do my best.

First and foremost, I would like to express my eternal gratitude to my four supervisors Andrew King, Kerry Walker, Ben Willmore and Nicol Harper.

Andy, thank you for taking me into your lab and giving me the freedom and space to grow as a scientist, intellectually, professionally and personally. Your mentorship has been invaluable to me. I have learned a lot from you, and I have always admired your scientific rigour, scholarship and leadership skills.

Kerry, I cannot thank you enough for your patience, support and mentorship. You have always been there for me, even when it has not been easy. You have taught me so much. From what it means to be a good scientist to how to handle adversity and setbacks, staying always positive and humble. You have been a real inspiration and I could not have wished for a better supervisor.

Ben, I am really grateful for all your technical help and the many discussions we have had. I would always remember my very first lecture in Computational Neuroscience, where you showed me how receptive fields can be derived from white noise, a sort of cosmos arising from the chaos. Your critical outlook has been invaluable.

Nicol, thank you for all your patience and dedication. I have learned so much from you and our interactions have allowed me to appreciate the beauty and power of Computational Neuroscience. Your rigour, attention to detail, and insight into the workings of the brain have been truly inspiring.

I would also like to thank all the other members of the auditory neuroscience group (ANG). Each one of them has contributed to my development as a scientist and person in various different ways. I would never forget the crazy ferret experiments and my partners in crime, Quentin and Mari. The many discussions and the guidance of the

more experienced DPhil students in the lab, especially Yossi, Michael and Yves. My brother from another mother Samuel, who made the whole DPhil journey much more pleasant and interesting. I am especially grateful to the two founding members of “Tea Club”, Monzi and Tai. The two of you have been there for me when I really needed it, more times than I can count. I hope we can keep our wonderful and nourishing friendship forever.

Outside the lab, Oxford has been very generous to me in terms of friends and acquaintances too. I had a lot of fun diving deeper into the world of Math, establishing societies, organising concerts and many more with Plamen. I have many happy memories with the “Eastern block”, Paula, Dima and Olga, from trekking mountains to stealing Gubs. The honourable mentions would not be complete without my MSc buddies, Alex, Freddy, Evie, Seb and Maria. I can also not forget mi hermano mexicano, Adrian. Special thanks also to Tedi, who has been an important friend for me.

Back home in sunny Bulgaria, there are an equal number of people important to me. My interactions with the members of the “NPMG Alumni Club”, especially Lyu, Bobi, Iva, Maria, Mimo, Kalo, Nasko, Vicky have made my stays there, while taking breaks from my DPhil, immeasurably better. I am also very grateful for the friendship of my childhood buddies Slavi and Vesko. Special thanks also goes to the evil Zlo.

Like any good Academy Awards speech, I reach the point where I thank my relatives. I would not be where I am today without the unwavering support of my parents. My father, who has always been a shining example of what it means to be a good man, husband, father, friend, and hard worker. My mom, whose warm embrace I have always felt, even in the darkest of hours. My brother, who has raised me and shaped as much as my parents. Last but not least, my wonderful grandparents.

Finally, I would like to thank Nadya. Your presence in my life has been incredibly important, from getting me interested in Neuroscience in the first place to your unconditional love and support. I could not have completed my DPhil without you. Your positive attitude, resolve, resourcefulness, passion and kindness have carried me through.

## Scientific Acknowledgements

The 2-photon imaging data from ferrets in Chapter 2 was collected and analysed together with Quentin Gaucher, Mariangela Panniello and Kerry Walker. In addition, Mariangela Panniello collected and analysed the imaging data from mice. Mariangela Panniello and Kerry Walker designed the study. The work presented in Chapter 2 has been published in eLife (Gaucher et al., 2020). Jasper Poort provided code and guidance for the image registration analysis in Chapter 3. I would also like to thank Dominic Gray and Miranda de Saint-Rome who helped with spike sorting data, and contributed to the electrophysiology analysis pipeline for separating excitatory and inhibitory neurons in Chapter 3. The electrophysiology data in Chapters 2, 3 and 4 was collected together with Quentin Gaucher, Kerry Walker, Nicol Harper, Benjamin Willmore and Monzilur Rahman. The “Neuropixels” electrophysiological setup was established together with Benjamin Willmore, Sundeep Teki and Florencia Iacaruso. I am also grateful to Matteo Carandini, Adam Kampff, Nicholas Steinmetz, Marius Pachitariu and the other organisers of the 2017 Neuropixels course where I received technical training. Nicol Harper designed the work presented in Chapter 4. Thanks also to the MICRON imaging facility (supported by Wellcome Strategic Awards 091911/B/10/Z and 107457/Z/15/Z) for technical advice and access to equipment. Greg Daubney provided help with histological work. I am also very grateful to the members of the DPAG mechanical workshop Daniel Heaton and Marcos O’Bardain. I am also grateful to Andrew King and Kerry Walker for providing funding for the work from the Wellcome Trust and BBSRC, in addition to the Christopher Welch Scholarship in Biological Sciences from the University of Oxford. I would also like to thank Deyan Mihaylov for advice on typesetting.



# List of Abbreviations

**5-HT3A** 5-Hydroxytryptamine 3A receptor neurons

**A1** Primary Auditory Cortex

**AAF** Anterior Auditory Field

**AC** Auditory Cortex

**ADF** Anterior Dorsal Field

**AEG** Anterior Ectosylvian Gyrus

**AM** Amplitude Modulation

**ANOVA** Analysis of variance

**AVCN** Anteroventral Cochlear Nucleus

**AVF** Anterior Ventral Field

**BF** Best Frequency

**BRIR** Binaural Room Impulse Response

**CCG** Crosscorrelogram

**CF** Characteristic Frequency

**CNIC** Central Nucleus of the Inferior Colliculus

**CN** Cochlear Nucleus

**COM** Center of Mass

- DCIC** Dorsal Cortex of the Inferior Colliculus
- DRR** Direct-sound to Reverberant-sound Ratio
- EEG** Electroencephalogram
- EG** Ectosylvian Gyrus
- EXC** Excitatory neurons
- FRA** Frequency Response Area
- GAD** Gamma-Aminobutyric acid neurons
- GCAMP** Genetically encoded calcium indicator
- GLM** Generalised Linear Model
- HRTF** Head-Related Transfer Function
- IC** Inferior Colliculus
- ILD** Interaural Level Difference
- INH** Inhibitory neurons
- IR** Impulse Response
- ITD** Interaural Time Difference
- LCIC** Lateral Cortex of the Inferior Colliculus
- LFP** Local Field Potential
- LGN** Lateral Geniculate Nucleus
- LSO** Lateral Superior Olive
- MEG** Middle Ectosylvian Gyrus
- MGB** Medial Geniculate Body
- MSE** Mean Squared Error
- MSO** Medial Superior Olive

- 
- NC** Noise Correlation
- NLL** Nucleus of the Lateral Lemniscus
- PBS** Phosphate-Buffered Saline
- PEG** Posterior Ectosylvian Gyrus
- PH** Peak Height
- PPF** Posterior Pseudosylvian Field
- PSF** Posterior Suprasylvian Field
- PSTH** Peristimulus time histogram
- PT** Peak Time
- PVCN** Posteroventral Cochlear Nucleus
- PV** Parvalbumin neurons
- SC** Signal Correlation
- SEM** Standard Error of the Mean
- SOC** Superior Olivary Complex
- SOM** Somatostatin neurons
- SPL** Sound Pressure Level
- STRF** Spectro-Temporal Receptive Field
- V1** Primary Visual Cortex
- VAS** Virtual Acoustic Space
- VIP** Vasoactive Intestinal Peptide neurons
- ICSD** inverse Current Source Density



# Table of contents

List of figures	xix
List of tables	xxiii
<b>1 General Introduction</b>	<b>1</b>
1.1 What is the role of auditory cortex? . . . . .	1
1.2 Subcortical auditory processing . . . . .	4
1.2.1 Cochlea and Auditory Nerve . . . . .	4
1.2.2 Cochlear Nucleus . . . . .	9
1.2.3 Superior Olivary Complex . . . . .	9
1.2.4 Nucleus of the Lateral Lemniscus . . . . .	10
1.2.5 Inferior Colliculus . . . . .	11
1.2.6 Medial Geniculate Body . . . . .	12
1.3 Cortical auditory processing . . . . .	13
1.3.1 Auditory cortex: Bird's-eye view . . . . .	13
1.3.2 Ferret auditory cortex . . . . .	20
1.4 Cortical tonotopic maps . . . . .	23
1.5 Cortical inhibitory neurons . . . . .	24
1.6 Reverberation . . . . .	27
1.6.1 Reverberation fundamentals . . . . .	27

---

1.6.2	Impact of reverberation . . . . .	30
1.7	Thesis overview and structure . . . . .	31
<b>2</b>	<b>Tonotopic organisation of auditory cortex</b>	<b>35</b>
2.1	Introduction . . . . .	35
2.1.1	Global tonotopy but local heterogeneity . . . . .	35
2.1.2	Species and methodological considerations . . . . .	36
2.1.3	Summary of the study . . . . .	37
2.2	Materials and methods . . . . .	38
2.2.1	Animals . . . . .	38
2.2.2	Viral vector injections . . . . .	38
2.2.3	In vivo two-photon calcium imaging . . . . .	40
2.2.4	Sound presentation . . . . .	41
2.2.5	Histology . . . . .	43
2.2.6	Confocal imaging . . . . .	43
2.2.7	Identifying the location of GCaMP6 injection sites . . . . .	43
2.2.8	Data analysis for two-photon imaging experiments . . . . .	44
2.2.9	In vivo electrophysiology . . . . .	48
2.3	Results . . . . .	51
2.3.1	Frequency response organization at the local scale . . . . .	52
2.3.2	Frequency response organization at the global scale . . . . .	58
2.3.3	Comparison of tonotopic organization in ferrets and mice . . . . .	62
2.4	Discussion . . . . .	66
2.4.1	Primary auditory cortex is tonotopically organized at the global scale . . . . .	66
2.4.2	Tonotopy is well preserved locally for neurons with simple, but not complex, frequency receptive fields . . . . .	67
2.4.3	Similarities and differences between mice and ferrets . . . . .	68

---

2.4.4	Comparison to previous electrophysiological studies . . . . .	70
2.4.5	Conclusions . . . . .	71
2.5	Supplementary figures . . . . .	72
<b>3</b>	<b>Functional properties of inhibitory neurons in auditory cortex</b>	<b>87</b>
3.1	Introduction . . . . .	87
3.2	Materials and methods . . . . .	90
3.2.1	Animals . . . . .	90
3.2.2	Viral vector injections . . . . .	91
3.2.3	In vivo two-photon calcium imaging . . . . .	91
3.2.4	Sound presentation . . . . .	91
3.2.5	Histology . . . . .	91
3.2.6	Confocal imaging . . . . .	92
3.2.7	Image registration . . . . .	93
3.2.8	Analysis of two-photon imaging data . . . . .	96
3.2.9	In vivo electrophysiology experiments . . . . .	97
3.2.10	Analysis of electrophysiology data . . . . .	97
3.3	Results . . . . .	100
3.3.1	Separating putative excitatory and inhibitory neurons based on extracellular spike waveforms . . . . .	100
3.3.2	Separating excitatory and subclasses of inhibitory neurons using imaging approaches . . . . .	108
3.3.3	Response properties of excitatory and inhibitory neurons . . . . .	116
3.4	Discussion . . . . .	128
3.4.1	High density microelectrode recordings allow us to separate puta- tive excitatory and inhibitory neurons based on spike waveform features . . . . .	129

---

3.4.2	Targeted viral delivery, calcium imaging, immunohistochemistry and registration techniques discriminate between excitatory and subclasses of inhibitory neurons . . . . .	131
3.4.3	The tuning properties of excitatory and inhibitory neurons in ferret auditory cortex are remarkably similar . . . . .	132
3.4.4	Signal and noise correlations in excitatory and subclasses of inhibitory neurons reveal distinct interactions . . . . .	134
<b>4</b>	<b>Adaptation to reverberation in auditory cortex</b>	<b>139</b>
4.1	Introduction . . . . .	139
4.1.1	Adaptation and robustness to reverberation in the auditory system	139
4.1.2	Summary of the study . . . . .	141
4.2	Materials and methods . . . . .	143
4.2.1	Animals and surgical procedures . . . . .	143
4.2.2	Sound presentation . . . . .	144
4.2.3	Spike sorting and data analysis . . . . .	144
4.2.4	Virtual acoustic space . . . . .	145
4.2.5	Sound stimuli . . . . .	146
4.2.6	Cochlear model . . . . .	147
4.2.7	Model kernels . . . . .	147
4.2.8	Neuronal STRFs . . . . .	148
4.2.9	Quantification of model kernels and STRFs . . . . .	149
4.2.10	Simulated model neuron . . . . .	151
4.3	Results . . . . .	152
4.3.1	Normative dereverberation model . . . . .	152
4.3.2	Comparison of model kernels to neuronal receptive fields . . . . .	154
4.3.3	Model kernels capture the temporal adaptive properties of auditory cortical neurons . . . . .	155

---

4.3.4	The observed changes are consistent with a truly adaptive process	161
4.3.5	Frequency dependence of the temporal profile of adaptation . . .	164
4.3.6	Exploring the role of putative excitatory and inhibitory neurons in the adaptation to reverberation . . . . .	166
4.4	Discussion . . . . .	171
4.4.1	Auditory cortical neurons adapt their responses to reverberation	172
4.4.2	Largely inhibitory phenomena underlie the adaptation to reverberation . . . . .	172
4.4.3	Adaptation to reverberation is frequency dependent . . . . .	174
4.4.4	Conclusion . . . . .	174
<b>5</b>	<b>General Discussion</b>	<b>177</b>
5.1	Summary of findings . . . . .	177
5.2	Cortical tonotopic maps: disorder without rhyme or reason? . . . . .	179
5.3	Cortical excitatory-inhibitory interactions and their relationship to receptive field properties . . . . .	183
5.4	Possible biological implementations of the adaptation to reverberation .	185
5.5	Subcortical and cortical contributions to the adaptation to reverberation	187
5.6	Possible confounds from the use of anaesthesia . . . . .	187
	<b>Bibliography</b>	<b>191</b>



# List of figures

1.1	The Auditory Pathway . . . . .	5
1.2	The Cochlea . . . . .	7
1.3	Auditory cortex properties . . . . .	14
1.4	Auditory cortex of the ferret . . . . .	20
1.5	Cortical excitatory-inhibitory circuits . . . . .	25
1.6	The basics of reverberation . . . . .	29
2.1	Selecting cells for downstream analysis . . . . .	45
2.2	Example well isolate single unit and poorly isolated multi unit . . . . .	50
2.3	Imaging neuronal responses to pure tones in ferret A1 . . . . .	53
2.4	Example FRA shapes . . . . .	54
2.5	Complexity of frequency receptive fields predicts local variability in ferret A1 tonotopy . . . . .	55
2.6	Correlations in neural activity are weaker and less spatially ordered for neurons with more complex FRAs . . . . .	59
2.7	Global tonotopic organization of frequency preferences in ferret A1 . . . . .	60
2.8	Similar tonotopic organization in the primary auditory cortex of ferrets and mice . . . . .	63
2.9	GCaMP6 expression in ferret auditory cortex . . . . .	72
2.10	Imaging in layer 2/3 of ferret A1 . . . . .	73
2.11	Spatial maps of best frequency in A1 of 4 ferrets . . . . .	74

---

2.12	BF variance, response strength and response reliability are comparable in GCaMP6m and GCaMP6f ferret datasets . . . . .	75
2.13	Determining cortical layers from an iCSD map of electrophysiological responses . . . . .	76
2.14	Firing rates per FRA type from Electrophysiology . . . . .	76
2.15	Noise correlations in neurons with single-peaked, double-peaked and complex FRAs . . . . .	77
2.16	Mapping data from individual ferrets onto a common template of auditory cortex . . . . .	78
2.17	Mapping the peak-to-peak distances in double-peaked neurons . . . . .	79
2.18	Comparison between deconvolved and non-deconvolved traces . . . . .	80
2.19	Effects of neuropil contamination on local and global tonotopic organization . . . . .	81
2.20	All 3 FRA types are present in the mouse primary auditory cortex . . . . .	82
2.21	FRA bandwidth does not predict local tuning heterogeneity . . . . .	83
2.22	The BF variance within an imaging field does not depend on the number of neurons in the field . . . . .	84
2.23	Effects of FRA complexity on BF organization hold across all imaged neurons . . . . .	85
3.1	Types of affine transformations . . . . .	94
3.2	Separating excitatory and inhibitory neurons using extracellular spike features . . . . .	99
3.3	Distributions of spike features across neurons . . . . .	102
3.4	K-means clustering provides good separation of 2 neuronal groups based on their spike features . . . . .	104
3.5	High degree of overlap in the clustering of neurons . . . . .	105
3.6	Functional confirmation of the spike feature clustering . . . . .	107
3.7	Overlap in expression between GAD and mDlx-GCaMP . . . . .	109
3.8	General imaging workflow . . . . .	110

---

3.9	Optimising antibody staining for different inhibitory subclasses . . . . .	112
3.10	Registration of confocal and in vivo 2-photon images . . . . .	114
3.11	Viral expression of GCaMP6f via mDlx enhancer yields specific labeling of GABAergic neurons in ferret auditory cortex . . . . .	115
3.12	Similar proportions of excitatory and inhibitory neurons are responsive to tones . . . . .	117
3.13	The tuning properties of excitatory and inhibitory neurons are similar .	120
3.14	The best frequency preferences within an imaging field are similar for excitatory and inhibitory neurons. . . . .	122
3.15	Electrophysiology and calcium imaging reveal similar complexity of frequency receptive fields between excitatory and inhibitory neurons . .	123
3.16	Correlations in neural activity from electrophysiology and calcium imag- ing data reveal higher connectivity between certain classes of neurons .	125
3.17	Signal and noise correlations in excitatory and subclasses of inhibitory neurons reveal distinct interactions . . . . .	127
4.1	Dereverberation normative model . . . . .	153
4.2	Comparison of normative model kernels and neuronal STRFs . . . . .	156
4.3	Frequency breakdown of the normative model kernels and neuronal STRFs	157
4.4	Frequency breakdown of the normative model and neuronal temporal profiles . . . . .	158
4.5	Quantification of the temporal profiles of the normative model kernels and neuronal STRFs . . . . .	160
4.6	Simulated model neurons confirm the adaptive nature of the real neuronal responses to reverberation . . . . .	162
4.7	The excitatory and inhibitory COM profiles are frequency dependent and scale with the amount of reverberation . . . . .	165
4.8	Excitatory and inhibitory neuronal STRFs . . . . .	168
4.9	Excitatory and inhibitory neuronal temporal profiles . . . . .	169

- 5.1 Is the locally heterogeneous structure of auditory cortex a consequence of its input? . . . . . 180
- 5.2 Modelling the interactions between connectivity and receptive fields . . 184

# List of tables

2.1	Details of stimuli presented to ferrets . . . . .	42
3.1	Key histological resources . . . . .	92
3.2	Spike features used for clustering . . . . .	98
3.3	Comparison of responsivity and selectivity of 4 neuron subtypes in calcium imaging data . . . . .	118



# Chapter 1

## General Introduction

### 1.1 What is the role of auditory cortex?



What does the Auditory Cortex (AC) do<sup>1</sup>? More generally, how do the structure of auditory cortex, and the information flow and processing happening within it determine our perception? What is the more general purpose of sensory cortices? These are the kind of questions that motivated me going into my DPhil.

The idea that cerebral cortex might have a more general function, or at least that the operations that are performed in the different cortical areas might have

---

<sup>1</sup>This picture was taken from the [website](#) of Prof Adam Kampff's lab. Originally based on an [artwork](#) made by the artist under the moniker of "Banksy".

something in common, is implicit in a great deal of theoretical and experimental brain literature. Historically, to [F. J. Gall \(1815\)](#) the cortex housed the faculties listed on his phrenological maps, e.g., mirthfulness, acquisitiveness, amativeness, etc. To [Flourens \(1842\)](#) the cortex was the interface with the mind and so must have mental attributes, such as intelligence and will. According to Pavlov and Loeb the cortex was the organ of associative learning ([Pavlov, 1927](#); [Loeb, 1900](#)). From the point of view of [Goltz \(1892\)](#) the cortex was the seat of understanding, reasoning power, and intelligence. [Herrick \(1926\)](#) expressed a very similar sentiment, but stated his position if not as clearly at least more prosaically. For [Lashley \(1950\)](#) the cortex was the organ of learning and memory. Other more modern views such as that of [Oakley \(1981\)](#) depicts cortex as the organ of representational and abstract memory, while [MacLean \(1982\)](#) sees it as the organ of emotion and expression. Amusingly, [Phillips et al. \(1984\)](#) speculate that the cortex is a Sherlock Holmes of sorts, sensitive to “conspicuous coincidence”. A general theme of all these theories is that they more or less substitute the neocortex for a facsimile of the mind as described by [Descartes \(1664\)](#). Probably all of these views have some truth in them, and we cannot state succinctly and explicitly a single purpose or all encompassing theory that describes sensory cortices and cortex more generally. Despite this, there have been some more modern attempts to generate a general theory for the functioning and roles of sensory cortices, including AC. Attneave and Barlow suggested that the goal of the early sensory neurons is to represent natural inputs efficiently by eliminating statistical redundancy ([Attneave, 1954](#); [Barlow, 2012](#)). This idea has formed the theoretical basis for several frameworks including “predictive coding” and “sparse coding”.

Probably one of the most prominent current theories is the predictive coding framework ([Rao and Ballard, 1999](#)). According to this framework, biological brains are constantly making predictions about the incoming sensory data from the outside world. These predictions are made using an internal “generative model” of the outside world, and these predictions are fed from higher brain regions to earlier brain regions closer to the sensory periphery where they are compared against the actual incoming sensory data. The disparity in these comparisons generates the so called “prediction error”, which highlights salient and surprising features in the outside world, presumably important for the animal. Such prediction errors lead to an improvement in the sensory data estimates (recognition), and over a longer period of time, to a modified generative model (perceptual learning) ([Spratling, 2017](#)). Naturally, predictive coding has been applied to AC in order to explain its function and role in perception (for a detailed

review see [Carbajal and Malmierca \(2018\)](#) and [Heilbron and Chait \(2018\)](#)). There have been studies which successfully found neural correlates in AC of some of the key variables of predictive coding models, such as surprise, prediction error, prediction change, and prediction precision ([Kumar et al., 2011](#); [Sedley et al., 2016](#)). However, the fundamental principles of the theory are in the best case scenario not verified yet ([Egner and Summerfield, 2013](#)) or, in the worst case scenario, they cannot be tested even in principle ([Kogo and Trengove, 2015](#)).

The second theory we mentioned, that of sparse coding, also derives from the efficient coding framework. Originally developed to explain the emergence of visual receptive fields, it posits that early sensory areas are optimized to accurately represent current or past inputs using a sparse, distributed code. Within the sparse coding framework, all neurons participating in the coding schema have an equal probability of being active over the full class of natural stimuli, but a relatively low probability of being active for any specific stimulus ([Field, 1994](#); [Olshausen and Field, 1996](#)). Here, especially when an overcomplete code is used (that is, when there are more neurons than input dimensions), it is efficient despite not having a reduction in the dimensionality of the code because only a few of those neurons are active to represent a given stimulus. The sparse coding framework has been put forward as a potential organising principle of AC ([Willmore and King, 2009](#)) and there has been some experimental data to support it ([Hromádka et al., 2008](#); [Zhang et al., 2019](#)). However, similar to predictive coding, the evidence is far from conclusive ([Dodds and DeWeese, 2019](#)).

Although such big questions are very tempting for the naïve graduate student, one must ground oneself into more concrete ventures in order to achieve tangible progress, hence in this thesis I have focused on three specific aspects of auditory cortical information processing - tonotopy, inhibition, and adaptation to reverberation. There has been theoretical work that links the ideas from sparse coding to the organisation of the AC, suggesting that the sparse coding framework might provide a possible explanation for the topographic distribution of frequency selectivity of auditory cortical neurons over the span of the auditory cortex, i.e. their “tonotopy” ([Terashima and Okada, 2012](#)). In addition, local inhibitory neuronal circuits have been proposed as key players in the implementation of predictive coding in sensory circuits, including the AC ([Hertäg and Sprekeler, 2020](#)). Finally, approaches utilising ideas from both predictive and sparse coding have been applied to clean speech from the corruptive influences of the non-stationary complex noise phenomenon of reverberation ([Gaubitch et al., 2010](#);

A. Jukić et al., 2014). During my DPhil, I explored these three different but related aspects of AC organisation and function, with the hope that the detailed concrete investigation of three specific questions may contribute, in time, to the understanding of the larger scale picture of auditory cortical information processing. The three specific questions that I explored were: 1) How is the ferret AC organised with regards to its representation of auditory frequency space (“tonotopy”), how does this compare to the mouse and how can we reconcile some of the previous methodological discrepancies in the literature? 2) What are the tuning properties of different classes of inhibitory neurons within ferret AC, how do they compare to those of excitatory neurons and how do these classes interact with one another? 3) How does the ferret AC achieve invariant representation of natural sounds in real world listening conditions, where sounds are masked by reverberation?

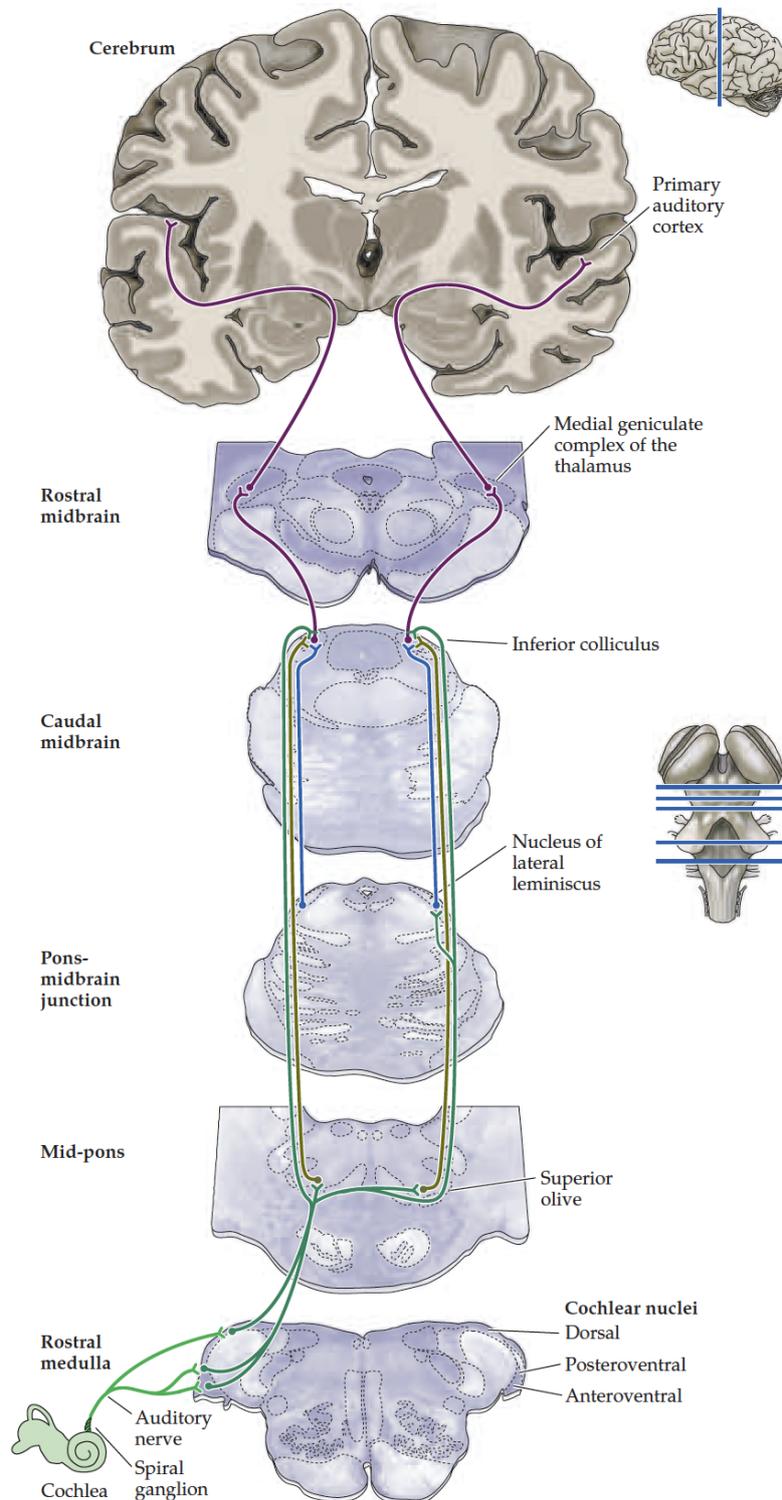
## 1.2 Subcortical auditory processing

Before proceeding to explore in more detail the current knowledge in the literature on AC, we will do a quick whistle-stop tour of the subcortical auditory pathways leading to AC (Figure 1.1). It is important to have a good appreciation of the subcortical auditory processing because it shapes to a large extent the organisation and computations within AC itself. Thus, we begin our tour at the cochlea.

### 1.2.1 Cochlea and Auditory Nerve

The cochlea is the first place where the auditory system interacts with the outside world. From here, each sound makes its journey through several subcortical stages until it reaches primary AC (Figure 1.1). Being at the interface, the cochlea “translates” each incoming sound from the “language” of pressure waves in the air to the “language” of neuronal activity. This transition is very important as it sets the stage for all subsequent processing that happens subcortically and in the AC.

Sound first enters the ear canal where it causes the vibration of the tympanic membrane. The tympanic membrane in turn transmits the sound further via three small bones (*ossicles*): the *malleus* (hammer), the *incus* (anvil) and the *stapes* (stirrup). The stirrup then causes the vibration of the *oval window* which in turn transmits the motion to the inside space of the cochlea, the *scala vestibuli* and then *scala tympani*



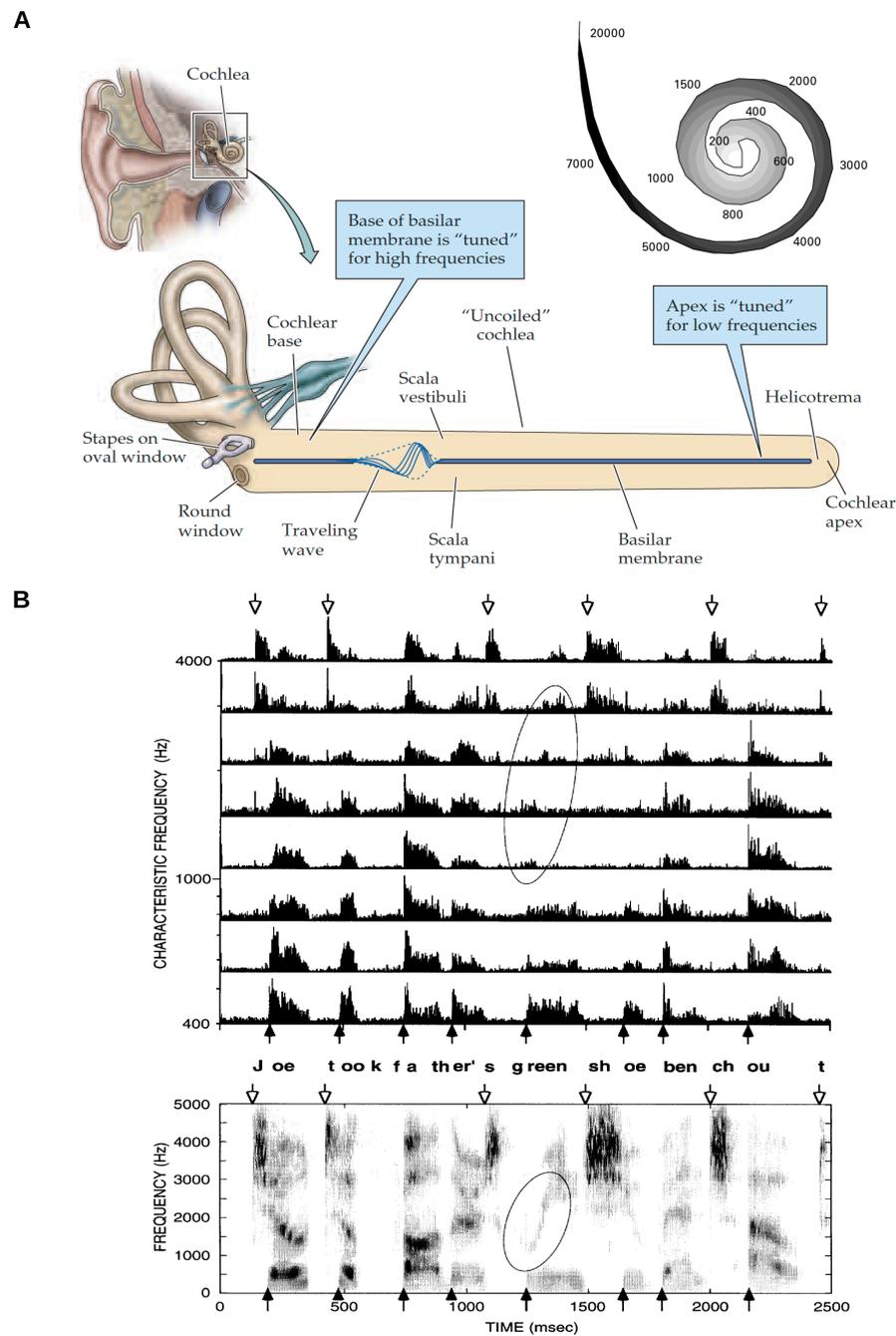
**Fig. 1.1 The Auditory Pathway.**

Afferent pathways from the cochlea up to the auditory cortex are represented. Copied from Purves et al. (2004).

(Figure 1.2A). Both *scala vestibuli* and *scala tympani* are filled with a liquid called “perilymph”. The basilar membrane itself is shaped like a door wedge, with a more thick and stiff base and a thinner and more flexible apex (Figure 1.2A). Thus, there are gradients running from the base to the apex, an inertial gradient of the perilymph caused by the fact that sounds need to travel through more liquid in order to reach the apex, and a stiffness gradient of the basilar membrane which runs in opposite direction. These gradients work together to determine the vibrational properties of the basilar membrane. The vibration of the oval window generates a traveling wave along the basilar membrane where higher frequencies excite the base while lower ones cause vibrations closer to the apex. Due to the biophysics of the system, the spacing of the frequencies is logarithmic, with lower frequencies having a relatively larger representation along the basilar membrane (Figure 1.2A, see inset). This arrangement gives rise to an auditory map: the “tonotopic” or “cochleotopic” one, which is preserved along the core regions of the subcortical auditory pathway up to the primary AC (Linden and Schreiner, 2003; Schreiner and Winer, 2007).

Due to the features described above, the cochlea has often been referred to as a “frequency analyzer”, akin to a device performing Fourier transformation so as to extract a frequency spectrum. However, a more accurate analogy is that of a mechanical band-pass filter, with a popular model being that of a “gamma-tone filter bank”. A gammatone filter is a linear filter described as the product of a gamma distribution and sinusoidal wave. It is essentially a bandpass filter. A gamma-tone filter bank is collection of such bandpass filters, each with a different center frequency and a certain bandwidth. It is a widely used model of auditory filters in the auditory system as, to a first approximation, it captures well the mechanical filtering properties of the cochlea. However, other filter banks have been found to more accurately describe the cochlear filtering in recent years (explored in Rahman et al. (2020)). Such filters have also been derived theoretically using ideas from information theory, where the mutual information between the input sounds and model neuronal responses was maximized by filters resembling a gamma-tone filter bank (Smith and Lewicki, 2006).

So called “hair cells” along the basilar membrane located within the organ of Corti are activated by the local mechanical displacement of the cochlea, and are responsible for translating this motion into changes in membrane potential. They come in two flavours, inner and outer hair cells, with the outer hair cells being much more numerous. Just like people, they have different “personalities”: inner hair cells like to “chat” and



**Fig. 1.2 The Cochlea.**

**A**, The basic anatomy of the cochlea is summarised. The biophysical properties of the basilar membrane and surrounding tissues allow it to function as a frequency analyzer with high frequencies represented at the base and lower ones towards the apex. These are logarithmically spaced with the lower frequencies being over-represented as shown by the inset. Adapted from [Purves et al. \(2004\)](#) and [Schnupp et al. \(2011\)](#). **B**, The upper panel represents a "Neurogram" of the sentence "Joe took father's green shoe bench out". Each row represents a Peristimulus time histogram (PSTH) constructed from the firing rates of auditory nerve fibers, which are arranged by their Characteristic Frequency (CF) (high-to-low). The lower panel shows a sound spectrogram of the same sentence to emphasize the similarities between the two, including finer details such as the format transitions in the word "green" (see ellipses). Copied from [Delgutte \(2002\)](#).

transmit the majority of information to the auditory nerves, while outer hair cells like to “dance” and are involved in active amplification of the motions of the basilar membrane. The release of glutamate from the base of the inner hair cells activates neurons in the spiral ganglia, whose axons are fibers of the auditory nerve<sup>2</sup>. Each auditory nerve fibre is most sensitive at a specific frequency - the Characteristic Frequency (CF) of the cell. If the sound level threshold for firing of the nerve fiber for each frequency is measured, the CF is the frequency where this threshold occurs at the lowest sound level.

Within auditory nerves, there are broadly three general information encoding strategies: “place code”, “rate code” and “temporal code”. The tonotopic arrangement of CFs amongst auditory fibers represents a “place code” for frequency. The “rate code” provides the basis for the processing of sound intensity in the auditory system. As the sound level increases, the fibers of the auditory nerve fire more vigorously (higher number of spikes per unit time i.e. “spike rate”), up to an intensity which will cause the firing rate to saturate. Different groups of fibers have different intensity thresholds, sensitivity and dynamic ranges. Interestingly, a recent paper by [Rahman et al. \(2020\)](#) found that incorporating a multi-fiber model of the cochlear processing allowed for more accurate predictions of the cortical neuronal responses thus validating its computational importance. The combination of different frequencies and intensities that each auditory nerve fibre responds to can be represented by its Frequency Response Area (FRA). We will use this simple yet powerful tool extensively in Chapters 2 and 3. The FRA allows for the identification of two key properties of the auditory nerve fibers, but also auditory neurons more generally: (1) CF, the pure tone frequency eliciting a response at the lowest sound level the fibre is responsive to (threshold level); (2) Best Frequency (BF), the frequency eliciting the maximum response across frequency for a fixed sound level. The third coding strategy, “temporal coding”, can be seen in the phase-locked firing of nerve fibers in response to sound frequencies up to ~1.5kHz, going to ~4kHz in different species ([Palmer and Russell, 1986](#); [Köppl, 1997](#); [Kettner et al., 1985](#); [Sumner and Palmer, 2012](#)). Each nerve fiber has a higher probability of firing at a particular phase of the sine wave for a given frequency. This phase locking is somewhat stochastic as the firing of the nerve fibers may coincide with a phase of the waveform which is not necessarily its peak. Although an individual fiber cannot fire at every sound wave peak, the combined effort of many fibers together can allow reliable phase-locking by the so called “volley principle” or “volley theory”. First proposed in

---

<sup>2</sup>Part of the VIIIth cranial nerve

the 1930s by Bray and Wever ([Wever and Bray, 1937](#)), the volley principle states that auditory nerve fibers fire action potentials slightly out of phase with one another, such that when their spiking is combined, they can phase-lock to frequencies much higher than the individual fibers.

Due to the combined efforts of these three encoding strategies, the distribution of activity across the population of auditory nerve fibers provides a sort of “neurogram” of the sound (Figure 1.2B). Thus, there is a strong correspondence between a sound spectrogram and the neurogram composed from the auditory nerve fibers indicating that many of the sound features, including fine spectral details and modulations, are faithfully transmitted.

### 1.2.2 Cochlear Nucleus

The next step along our auditory tour brings us to the Cochlear Nucleus (CN). The auditory nerve fibers synapse onto cells in the CN of the brainstem, which has three distinct regions, each of which preserves the tonotopy established earlier. One branch of the auditory nerve enters the Anteroventral Cochlear Nucleus (AVCN), while another branch synapses in the Posteroventral Cochlear Nucleus (PVCN) and then proceeds onto the dorsal division (DCN). CN neurons come in several different types, which are characterised by their anatomical location, morphology, physiology, synaptic inputs, and temporal and spectral response properties. A detailed discussion of these cell types is beyond the scope of this thesis. The subdivision of the ventral CN represents two main pathways to extract and enhance frequency and timing information: the sound localization path (AVCN) and the sound identification path (PVCN) ([Young and Oertel, 2004](#)). The AVCN provides input to the superior Superior Olivary Complex (SOC), the first structure where information from both ears is combined, where Interaural Time Difference (ITD) and Interaural Level Difference (ILD) are represented ([Carr, 1993](#)). Broadly tuned cells in the PVCN compute estimates of a level invariant spectral representation of sound ([May et al., 1998](#)).

### 1.2.3 Superior Olivary Complex

Auditory signals from the two ears are first combined at the level of the SOC of the brainstem, which thus plays an important role in extracting binaural information

related to spatial hearing. The Medial Superior Olive (MSO) is dedicated to the processing of ITDs, while ILDs are encoded by neurons in the Lateral Superior Olive (LSO). The MSO contains cells with bipolar dendrites that extend both medially and laterally (Purves et al., 2004). The lateral dendrites receive input from the ipsilateral AVCN, and the medial dendrites receive input from the contralateral AVCN (both inputs are excitatory). As might be expected, the MSO cells work as coincidence detectors, responding when both excitatory signals arrive at the same time (Brand et al., 2002). This requires a faithful representation of the timing of peripheral events on a microsecond time scale, which is made possible by low threshold voltage-gated potassium channels, unusually rapid-acting transmitter-gated channels, and exceptionally large nerve terminals (Trussell, 1999). The LSO in each hemisphere receives direct excitatory inputs from the ipsilateral AVCN and indirect (via the trapezoid body) inhibitory inputs from the contralateral CN, which allows ILDs to be computed by integrating level information from both ears (Sanes, 1990). Spatial information is best represented for stimuli originating in the frontal hemi-field of space, especially closer to the midline by neurons in both the MSO and LSO (Phillips and Irvine, 1981; McAlpine et al., 2001). It is also worth noting that apart from processing binaural spatial information, neurons within the SOC could play an important role in preserving the dynamic range of the input inherited from the CN, facilitated by inhibitory descending projections (May and Sachs, 1992).

#### 1.2.4 Nucleus of the Lateral Lemniscus

A second major pathway from the CN bypasses the SOC and terminates in the Nucleus of the Lateral Lemniscus (NLL) on the contralateral side of the brainstem (see Figure 1.1). These particular pathways respond to sound arriving at one ear only and are thus referred to as monaural. Some cells in the lateral lemniscus nuclei signal the onset of sound, regardless of its intensity or frequency, while other cells in the lateral lemniscus nuclei process other temporal aspects of sound, such as duration (Nayagam et al., 2005). The precise role of these pathways in processing temporal features of sound is not yet known. As with the outputs of the SOC, the pathways from the nuclei of the lateral lemniscus converge at the Inferior Colliculus (IC).

### 1.2.5 Inferior Colliculus

Next, we arrive at the IC. At the level of IC, temporal and spectral information have been proposed to be topographically but mutually orthogonally mapped (Langner, 1992). The IC also play a central role in multisensory and other non-auditory contributions to hearing as it is a necessary relay for most of the ascending and descending auditory information (for a detailed overview see Casseday and Winer (2005)).

Positioned fairly early on in the auditory system, the IC consists of the Central Nucleus of the Inferior Colliculus (CNIC) and a surrounding shell, which can be subdivided into the Dorsal Cortex of the Inferior Colliculus (DCIC) and Lateral Cortex of the Inferior Colliculus (LCIC). The CNIC primarily sends ascending auditory information to the thalamus (e.g. Kudo and Niimi (1980) and Calford and Aitkin (1983)) which then reach AC. The shell portions of the IC also project to the auditory thalamus (Kudo and Niimi, 1980; Calford and Aitkin, 1983) as well as the superior colliculus (SC) (from the dorsomedial part of the IC: Druga and Syka (1984); shell of the IC: Zhang et al. (1987); rostral pole of the IC: K Harting and Van Lieshout (2000)), and descending information projects back to the auditory brainstem (see Huffman and Henson (1990)).

In the mouse, the DCIC is located superficially and thus it has been possible to investigate its function using *in vivo* two-photon calcium imaging. Such studies have highlighted the boundary between the DCIC and the CNIC (Barnstedt et al., 2015) more precisely comparing to prior microelectrode recordings (Portfors et al., 2011). The Barnstedt et al. (2015) study confirmed the tonotopic organization of the CNIC, which can be divided into iso-frequency laminae defining a low-to-high, dorsal-to-ventral gradient of BFS (Merzenich and Reid, 1974; Portfors et al., 2011). On the other hand, tonotopy in the DCIC appeared to be less clear in both electrophysiological and imaging studies (Barnstedt et al., 2015; Ito and Oliver, 2014; Casseday and Winer, 2005).

The processing that occurs within IC and earlier nuclei is thought to sharpen the spectral peaks of natural complex sounds (Shamma, 2000). This is accomplished via lateral inhibition across neurons, according to their frequency selectivity. In general, the first evidence of non-monotonic rate-level functions in the auditory system are observed in IC (Irvine and Gago, 1990), although in the guinea pig these have been observed as early as the CN (Stabler et al., 1996). A rate-level function defines the relationship between the sound level and the firing rate of the neuron. A non-monotonic rate-level function is a function, where the firing rate of the neuron is not consistently increasing

or decreasing with increasing sound level, defining a more complex relationship between the two (Heil et al., 2011). Many of the neurons in the IC show bandpass sensitivity to the sinusoidal amplitude modulation of sounds (George et al., 2016).

### 1.2.6 Medial Geniculate Body

Our final stop of the subcortical auditory tour is the Medial Geniculate Body (MGB). The MGB in the thalamus is the auditory relay station between the IC and AC and receives convergent inputs from the separate spectral and temporal pathways in lower areas. Consequently, it has been argued that the MGB is the first structure where cells are found to respond to more complex spectro-temporal patterns (Sharpee et al., 2004; Atencio et al., 2012; Atencio and Sharpee, 2017) (see Figure 1.3C). While preserving a tonotopic organization (Imig and Morel, 1985), cells in the MGB are also selective to specific combinations of frequencies and specific time intervals between frequencies (on the order of milliseconds) (Bartlett, 2013). Although it is often described as a sensory “relay” in textbook accounts, the MGB is a much more rich structure akin to a “funhouse mirror” (Bartlett, 2013), due to its filtering and distortion of incoming inputs so as to facilitate the representation of relevant sensory features in AC later on (Sherman and Guillery, 2002). The MGB also plays an important role in multisensory integration, e.g. facilitating interactions between the somatosensory and auditory systems (Lohse et al., 2020b).

Growing evidence suggests that specialized thalamic circuits involving MGB can play an important role in determining the behavioural significance of sound stimuli (Weinberger, 2011). Recently, projections from the amygdala to the thalamic reticular nucleus (TRN) of MGB were found to amplify cortical sound responses (Aizenberg et al., 2019). Conversely, neurons within the MGB project onto the amygdala themselves and can modulate its activity. These projections are thought to underlie the conditioning of emotional reactions to acoustic stimuli (Clugnet et al., 1990; Farb and LeDoux, 1997).

It is also worth pointing out that the ventral divisions of MGB maintain a clear tonotopic organisation, similar to CN, SOC, NLL and the core parts of IC. In contrast, the dorsal divisions of the MGB lack a clear tonotopic order similar to the shell regions of the IC described before. These tonotopically organized auditory midbrain and MGB structures are sometimes referred to as “lemniscal”, and those lacking clear tonotopy as “nonlemniscal”. In addition, despite the fact that most ascending thalamic inputs

come from IC, there also exists a direct, non-tonotopic, short-latency projection from the dorsal CN to the medial division of the MGB (Malmierca et al., 2002; Schofield et al., 2014), which may play a role in multisensory integration.

## 1.3 Cortical auditory processing

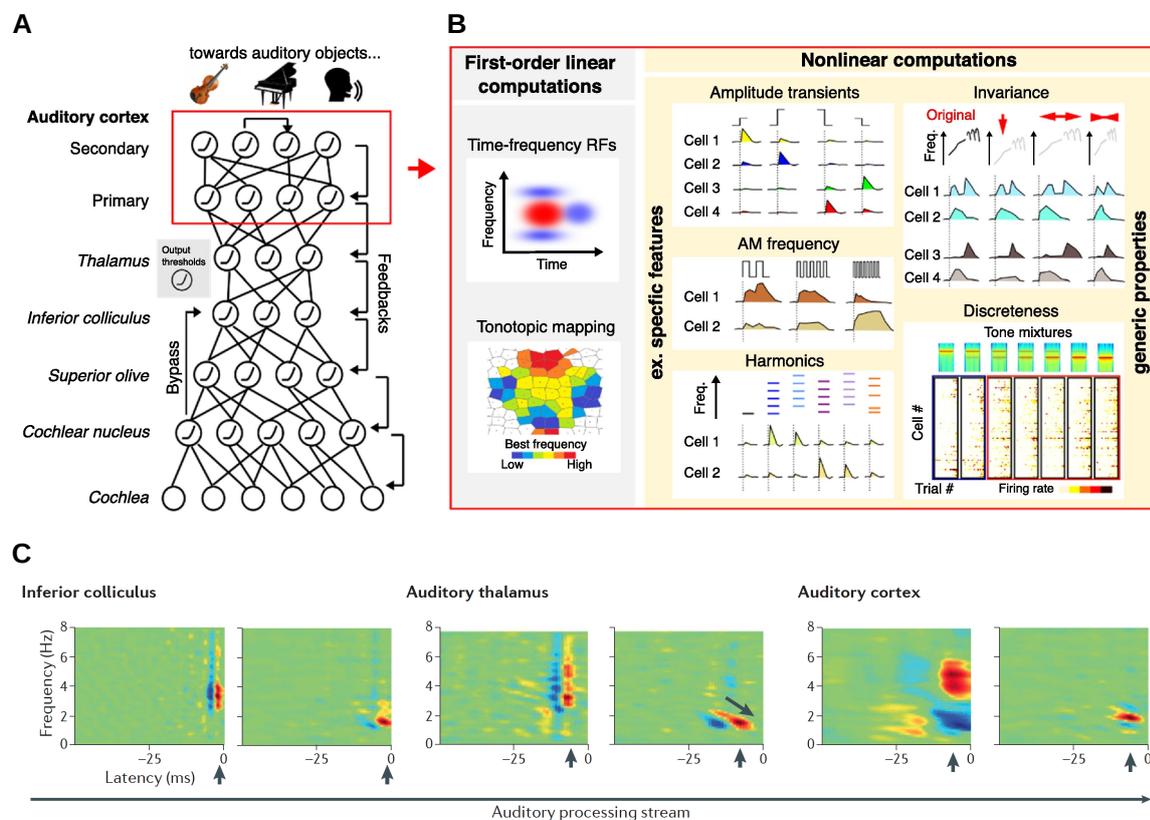
The mammalian AC typically consists of a core area composed of primary AC and potentially other fields, for example the Anterior Auditory Field (AAF), surrounded by other areas, often named belt or secondary fields, and tertiary areas (parabelt) (Malmierca, 2013). While auditory activity in cortex occurs far beyond AC proper (Meredith et al., 2009), in this chapter I will focus on primary AC and its surrounding fields.

### 1.3.1 Auditory cortex: Bird’s-eye view

Whether AC represents just acoustic features or building blocks for ‘auditory objects’ is a longstanding question (Nelken et al., 2003; Mizrahi et al., 2014; Nelken et al., 2014; Wang, 2013; Chechik and Nelken, 2012). The AC is an integral part of the network of brain regions responsible for generating meaning from sounds, auditory perceptual decision-making, and learning (King et al., 2018). Here I will review how AC builds progressively more complex acoustic features, as compared to simpler subcortical processing steps, in addition to exploring the role of AC in cognition, learning, and decision making (Figure 1.3).

As we described in subsection 1.2.1, the tonotopic organisation of auditory space is established early in the auditory pathway at the level of the cochlea and maintained throughout the subcortical processing stages. Within primary AC, the local circuits are organized tonotopically (Figure 1.3B, left column, bottom), although there has been a substantial controversy as to how tight this organization is (Rothschild et al., 2010; Guo et al., 2012; Kanold et al., 2014; Zeng et al., 2019; Tischbirek et al., 2019). We will hold off a detailed discussion of this issue for section 1.4 and Chapter 2. For now we will focus on the spectro-temporal tuning properties of cortical neurons.

The tuning properties of auditory neurons, including those of AC, are defined by their receptive fields, which describe the stimulus features to which they are most



**Fig. 1.3 Auditory cortex properties.**

**A**, Throughout the auditory system raw cochlear inputs are structured into biologically relevant percepts (auditory objects). This transformation requires a complex ensemble of non-linear computations, which are schematized as a multilayer network linking simple non-linearities (e.g. spike threshold) with an elaborate connection graph that includes feedforward, feedback and lateral connections. **B**, Beyond simple computations such as spatially organized frequency tuning, i.e. tonotopy (left), the AC displays a number of non-linear computations sketched in the middle column, including sensitivity to sound onsets and offsets of particular amplitude, Amplitude Modulation (AM) frequencies, or even harmonicity. Also, generic response properties, such as invariance to modification of acoustic parameters (top right) and discrete coordinated population response switches (bottom right). **C**, Spectro-Temporal Receptive Field (STRF) at different levels of the auditory system. In IC, some neurons have STRFs that are narrowly tuned in time, but broadly tuned in frequency (left panel) or narrowly tuned in both time and frequency (right panel). Thalamic neurons have greater latencies than IC neurons (left panel) and also demonstrate selectivity to patterns in time, e.g. a descending tone sweep (right panel). STRFs are slower and more complex in the AC. A and B copied from [Kuchibhotla and Bathellier \(2018\)](#), C copied from [Theunissen and Elie \(2014\)](#).

responsive. Auditory neurons are most commonly characterized by their sensitivity to sound frequency and a popular method is the FRA, which captures the neuronal responses to different frequency-level combinations (Schnupp et al., 2011). The STRF is the dominant computational tool for modeling the responses of auditory neurons, which captures not only their frequency preference but also sensitivity to the recent history of the sound stimulus (de Boer and Kuypers, 1968; Aertsen and Johannesma, 1981; Theunissen et al., 2001). Most widely used as part of a linear-nonlinear (LN) model, an STRF comprises a set of coefficients that describe how the response of the neuron at each moment in time can be modelled as a linear weighted sum of the recent history of the stimulus power in different spectral channels (Figure 1.3B,C). So how do the STRFs of AC differ from other subcortical structures like the IC? Using ideas from information theory, relying on the maximally informative dimensions (MID) approach, it has been found that multiple stimulus dimensions are required to describe the responses of neurons in primary AC (Sharpee et al., 2004; Atencio and Sharpee, 2017), whereas a similar approach requires only a single dimension to describe the majority of neurons in the IC (Atencio et al., 2012). This suggests that neuronal complexity is higher in the cortex and that this complexity can be captured by nonlinear combination of the responses of multiple simpler units (Harper et al., 2016), a finding that also seems to hold true in the AC equivalent in the songbird (Figure 1.3C) (Kozlov and Gentner, 2014, 2016).

The idea that sensory systems build a hierarchical feature representation (Figure 1.3A) has been quite prominent since the seminal discoveries of Hubel and Wiesel (1962). In the visual system, inputs from neurons within the Lateral Geniculate Nucleus (LGN) can be combined to generate the receptive fields of “simple cells” in Primary Visual Cortex (V1), and in turn the output of several “simple cells” can be pooled to create a phase-invariant “complex cell” (Carandini, 2005). The connectivity and structure of the visual system has been an important inspiration for the creation of convolutional neural networks, one of the workhorses of modern deep learning, where increasingly more complex features are developed with each successive layer (Lecun et al., 1998). In turn, deep convolutional neural networks have become one of the most successful quantitative models of various aspects of human and animal vision, predicting well both task performance and neuronal activity (Yamins et al., 2014; Lindsay, 2020).

In the auditory system, a similar hierarchical convolutional neural network has proven equally successful at capturing behavioural and physiological properties of AC (Kell et al., 2018). Interestingly, Kell et al. (2018) found that a deep network with random weights performed well above chance in both some of the auditory tasks and in predicting AC activity, highlighting the importance of deep hierarchical representation (Figure 1.3A). Other authors have suggested that a deep recurrent neural network might be a better model of auditory processing (Thompson, 2020). Combining acoustic (spectro-temporal) features in the auditory system is not as straightforward as in the visual system, e.g. there are no obvious equivalents of simple and complex cells. Nonetheless, there are some interesting examples. A recent study showed that the core AC of primates includes cells that specifically detect patterns of frequency harmonics and act as “harmonic templates” (Figure 1.3B, middle column, bottom), i.e. they can pull information from other cells with BFs that are multiples of one another (harmonics) (Feng and Wang, 2017). Other emergent properties include the detection of sound onsets and offsets of particular amplitude, which reflected multi-layered non-linearities (Deneux et al., 2016), and also sensitivity to AM frequencies (Gao and Wehr, 2015) (Figure 1.3B, middle column, top and middle). There is a build-up in the complexity of the sound representation within AC itself. In a recent study, Norman-Haignere and McDermott (2018) used an iterative procedure to synthesize artificial sounds which would give identical responses to natural sounds when presented to a model of cortical processing based on a bank of STRFs. When presented to human participants, the synthetic “model-matched” stimuli were indistinguishable from natural sounds in the responses of primary AC, while the natural stimuli were much more effective in driving secondary AC, suggesting that it represents sound properties beyond simple spectro-temporal features.

In addition to becoming sensitive to progressively more complex sound features, auditory cortical neurons exhibit more generic properties associated with perceptual phenomena, such as invariance. Invariance refers to the stability of representations with respect to small changes in acoustic parameters. Two recent studies in rats have shown that auditory cortex neurons respond to vocalizations or water sounds with a certain degree of robustness against various acoustic modifications (Carruthers et al., 2015; Blackwell et al., 2016). Interestingly, invariance for vocalization was tested both in primary and non-primary auditory cortex and was found to be more pronounced in non-primary areas (Figure 1.3B, right column, top), suggesting that invariance properties progressively emerge along the cortical hierarchy, correlating with the weakening of the

tonotopic map (Issa et al., 2014). Similarly, invariant representation of natural stimuli corrupted by background noise has been found to be more prominent in non-primary human AC compared to primary (Kell and McDermott, 2019). Neurons in the Primary Auditory Cortex (A1) of the ferret (Rabinowitz et al., 2011, 2012) and mice (Cooke et al., 2018) show compensatory adaptation to sound contrast (defined as the variance of the sound level distribution). When the contrast of the input to a given neuron is high, the gain of the neuron is reduced, thereby making it relatively insensitive to changes in sound level. When the contrast of the input is low, the gain of the neuron rises, increasing its sensitivity. In another ferret study, Rabinowitz et al. (2013) found that the subcortical structures are affected more by the addition of noise than AC and that noise tolerance was correlated with the adaptation to stimulus statistics, so that populations that showed the strongest adaptation to stimulus statistics were also the most noise-tolerant. This study shows that the increase in adaptation to sound statistics from auditory nerve to midbrain to cortex is helping in the construction of noise-invariant sound representations in the higher auditory brain. In Chapter 4 we will explore such invariant representation in ferret AC to a non-stationary auditory noise caused by echoes (reverberation).

Going beyond representations of natural sounds, the AC is believed to have important cognitive roles such as auditory scene analysis, prediction and decision making, in addition to being critical for many types of learning. We will briefly review those below before proceeding to describing the ferret AC.

One of the main functions of the auditory system is to separate attended sound sources from an often complex mixture of competing stimuli. Important acoustic features like sound intensity, pitch, timbre and location need to be processed separately and then combined so as to form a coherent auditory percept, which ultimately leads to the isolation of distinct auditory objects (Griffiths and Warren, 2004). Despite the fact that some elements of auditory scene analysis have been observed at various subcortical stages (Pressnitzer et al., 2008; Kondo and Kashino, 2009; Yao et al., 2015), it is ultimately at the level of AC where the features extracted from sound stimuli and top-down attentional mechanisms converge (Shamma et al., 2011). Thus, many recent studies have highlighted the crucial role of AC in delivering perceptual representations that are stable, created on the basis of grouping and segregation of spatial, temporal and spectral regularities inherent in the auditory environment (Puvvada and Simon,

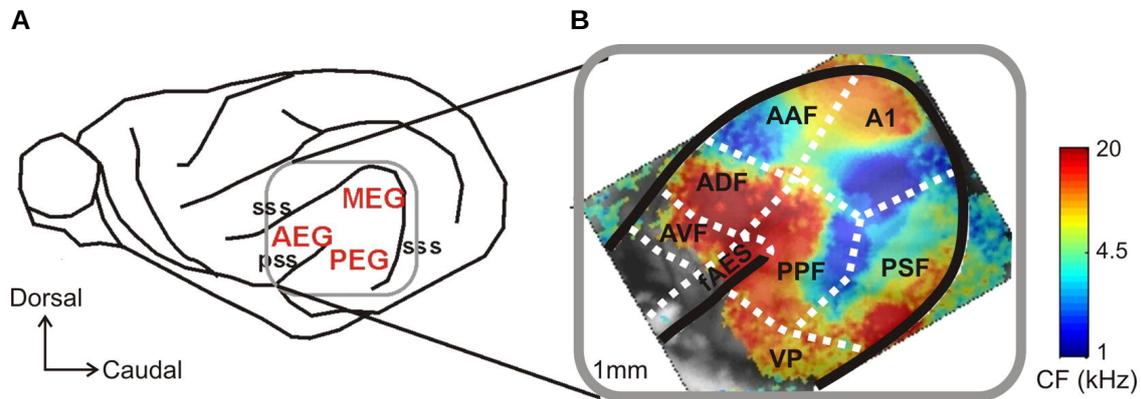
2017; Alain et al., 2017; Teki and Griffiths, 2016; Teki et al., 2016; Christison-Lagay et al., 2015; Fishman et al., 2014).

Another important aspect of the functioning of the auditory and other sensory systems is the anticipation and prediction of future input. A very influential paradigm in the field has been that of predictive coding, first championed by Rao, Ballard and Friston. In the predictive coding framework the nervous system is believed to learn to anticipate future input, which is then compared to the incoming sensory information and a “prediction error” signal is generated which contains the more salient information from the external world (Rao and Ballard, 1999; Friston, 2005). Due to their periodic nature, stimuli evoking pitch perception are inherently predictable and predictive coding has been successful in explaining pitch representation within various areas of the AC in a hierarchical and distributed fashion (Kumar et al., 2011). Interestingly, many of the key components of the predictive coding framework, such as prediction precision, prediction change, surprise and prediction error have all been observed in the AC (Sedley et al., 2016). Other studies have shown that neurons in primary AC are sensitive to surprise and deviations from a repeated sound pattern (Polterovich et al., 2018) and that they can encode prediction error (Rubin et al., 2016). It has also been observed that the signals associated with prediction error tend to increase along the auditory system, becoming more prominent in AC (Parras et al., 2017). In addition, Singer et al. (2018) have shown that a neural network trained to predict future sound stimuli from those in the recent past produces receptive fields (STRFs), which are qualitatively and quantitatively similar to those found in the AC, underscoring the importance of prediction in the auditory system.

In classical textbook accounts of decision making, this ability has been ascribed to higher cortical areas, such as prefrontal and parietal cortices, in conjunction with certain subcortical structures like the basal ganglia (Purves et al., 2004). Nonetheless, new experimental evidence is emerging arguing that AC represents decision-related signals and behavioural choice variables, in addition to its more “traditional role” of encoding spectro-temporal and other physical sound features (Francis et al., 2018; Niwa et al., 2012; Bizley et al., 2013; Runyan et al., 2017). An elegant study by Tsunada et al. (2016) used micro-stimulation in macaque AC to demonstrate a causal contribution to auditory perceptual decision-making.

Finally, AC is also crucial for various types of learning. Some of the learning is of a more passive nature, relying on the identification of certain statistical regularities in

the sounds being presented. In the songbird, microelectrode recordings from forebrain areas that functionally correspond to the AC in mammals have delivered evidence for statistical learning, which manifested as a reduction in the firing rate of neurons for familiar versus novel sequences (Lu and Vicario, 2014). In humans and monkeys that were exposed to passive learning involving sequences of pseudo-speech stimuli governed by an artificial grammar structure, it was observed that violating the rules in these sequences activated homologous cortical areas, in addition to modulating nested low-frequency phase and high-gamma amplitude coupling in the AC (Wilson et al., 2015; Kikuchi et al., 2017). Such oscillatory coupling is believed to facilitate speech processing in the human AC, thus potentially representing a mechanism for the analysis of auditory sequences that is conserved evolutionarily (Giraud and Poeppel, 2012). There have been also a plethora of studies that demonstrate that changes in the encoding of stimulus properties in AC neurons is associated with perceptual learning (Irvine, 2018). Such learning commonly involves the overrepresentation of stimuli in AC which were used for training, and the amount of plasticity in these representations is believed to be indicative of both the strength of the associative memory and the behavioural importance of the auditory stimuli (Bieszczad and Weinberger, 2010). An interesting recent study found that when gerbils are trained to detect changes in an amplitude-modulation task, there was a strong correlation in the time course of cortical and behavioural plasticity, in addition to a diminished learning capacity when AC was inactivated, without affecting detection thresholds (Caras and Sanes, 2017). The adaptation observed after training on a spatial task with one of the ears plugged was impaired when AC was inactivated in ferrets (Nodal et al., 2012; Bajo et al., 2019). A few neuromodulatory systems have been proposed as plausible candidates for the mediation of reinforcement signals and information about behavioral context to AC which thus would play a pivotal role in sound processing and plasticity. These include the cholinergic basal forebrain (Froemke et al., 2013) and the noradrenergic locus coeruleus (Glennon et al., 2019). In line with this view, recent studies have shown that cholinergic inputs primarily target inhibitory interneurons in the AC (Nelson and Mooney, 2016; Kuchibhotla et al., 2017), indicating that the excitatory-inhibitory balance might be affected transiently during learning (Carcea and Froemke, 2013).



**Fig. 1.4 Auditory cortex of the ferret.**

**A**, Location of ferret auditory cortex on the middle, anterior and posterior ectosylvian gyri (MEG, AEG and PEG, respectively). The suprasylvian (sss) and pseudosylvian (pss) sulci are also labelled. **B**, The inset shows the location of eight auditory cortical fields: the primary and anterior auditory fields (A1 and AAF), the tonotopically organized Posterior Suprasylvian Field (PSF) and Posterior Pseudosylvian Field (PPF), and the non-tonotopic Anterior Dorsal Field (ADF) and Anterior Ventral Field (AVF), ventroposterior area (VP) and anterior ectosylvian sulcal field (faES). The color scale shows the tonotopic organization of these auditory cortices, as visualized using optical imaging of intrinsic signals. Copied from [Bizley and King \(2009\)](#).

### 1.3.2 Ferret auditory cortex

After this quick tour of the auditory system and more general overview of AC we will briefly review the ferret as a model organism and describe some basic anatomical and physiological features of its AC since the majority of the experiments during my DPhil involved ferrets and I focused on the properties of AC.

Ferrets (*Mustela putorius*) are a popular animal model in auditory science, as they are well suited to both behavioural and physiological studies of hearing. They can be trained to perform a variety of sound localization ([Nodal et al., 2008](#); [Bizley et al., 2007b](#); [Kacelnik et al., 2006](#); [Kelly and Kavanagh, 1994](#)), discrimination ([Walker et al., 2009](#)), and detection tasks ([Hine et al., 1994](#); [Kelly et al., 1986b, 1996](#); [Fritz et al., 2007](#); [Fritz, 2005](#)). The range of sound frequencies that are audible to ferrets (approximately 20Hz–44 kHz) entirely overlaps, and extends beyond, that of humans ([Nodal and King, 2014](#)). In contrast to many rodents ([Kelly and Kavanagh, 1994](#)), particularly mice, whose audible range is shifted to higher frequencies, the sensitivity of ferrets to low-frequency sounds allows the study of various aspects of hearing, including pitch

perception, sound localization, and the processing of communication calls, within a similar frequency range to that used by humans (Nodal and King, 2014).

Similar to other mammals, there are several ferret auditory cortical areas that possess different functional properties (Kowalski et al., 1995; Nelken et al., 2004; Bizley et al., 2005) (Figure 1.4). The Ectosylvian Gyrus (EG) of the ferret is divided into Middle Ectosylvian Gyrus (MEG), Anterior Ectosylvian Gyrus (AEG), Posterior Ectosylvian Gyrus (PEG) (Figure 1.4A). The MEG comprises the primary auditory cortex (A1) and the anterior auditory field (AAF). The neurons that are located in these areas tend to have short latencies when responding to sounds. In addition, the A1 and AAF neurons are organised in topographic fashion according to their BF in “tonotopic” maps, with the tip of the MEG representing higher BFs, which get progressively lower as we go further away towards the AEG and PEG (Figure 1.4B). It is well established that there are homologous primary areas in other mammals which resemble A1 and AAF, including humans (Nodal and King, 2014). However, figuring out the homologues to the ferret secondary AC areas has been less straightforward. In addition to the primary A1 and AAF, there are two more secondary areas along the PEG which are tonotopically organised, the PPF and PSF (Figure 1.4B). The neurons located in PPF and PSF can be delineated from the primary areas according to the temporal characteristics of their responses (Bizley et al., 2005). There are also two other secondary areas located in the AEG, the anterior dorsal and ventral fields (ADF and AVF), where neurons are sound responsive but do not exhibit the same tonotopic structure as PPF and PSF. These areas are primarily tuned to higher frequencies. Most of the areas in ferret AC in addition to responding to sounds, have been shown to receive visual inputs which originate in different parts of the visual cortex (Bizley et al., 2007a). In light of this, the ventral-most region in the AEG, the AVF (Figure 1.4B), is probably more accurately described as a higher level multisensory area, owing to the fact that it houses a fairly large number of visually responsive cells and it is in very close proximity to a somatosensory field (Keniston et al., 2009). Additional studies have underscored the multisensory properties of neurons located in the AEG in general (Ramsay and Meredith, 2004; Manger et al., 2005). Multisensory inputs from other modalities to AC have been reported in other mammals too, including primates (Kayser et al., 2009). The roles of such connections are not well understood. One possibility is that they improve the ability of auditory cortical neurons to encode the identity (Ghazanfar et al., 2008; Kayser et al., 2010) or location (Bizley and King, 2008) of the sound source.

Studies of the ferret AC have provided important insights into the encoding and representation of various sound properties such as spatial location, periodicity and timbre (Schnupp et al., 2001; Bizley et al., 2009; Walker et al., 2011) in addition to furthering our understanding of how cortex represents natural sounds like speech (Rabinowitz et al., 2013; Mesgarani et al., 2008, 2009). Using artificial vowel sounds to explore pitch and timbre perception, Bizley et al. (2009) discovered that different cortical areas represent both spatial and non-spatial sound features in a distributed fashion. However, the representation was not entirely diffuse as neurons within certain areas contained more information in their firing patterns about pitch or timbre (“what”), while the firing of neurons in other areas was more strongly modulated by sound location (“where”). These findings are broadly in line with findings in both human and non-human primates, where the concept of distinct ventral and dorsal processing streams is widely accepted. These were initially assigned to “what” and “where” functions, respectively, but their precise functions, and the extent to which they interact, continue to be debated (Cohen et al., 2016; Albouy et al., 2017; Da Costa et al., 2018; Rauschecker, 2018). At the same time, recordings in the ferret AC have revealed that sound stimulus attributes are not only coded by different spatial patterns of neural activity but also by different temporal patterns within the same population (Walker et al., 2011). Concretely, the firing rate of a given neuron can be modulated by the identity (i.e., spectral timbre) of an artificial vowel sound in one time window and the periodicity of the vowel in a later time window. This type of coding strategy is known as “multiplexing” and it means that neurons can perform mutually invariant information coding of two or more perceptual features. It was shown quite elegantly by Walker et al. (2011) that the faster temporal encoding for vowel identity compared to periodicity in the neural data was mirrored in the behaviour of the ferrets where they detected changes in vowel identity faster than those in periodicity. This cortical processing strategy potentially allows for a more dynamic representation of perceptual attributes or even the acquisition of representations of new attributes that become relevant or are learned through behavioral training.

After reviewing the more general properties and features of the auditory system and AC, we will now focus on more specific topics which will allow us to understand the context for the three results chapters in this thesis. The first topic that we will explore is that of cortical tonotopic maps.

## 1.4 Cortical tonotopic maps

A common textbook view of sensory cortices is that they are organised according to the same general schema, regardless of what computation is performed in each area (Adesnik and Naka, 2018). For example, visual thalamic input rerouted to AC in the ferret led to the development of visual receptive fields in the otherwise “auditory neurons” (Roe et al., 1992). According to this canonical model of cortex (Douglas and Martin, 2004), information flows from the thalamus to cortical layer 4 via thalamocortical synapses, and in turn the layer 4 neurons send the information further to neurons in layers 2/3, which are thought of as the main “computational layer” of cortex, and finally layer 2/3 neurons connect to layer 5 (Dantzker and Callaway, 2000; Shepherd and Svoboda, 2005). The canonical model of cortical circuits represents the accumulation of neuroanatomical and neurophysiological data for a few decades, primarily from studies of the circuitry of visual cortex, where the microcircuitry was first explored more thoroughly (Adesnik and Naka, 2018). However, this simple view has also been challenged, e.g. we now know that in the somatosensory system thalamus can send direct projections to deeper cortical layers circumventing the layer 4-to-layer 2/3 route (Constantinople and Bruno, 2013).

As mentioned earlier, in the primary AC of all species studied to date, there is a characteristic map that mirrors the tonotopic layout of sound frequencies at the sensory periphery, the cochlea. We already discussed the tonotopic organisation of ferret AC in subsection 1.3.2. In the mouse, as in many other species, high frequencies are represented in the rostral part of the cortex and low frequencies in the caudal (Stiebler et al., 1997; Guo et al., 2012; Kanold et al., 2014). This 1D tonotopic axis in both species, ferrets and mice, can be considered the analog of 2D space in the somatosensory and visual cortices, in the sense that it reflects the way the receptors are organised in the sensory periphery. However, because the representation of sound frequency along the cochlea is intrinsically 1D, the organization of AC along the axis perpendicular to the cortical tonotopic axis cannot be inferred in an obvious way from the organization of the sensory periphery (Read et al., 2002). AC is therefore functionally anisotropic: the functional organization along the tonotopic axis is qualitatively different from the organization orthogonal to the tonotopic axis. Taking this into account, the AC differs from both the somatosensory and visual cortices, where the 2D organization of the peripheral receptors is reflected more directly within cortex itself (Oviedo et al., 2010; Oviedo, 2017).

In Chapter 2, we will address these questions in order to gain better understanding of the cortical tonotopic structure in the ferret, compare it to the mouse, try to elucidate the logic behind the local tonotopic variability, and attempt to reconcile some of the differences between electrophysiological and 2-photon calcium imaging studies.

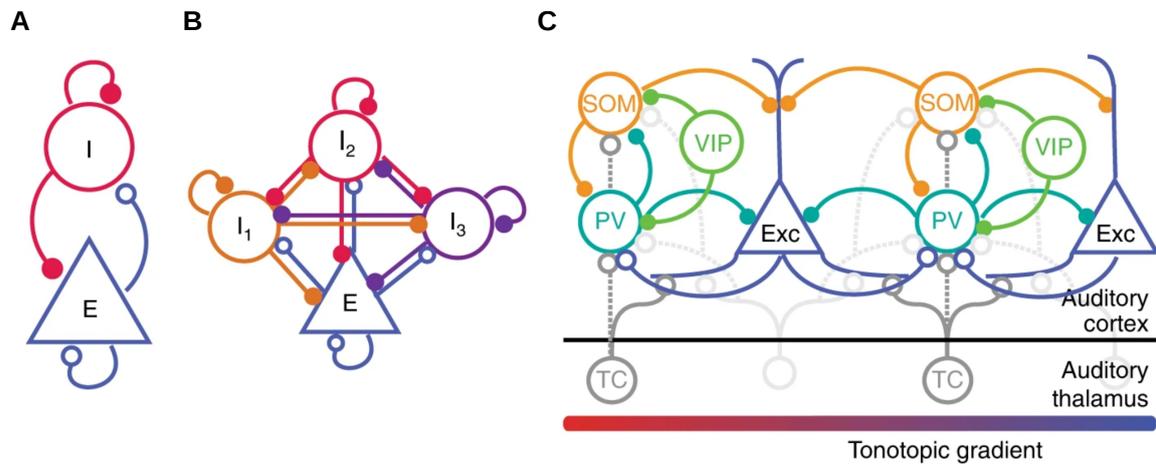
## 1.5 Cortical inhibitory neurons

The second question that I wanted to explore in my DPhil also relates to the organization and properties of the AC. Here, I wanted to investigate how different neuronal types, namely excitatory pyramidal and different types of inhibitory neurons found in AC, differ in their tuning properties and how they might interact with one another during the cortical processing of sound information. Therefore, we will briefly review the current knowledge on cortical neuronal types, holding off a more detailed introduction of excitatory-inhibitory circuits in audition for Chapter 3.

The realisation that neurons in the brain are highly diverse dates back to the seminal anatomical studies of Santiago Ramón y Cajal and Lorente de No ([DeFelipe and Jones, 1988](#)). Despite this incredible diversity, virtually all cortical neurons fall into one of two classes: 1) glutamatergic Excitatory neurons (EXC), or principal cells, which propagate signals within and among various brain regions; 2)  $\gamma$ -aminobutyric acid (GABAergic) Inhibitory neurons (INH)<sup>3</sup>, which gate signal flow and sculpt network dynamics (Figure 1.5) ([Connors and Gutnick, 1990](#); [McCormick et al., 1985](#); [Zeng and Sanes, 2017](#)). The observation that the abundance of “short-axon cells” increased during evolution, which we now know correspond mostly to INH, prompted Cajal to state that INH are responsible for the “functional superiority” of the human brain ([DeFelipe and Jones, 1988](#)). Recently, GABAergic neurons have been classified by electrophysiological, morphological and molecular characteristics (gene expression patterns). The INH can be broadly divided into three large families, characterized by single genetic markers which are non-overlapping and account for almost 100% of all GAD-expressing neurons: the

---

<sup>3</sup>Also referred to as GAD neurons due to the expression of the Glutamate decarboxylase (GAD) enzyme which produces GABA.



**Fig. 1.5 Cortical excitatory-inhibitory circuits.**

**A**, Diagram of excitatory–inhibitory circuit with recurrent connections. Theoretical and experimental studies demonstrate that inhibition stabilizes between excitatory and inhibitory neurons in the auditory cortex. **B**, Inhibitory–excitatory network can be extended to include several interneuron subtypes. **C**, Schematic diagram of connectivity between select neurons in the auditory cortex (note that layer-specific information is omitted here): Exc: Excitatory neurons; PV: parvalbumin-positive interneurons; SOM: somatostatin-positive interneurons; VIP: vasoactive intestinal peptide-positive interneurons; TC: Thalamo-cortical projection neurons. All neuron types receive additional inputs from other brain areas, which were omitted from the diagram for simplicity. Open circles: excitatory synapses; closed circles: inhibitory synapses. Solid lines indicate dominant projections; dashed lines indicate occasional connections. Copied from [Blackwell and Geffen \(2017\)](#).

calcium binding protein parvalbumin (PV)<sup>4</sup>, the neuropeptide somatostatin (SOM<sup>5</sup>), and the neurotransmitter 5-hydroxytryptamine 3a receptors (5-HT3A) (Figure 1.5C) ([Lee et al., 2010](#); [Rudy et al., 2011](#); [Zeisel et al., 2015](#); [Tasic et al., 2016](#)). These three cardinal classes are often further subdivided: Parvalbumin neurons (PV) into “basket cells” targeting the soma and proximal dendrites of EXC and other PV, and “chandelier cells” targeting the axonal initial segment of EXC. The majority of fast-spiking cells observed in electrophysiological experiments are PV, although the other classes can also sometimes show such firing properties. PV are the most numerous (~40%) and best-studied cortical INH. Somatostatin neurons (SOM) (~30%) are subdivided into

<sup>4</sup>Note that we will use GAD, PV, SOM, 5-HT3A, VIP font for the protein markers and GAD, PV, SOM, 5-HT3A, VIP font to indicate the types of neurons, to distinguish between proteins and neuronal types.

<sup>5</sup>Commonly referred to as Sst too.

Martinotti and non-Martinotti cells, which both target the distal dendrites of EXC, and are electrotonically isolated from the soma (some non-Martinotti cells are also known to target PV). 5-Hydroxytryptamine 3A receptor neurons (5-HT<sub>3A</sub>) (~30%) are subdivided into Vasoactive intestinal peptide (VIP) and non-VIP cells, which primarily target other INH (SOM and to a lesser degree PV) and are thus involved in disinhibition (see below). In addition, 5-HT<sub>3A</sub> can be heavily modulated by serotonin and acetylcholine.

Although INH constitute a minority of all cortical neurons, e.g. 10%–20% in rodents (Meyer et al., 2011), their local axons ramify extensively and have enormous influence on cortical computations (Isaacson and Scanziani, 2011; Kepecs and Fishell, 2014; Roux and Buzsáki, 2015; Hattori et al., 2017; Wood et al., 2017). Despite this bewildering complexity, some common themes in cortical circuits involving INH are beginning to crystallize in the form of circuit motifs: 1) Feedforward inhibition, where an afferent excitatory input source, in addition to contacting EXC, also synapses onto INH, which in turn provide disynaptic inhibition to the EXC receiving the excitatory input (Toyama et al., 1974; Miller et al., 2001a; Schiff and Reyes, 2012; Kloc and Maffei, 2014); 2) Feedback Inhibition, where unlike feedforward inhibition, the source of excitation is locally generated by EXC and INH synapse back to the local EXC population, either the same neuron (recurrent inhibition), or neighbouring neurons (lateral inhibition) (Adesnik et al., 2012; Roux and Buzsáki, 2015; Buzsáki and Wang, 2012; Couey et al., 2013); 3) Disinhibition, characterized by the release of inhibition of EXC by INH synapsing on other INH, in a double negative motif (Somogyi et al., 1983; Kisvárdy et al., 1993; Caputi et al., 2009; Lee et al., 2013).

A long-held idea in the field is that different INH execute some form of arithmetic operations, common ones being subtraction or division (Silver, 2010). Inhibition can thus change the input–output relationship between the excitatory drive and the resulting firing rate in EXC: either by decreasing its slope divisively or by a subtractive shift. In turn, these elementary operations are the building blocks for cortical computations such as normalization, an operation that provides divisive gain in proportion to the summed activity in a circuit (Schwartz and Simoncelli, 2001; Carandini and Heeger, 2012). In this context, the standard argument is that axons from PV form synapses onto the somata, axon initial segments, and proximal dendrites of pyramidal neurons (Kawaguchi and Kubota, 1993, 1997, 1998; Tamás et al., 1997). Because action potentials are generated near the soma, activating inhibitory conductances near the soma decreases the effective input resistance and thus divisively scales the magnitude of depolarization

evoked by a particular synaptic conductance. In contrast, axons from SOM form synapses onto dendrites, frequently the distal rather than apical dendrites (Kawaguchi and Kubota, 1997, 1998; Wang et al., 2004; Silberberg and Markram, 2007). Because dendrites are electrotonically isolated from the site of action potential generation, activating inhibitory conductances only in the dendrites will not substantially decrease the effective input resistance for more proximally delivered excitation. Because dendritic inhibition does not decrease somatic resistance, dendritic inhibition would be predicted to subtractively suppress the firing rate (Vu and Krasne, 1992; Jadi et al., 2012). However, this view has also been challenged (Gidon and Segev, 2012; Lovett-Barron et al., 2012; Mehaffey, 2005; Seybold et al., 2015).

In Chapter 3 we will examine the tuning properties of EXC and different classes of INH, in addition to their functional coupling and connectivity.

## 1.6 Reverberation

Following on from subsection 1.3.1, I wanted to not only investigate the properties and organisation of AC, but also its role in building an invariant representation of sounds (Figure 1.3B). Thus, I chose to explore the role of AC in the adaptation to reverberation. Below I give a brief description of this interesting and pervasive auditory phenomenon, providing the necessary background in order to understand Chapter 4. Here we focus on the basic principles behind reverberation and the motivation for its study, highlighting the importance of reverberation in health and disease. In the introduction to Chapter 4, we will examine the neural correlates of reverberation and the current understanding of the adaptive mechanisms in reverberation.

### 1.6.1 Reverberation fundamentals

Reverberations, or echoes, are an integral part of natural sounds<sup>6</sup>. Reverberation is the persistence of a sound after its initial production and it is caused by the reflection of the sound emitted by the source from the surfaces of nearby objects in space, such as the room walls and floor, furniture, buildings, trees in forests and people (Huisman and Attenborough, 1991; Sakai et al., 1998) (Figure 1.6A). Practically every sound

---

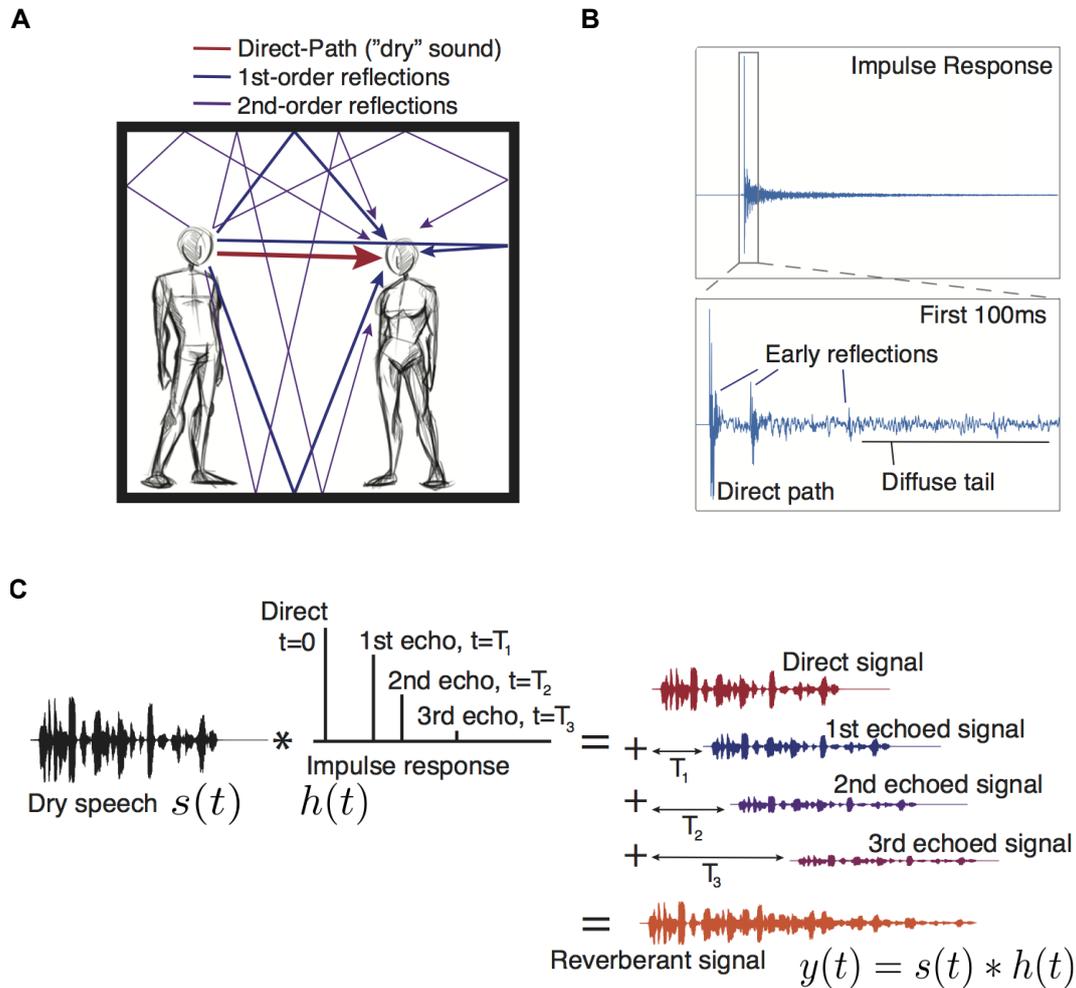
<sup>6</sup>echoes, reverberation and reverb jointly termed “reverberation” throughout the thesis.

that we hear has some amount of reverberation. Adding reverberation to virtual sounds significantly improves their perceived spaciousness and realism, and it is an important component of spatial auditory displays (Shinn-Cunningham, 2000; Trivedi et al., 2009). These reflections are attenuated (some absorption occurs at the reflecting surface), delayed (due to their increased path length), and distorted in frequency (owing to frequency-selective absorption by surfaces) copies of the original sound as shown in Figure 1.6C (Kuttruff, 2017). The physics of reverberation are well established (Schroeder, 1961; Blesser and Salter, 2009; Kuttruff, 2017). The acoustic properties of a space can be compactly described by its Impulse Response (IR), a linear filter  $h(t)$ . The reverberant sound is obtained by the convolution of the original anechoic source signal,  $s(t)$ , with the IR:  $y(t) = h(t) * s(t)$ . The IR itself can be divided into three distinct parts: (1) Direct arrival; (2) Early reflections; (3) Diffuse tail (late reverberation) (Figure 1.6B).

The first part of the IR corresponds to the direct sound path and if convolved with the anechoic sound will recover a scaled version of it. The early reflections follow the direct arrival and are highly direction and distance dependent resulting from sparse initial low-order reflections. These are followed by denser higher-order reflections which are much less structured, called the reverberant tail (late reverberation). Although the separation of early reflections from the reverberant tail is not always clear in realistic conditions, there are certain measures which aid this distinction. The amplitude of the tail can be accurately described by Gaussian statistics and follows an exponential decay profile (Schroeder, 1962; Blesser and Salter, 2009; Kuttruff, 2017). A recent extensive survey (over hundreds of IRs) of the temporal profile of IRs from natural and man-made environments found that the majority could be accurately described by an exponentially decaying Gaussian after  $\sim 50$ ms (Traer and McDermott, 2016). This is in excellent agreement with human psychophysical studies where measures setting the early-to-late boundary at 50ms seem to correlate highly with speech recognition performance (Nishiura et al., 2007). Another important observation made by Traer and McDermott (2016) is that the tail makes up the vast majority of the IR (measured as a fraction of either IR power or duration) and thus induces the majority of the distortion to a source signal. Reverberation time<sup>7</sup> ( $RT_{60}$ ) is the most commonly used metric to quantify the amount of reverberation in a given acoustic space (Schroeder, 1965).  $RT_{60}$  is the time necessary for the sound in a room to decay by 60dB relative to

---

<sup>7</sup>Commonly abbreviated as  $T_{60}$  too.



**Fig. 1.6 The basics of reverberation.**

**A**, Sound reaches a listener directly as well as via reflections off surrounding surfaces. **B**, Example room impulse response (IR) (top). Every peak corresponds to a possible propagation path; the time of the peak indicates how long it takes the reflected sound to arrive at the ear and the amplitude of the peak indicates the amplitude of the reflection, relative to that of the sound that travels directly to the ear. The first 100 ms of the IR (inset, bottom). Discrete early reflections (likely first- or second-order reflections) are typically evident in the early section of an IR, after which the reflections become densely packed in time, composing the diffuse tail. The tail makes up the vast majority of the IR (measured as a fraction of either IR power or duration) and therefore induces the majority of the distortion to a source signal. While the early reflections have a complex structure, the tail for most natural IRs can be accurately described by an exponentially decaying Gaussian. **C**, The effect of reverberation on an anechoic (“dry”) sound  $s(t)$  is described mathematically by the convolution ( $*$ ) of the sound with the IR of the environment,  $h(t)$ , thus producing the reverberant sound  $y(t)$ . The original sound is repeated, delayed, and scaled for every nonzero point in the IR and the resulting signals are summed. This process is illustrated for a schematic IR with 3 echoes. For clarity these echoes are more widely spaced than in a naturally occurring IR. Adapted from [Traer and McDermott \(2016\)](#).

the direct arrival sound. Another way of quantifying the amount of reverberation is the Direct-sound to Reverberant-sound Ratio (DRR) (Griesinger, 2009). DRR measures the power of the direct path of the IR relative to the rest. The  $RT_{60}$  measure is largely distance independent whereas the DRR strongly depends on distance, decreasing rapidly as source and receiver are separated.

## 1.6.2 Impact of reverberation

Because it is an integral part of natural and man-made environments, reverberation exerts a strong influence on the perception of sounds and auditory scene analysis in general. It impacts sound localization and sound recognition, especially that of complex sounds such as speech. The most obvious effect is that of the late reverberation, which increases the noise floor for the receiver, especially at a distance (Koenig et al., 1977). In addition, reverberation can be viewed as causing an attenuation of a signal's amplitude modulation, as the reflected energy fills the momentary dips in the anechoic signal (Figure 1.6C) (Houtgast and Steeneken, 1985). The reflections caused by reverberation follow the original sound and can substantially affect the acoustic waveform by reducing the distinctiveness of the temporal envelope and smearing the dynamic structure of sound over time. In a reverberant environment, reflected acoustic waves reach the listener from all directions, interfering with the direct sound. Under such conditions, the signals at the two ears become decorrelated (Beranek, 2004) and the ITDs, one of the main sound localization cues, fluctuates substantially (Shinn-Cunningham and Kawakyu, 2003). Thus, localization accuracy can degrade considerably in strong reverberation (Rakerd and Hartmann, 2005; Shinn-Cunningham et al., 2005). However, reverberation can be also beneficial for inferring room size and the distance to the sound source (Bronkhorst and Houtgast, 1999; Zahorik and Wightman, 2001; Cabrera et al., 2005; Brumm and Naguib, 2009). Reverberation also impacts negatively the perception of vocalizations and speech. Some of its effects on speech are to reduce its intelligibility (Knudsen, 1929; Nábělek et al., 1989; Guediche et al., 2014), limit our ability to selectively follow one voice when there are other voices in the background ("cocktail party effect") (Culling et al., 1994; Darwin and Hukin, 2000) and influence the perception of pitch in complex sounds, including speech (Sayles and Winter, 2008). The negative effects of reverberation seem to have a bigger impact on hearing-impaired people and especially those suffering from cochlear hearing loss (Humes et al., 1986; Helfer and Wilber, 1990). In addition, reverberation can pose challenges for the

effectiveness of auditory prostheses and hearing aids and substantially limit the benefits of such devices (Schweitzer, 2003; Qin and Oxenham, 2005; Poissant et al., 2006). Thus, better understanding of reverberation and mechanisms for its amelioration can help us build better hearing aids (Schweitzer, 2003).

It is also noteworthy that the two parts of the IR (early reflections and tail, Figure 1.6B) can have different effects with regards to reverberation and its perception. In terms of speech processing, early reflections have largely positive role by increasing the level of the signal at the ear, and boosting the effective ratio of signal-to-noise components in the sound (Nábélek, 1978; Bradley et al., 2003). In contrast, the reverberant tail is seen as largely having a negative impact, causing attenuation in amplitude modulation, increasing the noise floor and overall temporal smearing (Koenig et al., 1977; Houtgast and Steeneken, 1985). Cochlear implant users do not derive the same benefit from early reflections and are more severely affected by the reverberant tail (Hu and Kokkinakis, 2014). In directional sound localization, both early and late reflections tend to degrade localization cues although very short early reflections might be beneficial under certain circumstances (Gourévitch and Brette, 2012). Interestingly, Traer and McDermott (2016) found that human listeners cannot distinguish between the same IR with and without the early reflections, arguing that the majority of the acoustic and perceptual distortion is due to the tail.

## 1.7 Thesis overview and structure

In this thesis, I explore the organisation, structure and function of AC in the ferret. Although my main focus was on the study of the properties of ferret AC, I also spent significant amount of time establishing certain experimental techniques necessary for the execution of my research. Indeed, the start of my graduate studies in Prof King's and Dr Walker's groups coincided with the commencement of the novel high-density microelectrode ("Neuropixels") electrophysiological experiments in the lab, at a time, October 2016, when only a few dozen groups in the world had access to this novel technology. I spearheaded this effort in addition to establishing immunohistochemistry and image registration and reconstruction approaches for the study of inhibitory neurons in the ferret. Note, that there is no "General Methods" chapter since the individual "Results" chapters provide a detailed enough account of the various methods that have been used. Results Chapters 2 to 4 progress through investigations of the

research questions that I have undertaken throughout my DPhil studies, followed by a general discussion in Chapter 5. The structure is the following:

- **Chapter 2: Tonotopic organisation of auditory cortex.**

This first results chapter explores the question of how the ferret AC is organised with regards to its representation of auditory frequency space (“tonotopy”). In addition, an interspecies comparison is made between ferrets and mice in order to establish whether the more loose “salt and pepper” local tonotopic organisation of AC is a rodent peculiarity or a more general principle of cortical organisation. Finally, we attempt to reconcile some of the previous methodological discrepancies in the literature. We pursue these aims using two-photon calcium imaging in the auditory cortices of anaesthetised ferrets and mice, and by using high-density “Neuropixels” microelectrodes while presenting pure tone sound stimuli.

- **Chapter 3: Functional properties of inhibitory neurons in auditory cortex.**

In this next results chapter, we continue investigating the micro-circuitry of ferret AC. Here the emphasis is on the tuning properties of different classes of inhibitory neurons within ferret AC, and we compare them to those of excitatory neurons. We also investigate how these neuronal classes interact with one another using “signal” and “noise” correlation analyses. Similar to Chapter 2, this was achieved by presenting pure tone sound stimuli to anaesthetised ferrets and using 2-photon calcium imaging and microelectrode recordings to measure cortical responses. Given the lack of transgenic ferrets lines, it was necessary to first develop reliable electrophysiological and 2-photon calcium imaging procedures and analysis pipelines to separate the different types of neurons. Thus, I also describe the results of this process, going into detail behind the experimental and data analysis procedures involved. This is then followed by a direct comparison of the tuning properties of excitatory and inhibitory neurons in ferret AC and some preliminary analyses regarding their interactions.

- **Chapter 4: Adaptation to reverberation in auditory cortex.**

While the previous two chapters investigate the organization of ferret AC and its more basic tuning properties, in this final chapter we explore how AC contributes to the robust representation of more complex natural sounds. Specifically, we ask the questions: How does the the AC built an invariant representation to natural

sounds in realistic acoustic conditions where reverberation is present? What are the optimal linear filters that can “clean” reverberant sounds from echo, what properties do they have and do they resemble those of auditory cortical neurons? To achieve this, we first trained a linear model to retrieve “clean” anechoic sounds from versions which are corrupted by reverberation. We built our model in a way such that the model “units” can be directly compared to the properties of auditory cortical neurons using their spectro-temporal receptive fields. We used virtual acoustic space stimuli to present natural sounds in reverberant conditions to anaesthetised ferrets and recorded the spiking responses of auditory cortical neurons using microelectrodes. We then used this data to fit receptive fields and compare them to the linear model.

- **Chapter 5: General discussion.**  
General discussion and conclusions of the thesis.



# Chapter 2

## Tonotopic organisation of auditory cortex

“You can’t use an old map to explore a new world.”

---

*Albert Einstein*

The work presented in this Chapter has been published as: Gaucher Q, Panniello M, Ivanov AZ, Dahmen JC, King AJ, Walker KMM. 2020. **Complexity of frequency receptive fields predicts tonotopic variability across species.** eLife 9:e53462.

### 2.1 Introduction

#### 2.1.1 Global tonotopy but local heterogeneity

Early electrophysiological investigations revealed a tonotopic arrangement of neuronal frequency preferences in A1 in all species studied, including monkeys ([Merzenich and Brugge, 1973](#)), cats ([Hind, 1953](#)), ferrets ([Kelly et al., 1986a](#)), gerbils ([Steffen et al., 1988](#)), and rats ([Sally and Kelly, 1988](#)). More recent multielectrode recordings ([Imaizumi et al., 2004](#); [Bizley et al., 2005](#); [Recanzone et al., 1999](#)) and large-scale imaging experiments ([Nelken et al., 2008](#)) have confirmed the presence of a tonotopic gradient over large areas of A1 in carnivores and primates, while single-neuron recordings

in awake animals have reported clear frequency gradients in marmosets (Bendor and Wang, 2008) and weaker tonotopy in cats (Goldstein et al., 1970).

In vivo two-photon calcium imaging can measure the stimulus preferences of a large number of individual neurons within a cortical region compared to the more indiscriminate sampling of the above techniques, improving the spatial sampling of mapping studies. The results of such experiments in mouse A1 over the past decade have challenged our understanding of the cortical tonotopic map. While previous electrophysiological and large-scale imaging studies confirmed an overall tonotopic organization in mouse A1 (Hackett et al., 2011; Guo et al., 2012; Stiebler et al., 1997), more recent two-photon imaging studies have reported different degrees of frequency tuning variability among neighbouring neurons, from well-ordered local maps (Issa et al., 2014) to moderately heterogeneous tuning (Bandyopadhyay et al., 2010; Rothschild et al., 2010; Panniello et al., 2018; Romero et al., 2019). Similarly, two-photon imaging has revealed local heterogeneity in whisker selectivity in mouse primary somatosensory cortex (Kerr et al., 2007; Sato et al., 2007), and in orientation tuning in mouse V1 (Bonin et al., 2011).

### 2.1.2 Species and methodological considerations

There are several possible explanations for the local variation of A1 frequency preferences revealed by two-photon imaging studies compared to the smoother tonotopy described in electrophysiological studies. First, the heterogeneity may be due to the higher spatial sampling rate of two-photon imaging. Microelectrode studies are often based on responses that are summed across several neurons, and this spatial averaging may lead to smoother tonotopic gradients (Guo et al., 2012). Even in single neuron recordings, the spacing between cells is at least  $50\mu\text{m}$  (South and Weinberger, 1995), and typically  $\sim 100\mu\text{m}$  or greater, so variations in stimulus preferences between neighbouring cells are not usually examined (Kanold et al., 2014). Secondly, multielectrode recordings may be biased more towards the most robustly responding neurons in thalamorecipient layers, whereas most two-photon imaging studies have been restricted to the superficial layers of the cortex. Studies investigating layers 2/3 and 4 have usually found smoother tonotopy in the deeper layers (Guo et al., 2012; Winkowski and Kanold, 2013) (but see Tischbirek et al. (2019) reporting similar tonotopic organization across all layers of mouse AC).

The variations in tonotopy across studies may also partly reflect a species difference in the cortical organization of rodents and higher mammals, particularly as local thalamic inputs to A1 in mice are also heterogeneous in their frequency tuning (Vasquez-Lopez et al., 2017). Two-photon imaging studies of primary visual cortex have revealed poorer spatial organization of orientation tuning in rodents than in cats (Bonin et al., 2011; Ohki et al., 2005), tree shrews (Lee et al., 2016), and ferrets (Wilson et al., 2017). The same may be true of tonotopy in A1. This view is supported by a recent study in marmosets, in which A1 was reported to be more tonotopically organized than in rats (Zeng et al., 2019).

In addition, it is not clear how spatial organization of tuning to a single sound frequency coincides with the known functions and complex frequency receptive fields of AC. Throughout the ascending auditory pathway, neurons progressively integrate spectral and temporal features of sound (Linden and Schreiner, 2003), and the receptive fields of many A1 neurons are poorly predicted by a model of linear tuning to a single sound frequency (Ahrens et al., 2008; Sadagopan and Wang, 2009). Recent studies have shown that the preferred frequencies of A1 neurons with irregular tuning curves are poorly mapped (Romero et al., 2019; Tischbirek et al., 2019).

### 2.1.3 Summary of the study

The study in this chapter is the first application of 2-photon calcium imaging and Neuropixels recordings to study auditory processing in ferrets. The present experiments further examine whether local heterogeneity in tonotopy is a general feature of mammalian A1, or a peculiarity of rodents. We conduct 2-photon imaging experiments in mice and ferrets, and compare their local tonotopic organization. Here we show that local heterogeneity of frequency preferences is not unique to rodents. Using two-photon calcium imaging in layers 2/3, we found that local variance in frequency preferences is equivalent in ferrets and mice. Neurons with multi-peaked frequency tuning are less spatially organized than those tuned to a single frequency in both species. Our present results further demonstrate how A1 neurons with multi-peaked frequency tuning curves impact on tonotopic organization. Furthermore, we show that microelectrode recordings may describe a smoother tonotopic arrangement due to a sampling bias towards neurons with simple frequency tuning. These results help explain previous inconsistencies in cortical topography across species and recording techniques.

## 2.2 Materials and methods

### 2.2.1 Animals

All animal procedures were approved by the local ethical review committee of the University of Oxford and performed under license from the UK Home Office (PPL license number: PC27AFFA2). Eight female ferrets (*Mustela putorius furo*; Marshall BioResources, UK) and 11 female mice (C57BL/6J; Harlan Laboratories, UK) were used in the two-photon imaging experiments, and six ferrets (two male) were used in the electrophysiology experiments.

### 2.2.2 Viral vector injections

#### Ferret surgery

At age 8–12 weeks, ferrets were put under general anesthesia with an intramuscular injection of ketamine (Vetalar; 5 mg/kg) and medetomidine (Domitor; 0.02 mg/kg), and medicated with buprenorphine (Vetergesic; 0.02 mg/kg i.m.), atropine (Atrocare; 0.06 mg/kg i.m.) and meloxicam (Metacam; 0.2 mg/kg i.m.). Ferrets were then intubated and artificially ventilated. A mixture of oxygen and isoflurane (IsoFlo; 0.5 – 2%) was continuously delivered throughout the surgery to maintain anesthesia. Respiratory rate, end-tidal CO<sub>2</sub>, electrocardiogram, blood pressure, and blood oxygenation were continuously monitored, and body temperature was maintained at 36–38°C (3MTM Bair Hugger™). The eyes were lubricated (Maxitrol, Alcon) to prevent corneal desiccation. An intravenous cannula was inserted to deliver Hartmann’s solution (54ml/kg/hr) continuously, a single dose of co-amoxiclav (Augmentin, 20mg/kg), and medetomidine (Domitor; 0.027 mg/kg i.m.) as required, throughout the surgery. After placing the ferrets in a stereotaxic frame (Model 900LS, David Kopf Instruments), the scalp was cleaned (ChlorPrep®; 2% chlorhexidine gluconate), bupivacaine (Marcain; 2mg/kg, s.c.) was injected into the scalp, the scalp was incised, and the right temporal muscle was retracted. A craniotomy and durotomy were carried out to expose the primary and secondary AC, based on stereotaxic coordinates (11 mm ventral to the midline and 8 mm frontal to Lambda) and visual confirmation of the ectosylvian gyrus.

Injections were carried out using a glass pipette and a custom-made pressure injection system. The viral vector consisted of a 1:1 solution of AAV1.Syn.GCaMP6m.WPRE.SV40

(Penn Vector Core) and Phosphate-Buffered Saline (PBS) (Sigma Aldrich) for ferrets 1-4, and a 1:1 solution of AAV1.Syn.GCaMP6f.WPRE.SV40 (Addgene) and AAV1.mDlx.GFP-GCaMP6f-Fishell-2.WPRE.SV40 (Penn Vector Core) for ferrets 5-8. Approximately 260 nl of the viral vector was slowly injected at depths of  $\sim$ 800, 600, 400 and  $200\mu\text{m}$  below the pial surface in ferrets 1-4 and at depths of  $200\mu\text{m}$  and  $400\mu\text{m}$  in ferrets 5-8. For each ferret, the viral vector solution was injected into 4-9 sites across primary and secondary AC.

The bone was then replaced over the craniotomy, the muscle, fascia and skin were sutured closed, and the animal was allowed to recover. Dexamethasone (0.5 mg/kg i.m.) and buprenorphine (0.01 mg/kg i.m.) were administered immediately after surgery. Meloxicam (0.2 mg/kg oral) and co-amoxiclav (20 mg/kg s.c.) were administered daily for 5 days, and buprenorphine (0.01 mg/kg i.m.) for two days.

### Mouse surgery

Mouse experiments were previously reported in [Panniello et al. \(2018\)](#). At age 5-6 weeks, mice were premedicated with dexamethasone (Dexadreson;  $4\mu\text{g}$  s.c.), atropine (Atrocare;  $1\mu\text{g}$  s.c.) and carprofen (Rimadyl;  $0.15\mu\text{g}$  s.c.), and put under general anesthesia with fentanyl (Sublimaze; 0.05 mg/kg i.p.), midazolam (Hypnovel; 5 mg/kg i.p.), and medetomidine hydrochloride (Domitor; 0.5 mg/kg i.p.). The mouse was then placed in a stereotaxic frame (Model 900LS, David Kopf Instruments) and maintained at  $36\text{-}37^\circ\text{C}$  body temperature throughout surgery (DC Temperature Controller, FHC). The scalp was incised, the temporal muscle retracted, and A1 was located using stereotaxic coordinates (70% of the distance from Bregma to Lambda, and 4.5mm lateral from the midline). Two small holes ( $\sim$ 0.4mm diameter), separated rostrocaudally by  $\sim$ 0.5 mm, were drilled over the right A1. A total of  $\sim$ 200nl of the viral construct AAV1.Syn.GCaMP6m.WPRE.SV40 (Penn Vector Core), diluted (1:2) in PBS, was injected at each site, spread equally across four depths spanning  $50\text{-}400\mu\text{m}$  below the pial surface. After injection, the skin was sutured and general anesthesia was reversed with flumazenil (Anexate; 0.5 mg/kg s.c.) and atipamezol (Antisedan; 2.5 mg/kg s.c.). Postoperative buprenorphine (Vetergesic; 1 ml/kg s.c.) and enrofloxacin (Baytril; 2 ml/kg s.c.) were administered immediately after surgery and again 24 hours later. To confirm that we were targeting mouse A1 with our viral vector injections, the coordinates used matched those of fluorescent retrobead injections that retrogradely labelled neurons in the ventral division of the medial geniculate body ([Panniello et al., 2018](#)). These injection sites also correspond to the location of A1 determined

by imaging the full tonotopic gradient over larger cortical areas of transgenic mice expressing Genetically encoded calcium indicator (GCaMP) 6f in cortical neurons under the CaMKII promoter (Panniello et al., 2018). Nevertheless, it is not possible to say with the same degree of confidence as for the ferret data that all mouse imaging reported in the present study was performed in A1.

### 2.2.3 In vivo two-photon calcium imaging

#### Ferret surgery

*In vivo* two-photon imaging was performed 3–6 weeks after viral injection. General anaesthesia was induced with an intramuscular injection of ketamine (Vetalar; 5 mg/kg) and medetomidine (Domitor; 0.02 mg/kg), and was maintained with a continuous intravenous infusion of these two drugs in Hartmann’s solution with 3.5% glucose and dexamethasone (0.5 mg/ml/hr). The animal was intubated and artificially ventilated. Respiratory rate, end-tidal CO<sub>2</sub>, electrocardiogram and blood oxygenation were continuously monitored throughout the imaging session. Eye ointment (Maxitrol; Alcon, UK) was applied throughout and body temperature was maintained at 36–38°C. Atropine (Atrocare; 0.06 mg/kg i.m.) was administered every 6 hours, or when bradycardia or arrhythmia was observed. Ferrets were placed in a custom-built stereotaxic frame and head stability was achieved using ear bars and a mouthpiece. After shaving the scalp and injecting bupivacaine (Marcain, 2mg/kg s.c.), the skin was incised and the temporal muscle removed. A steel holding bar was secured to the skull using dental cement (SuperBond; C&B, UK) and a stainless steel bone screw (Veterinary Instrumentation, UK). A circular craniotomy 10mm in diameter was drilled over the injection site, and any dura that had regrown over the AC was removed. A custom made titanium ring (10mm diameter) containing an 8mm glass coverslip (Thermo Fisher Scientific, UK) was inserted in the craniotomy and secured to the skull with dental cement. Ear bars were removed, and the ferret was placed under the microscope for imaging.

#### Mouse surgery

*In vivo* two-photon imaging was performed 3–6 weeks after viral injection. All mice were imaged at <12 weeks of age, before the development of high frequency hearing loss (Ison et al., 2007). General anaesthesia was induced with an intraperitoneal injection of ketamine (Vetalar; 100 mg/kg) and medetomidine (Domitor; 0.14 mg/kg), and was maintained with hourly subcutaneous injections of both agents (50 mg/kg/h Vetalar

and 0.07 mg/kg/h Domitor). Body temperature was maintained at 37–38°C with a heating pad. The mouse was moved to a stereotaxic frame (Model 900LS, David Kopf Instruments), the scalp was incised, the temporal muscle retracted, and a craniotomy of ~2.5mm diameter was performed over the injection sites in the right AC. The exposed area was covered with a glass coverslip, which was secured to the skull with cyanoacrylate adhesive (UltraGel; Pattex, DE). A steel holding post was attached to the left side of the skull using dental cement (UniFast Trad, GC Dental Products Corporation), ear bars were removed, and the mouse was placed under the microscope in the stereotaxic frame.

### Two-photon imaging

Imaging of calcium transients was performed using a B-Scope two-photon microscope (Thorlabs, Inc., UK) controlled by ScanImage 4.1 software (<http://scanimage.org>). Excitation light was emitted by a Mai-Tai eHP laser (SpectraPhysics, UK; 70 fs pulse width, 80MHz repetition rate) tuned to 930 nm. The beam was directed into a Conoptics modulator (laser power, as measured under the objective, was 15–30 mW) and scanned through an 8kHz resonant scanner in the x-plane and a galvanometric scanning mirror in the y-plane. The resonant scanner was used in bidirectional mode, at a resolution of  $512 \times 512$  pixels, allowing us to acquire frames at a rate of ~30Hz for our most common zoom. A 16X/0.80W LWD immersion objective (Nikon, UK) was used, and emitted photons were guided through a 525/50 filter onto GaAsP photomultipliers (Hamamatsu Photonics, Japan). Neuronal fields were between  $200 \times 200\mu\text{m}$  and  $300 \times 300\mu\text{m}$  in size. Neuronal activity was imaged in ferrets at  $176 \pm 26.83\mu\text{m}$  (median  $\pm$  s.d.) below cortical surface. In mice, imaging was performed at  $216 \pm 34.21\mu\text{m}$  below pial surface, corresponding to layer 2/3 in this species (Anderson et al., 2009; Dahmen et al., 2008). In both ferrets and mice, imaging was performed in the right hemisphere (right A1).

### 2.2.4 Sound presentation

Pure tones were generated via Matlab (MathWorks, Inc., USA), and an RZ6 multiprocessor (Tucker-Davis Technologies, USA) was used to synchronize the sound presentation with the microscope scanning.

Sound stimuli were presented binaurally to the animal via a customized closed acoustic delivery system comprised of either two Tucker-Davis Technologies EC1

electrostatic speakers (11 mice and three ferrets) or two Panasonic RPHV297 earphones (5 ferrets). Speakers were coupled to a 12-cm-long silicone tube leading into the ear canal. The output response of the speakers was measured using a Brüel & Kjær calibration system with a GRAS 40DP microphone coupled to the end of the silicone tube. An inverse filter was applied to the speaker output to produce a flat spectral response ( $\pm 3$ dB) over the tone frequency range. Sound intensity was calibrated with an Iso-Tech TES-1356-G sound level calibrator.

The microscope and experimental animals were enclosed within a sound- and light-attenuating box. The ambient room noise was  $<42$ dB Sound Pressure Level (SPL) inside this box, and primarily consisted of energy at  $<200$ Hz. The resonant scanner generated a constant acoustical tone of 8kHz during imaging that was  $<30$ dB SPL near the animal's head.

For each imaging field in the ferret experiments, 10-20 repetitions of pure tones were presented at frequencies with 0.25 octave spacing across the ferret hearing range, and levels of 30-90dB SPL, with 20dB spacing (50 to 125 ms onset and offset cosine ramps). Frequency/level combinations were presented in pseudorandom order. Details of stimulus parameters for each animal are given in Table 1. For mice, ten repetitions of 100-ms duration (5 ms onset and offset cosine ramps) pure tones were presented at 18 frequencies (1.9–50kHz, 0.6 octave spacing) and 4 levels (40, 60, 80, and 100 dB SPL), at a rate of 0.65Hz.

**Table 2.1 Details of stimuli presented to ferrets**

ferret	stim duration (ms)	repetitions	ISI (s)	frequency range (Hz)
1	100	10	0.75	1259-41687
2	500	12	1.50	1047-28840
3	100	10	1.33	1259-41687
4	100	10	1.33	1259-41687
5	500	12	1.50	346-28840
6	500	20	1.50	346-28840
7	500	12	1.50	1047-28840
8	500	12	1.50	1047-28840

### 2.2.5 Histology

At the end of each imaging session, experimental animals were overdosed (mice: 100mg/kg ketamine and 0.14mg/kg medetomidine, i.p.; ferrets: Euthatal, 1ml pentobarbital sodium, i.p.) and perfused transcardially, first with 0.01M PBS and heparin (20 units/ml), and then with 4% paraformaldehyde in PBS. Brains were removed and placed in 4% paraformaldehyde for two hours, after which they were stored in PBS with 0.01% sodium azide. Ferret brains were then cryoprotected in a 1:3 solution of sucrose and PBS for 24 hours. Sagittal brain sections (50 $\mu$ m thickness) were obtained using a freezing sliding microtome (Leitz Wetzlar). Sections were sliced parallel to the orientation of the *in vivo* cranial window to facilitate the reconstruction of imaged neurons. Sections were washed three times in PBS, after which they were mounted onto microscope slides using Vectashield mounting medium (Vector Laboratories Ltd., USA). A coverslip was placed on the slide for imaging and sealed with clear nail polish.

### 2.2.6 Confocal imaging

Brain sections were imaged using an inverted Olympus FV3000 six laser line spectral confocal microscope fitted with high sensitivity gallium arsenide phosphide (GaAsP detectors) and a 4x, 0.16 NA UplanSApo objective. The confocal pinhole was set to one airy unit to optimize optical sectioning with emission collection. Images were collected in resonant scanning mode at 512x512 pixels (pixel size 6.21 $\mu$ m) and 16x averaging. Tile scans were stitched using the Olympus FluoView software.

### 2.2.7 Identifying the location of GCaMP6 injection sites

A tracing of the ectosylvian and pseudosylvian sulci in ferret 7 was used as a template of AC, and imaging fields from all 8 ferrets were transferred onto this template. In ferrets 1, 5, 6 and 7, 1mm x 1mm images of the brain surface were acquired *in vivo* at the beginning of the two-photon imaging session, across the entire craniotomy. These images were then tiled together to reconstruct the surface of AC and locate the imaged fields. The tiled AC was then aligned with boundaries of the sulci in the auditory cortical template. For ferrets 2, 3, 4 and 8, GCaMP6 injection sites viewed on confocal images of the ferret ectosylvian gyrus (above) were aligned with the two-photon imaging fields acquired in the same ferret to precisely determine the location of each

imaged field on the gyrus. The confocal images were then aligned to the sulci in the template map of the ectosylvian gyrus (Figure 2.16), using custom Matlab scripts. These procedures allowed us to plot the coordinates of neurons imaged in different animals onto a common map of AC.

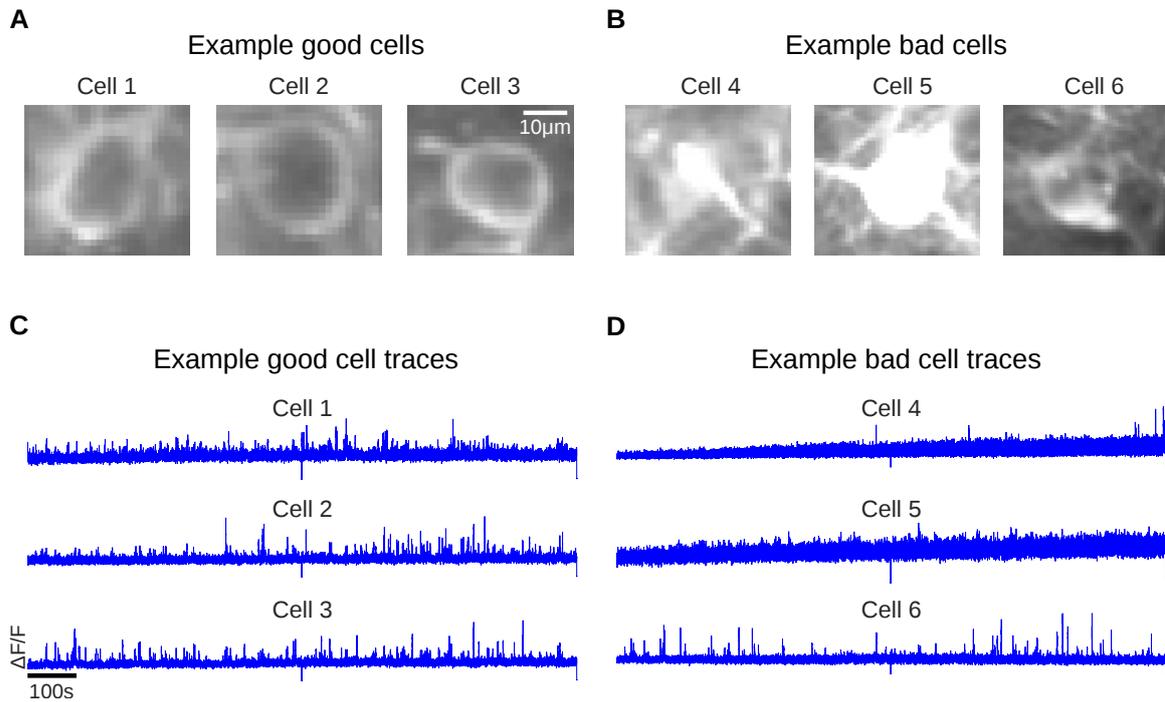
## 2.2.8 Data analysis for two-photon imaging experiments

### Isolating the responses of single neurons from imaging fields

To isolate the responses of single neurons from imaging fields, we analysed the responses of 3604 neurons imaged in 32 imaging fields across 8 ferrets, and 1962 neurons imaged in 42 imaging fields across 11 mice.

Videos of the imaging field during tone presentation were imported into Suite2p software (<https://github.com/MouseLand/suite2p>), which automatically performs mechanical drift correction, cell detection, and neuronal and neuropil trace extraction, and spike deconvolution. Parameters for the estimation of calcium transient templates were set separately for GCaMP6m and GCaMP6f, as the rise and decay times differ for these two indicators. These parameters were optimized using our own dataset. The most relevant parameter was the half decay time of the indicator (0.7s for GCaMP6f and 1.25s for GCaMP6m).

We manually inspected all regions of interest automatically detected by the Suite2p built-in classifier, in order to confirm if they were individual healthy neurons. This assessment was based on the frame-averaged image (e.g. the candidate neuron had a clear ring of fluorescence and was dark in the center, see Figure 2.1A), the alignment of the region of interest, e.g. the region did not expand beyond the cell membrane (Figure 2.1B Cell 4) or include other neurons (Figure 2.1B Cell 6), cells that were not over-expressing the indicator (Figure 2.1B Cell 5) and the activity of the calcium trace (e.g. no evidence of “bursting” that characterizes cell death). Healthy cells exhibited clear calcium transients (Figure 2.1C), while many of the bad cells had calcium traces that showed little modulation by the stimuli (Figure 2.1D). For each confirmed neuron, time series of the neuronal calcium trace ( $\Delta F/F_0$ ), neuropil calcium trace, and deconvolved spike probability were exported to Matlab for further analysis. Deconvolving the calcium trace allowed us to control for differences in the dynamics of the two calcium sensors used (GCaMP6m and GCaMP6f), and deconvolution parameters



**Fig. 2.1 Selecting cells for downstream analysis.**

**A**, Example healthy (“good”) cells (Cells 1-3) from one animal expressing GCaMP6f, characterised by their donut shape, i.e. the indicator is expressed mainly in the cytosol and not the nucleus of the cell. In addition, the whole cell is clearly visible in the field of view. **B**, Example unhealthy (“bad”) cells (Cells 4-6) from the same imaging area as **A**. These cells were excluded based on only part of the soma being visible (Cell 4), the cell being too full of indicator (Cell 5) or having two cells on top of one another (Cell 6). **C**, Example calcium traces ( $\Delta F/F$ ) from the 3 cells in **A**. Clear calcium transients are visible. **D**, Example calcium traces from the 3 cells in **B**. Notice the lack of clear transients (Cell 4 and 5). Cell 6 displays activity but here the signal comes from two cells on top of one another.

(e.g. half-decay constant) were adapted for each indicator, as described by Pachitariu et al. (2018).

### Identifying frequency sensitive neurons

To identify frequency sensitive neurons, neuronal responses to sound presentation were quantified within a time window starting at stimulus onset and lasting twice as long as the stimulus. The “evoked activity” was defined as the average of the inferred spike probability trace within the response window. A two-way Analysis of variance (ANOVA), with tone frequency and sound level as predictors, was used to determine if the evoked activity was significantly modulated by sound frequency or intensity ( $\alpha$

= 0.05). Neurons showing a significant main effect of frequency or frequency/level interaction were defined as “frequency sensitive”, and only these neurons were included in further analyses.

### **Calculating best frequency and classifying the frequency response**

To calculate best frequency and classify the frequency responses, FRA plots were constructed for a given neuron from the trial-averaged evoked activity in response to tones presented at each frequency/level combination. For visualization, FRA plots were smoothed using a two dimensional three-point Gaussian kernel.

Frequency profiles of neurons were calculated by averaging the FRA across all sound levels. The frequency eliciting the highest response in the frequency profiles was defined as the BF, as in previous studies (Guo et al., 2012; Panniello et al., 2018; Barnstedt et al., 2015). The FRA bandwidth, expressed in octaves, was defined as the continuous range of frequencies around BF that elicited a response greater than 50% of its maximal (i.e. BF) response.

Neurons were automatically categorized into three classes based on their frequency-response profiles. A neuron was defined as “single-peaked” if its frequency-response profile contained only one continuous region above threshold, where threshold was defined as 75% of the maximal response across all frequencies. “Double-peaked” neurons were those with two distinct regions above threshold in their frequency-response profile. For these FRAs, the frequency eliciting the strongest response was defined as  $BF_d$ , while the frequency eliciting the highest response in the other response region was defined as peak 2. A neuron was defined as “complex” if it showed more than two discontinuous regions above threshold in the frequency-response profile. We also visually inspected FRAs (both smoothed and unsmoothed) of neurons and manually classified them as single-peaked, double-peaked or complex. The classification of 83% of neurons was the same under the automated frequency profile or manual FRA classification procedures. For the remaining 17% of neurons, the final classification was based on visual inspection of the FRA and frequency-response profile together.

### **Trial-to-trial reliability**

The trial-to trial reliability of responses at BF was estimated for each neuron as the Fano Factor (FF) i.e. the variance in the evoked response to the BF across trials, divided by the average evoked activity at BF.

### **Quantification of local BF variability**

To assess the variability of frequency tuning within each imaging field, we computed the difference (in octaves) between each BF and the average BF of every other neuron in the imaging field, an approach similar to a recent study (Tischbirek et al., 2019). This analysis was performed independently on neurons from each of the three FRA classes, as well as on the three classes combined (i.e. all frequency-sensitive neurons). In this analysis, we included only imaging fields containing at least three single-peaked, three double-peaked and three complex neurons.

To compare the degree of local BF variability between mice and ferrets, it was necessary to account for differences in their hearing range and the length of their tonotopic gradient in A1. The mouse A1 is  $\sim 1$  mm in length along the tonotopic gradient (Guo et al., 2012; Stiebler et al., 1997) and the hearing range of C57BL/6J mice at 10 weeks is 2-60kHz (Heffner and Heffner, 2007; Ison et al., 2007), giving an A1 tonotopic map slope of 4.9 octaves/mm. In ferrets, the tonotopic gradient spans  $\sim 3.5$ mm in A1 (Kelly et al., 1986a; Bizley et al., 2005) and the hearing range is  $\sim 36$ Hz – 44kHz (Kelly et al., 1986b), so the tonotopic gradient in A1 has a slope of 2.9 octaves/mm. For each imaging field, we first found the average BF across all the relevant neurons within the field (i.e. all neurons of a given FRA class). The expected BF range was then calculated around this average BF, given the tonotopic slopes above and the size of the imaging field. The percentage of neurons having a BF within the expected range was used as a metric of variability of local frequency tuning within the field for that neuronal class.

### Signal and noise correlations

In keeping with previous studies (Rothschild et al., 2010; Panniello et al., 2018), signal and noise correlations were computed between the evoked responses of pairs of frequency-sensitive neurons recorded simultaneously. Briefly, noise correlations were estimated by first normalizing each neuron's response to sounds presented on individual trials by its signal response to that frequency/level combination. This normalization was obtained by subtracting, from each single trial response, the average response to all tones presented at that frequency/level combination. The noise correlation was then calculated across the normalized trial-by-trial responses of two cells.

Signal correlations were calculated as the correlation between the trial-averaged responses of any two neurons, minus the noise correlation computed for the neuronal pair.

### Analysing the effects of neuropil contamination

We also analysed the effects of neuropil contamination. Neuronal calcium traces extracted from two-photon imaging acquisitions can be contaminated by the fluorescent signal coming from the pixels immediately surrounding each soma. For this reason, it is common practice to subtract from the neuronal trace what is known as the neuropil signal (Chen et al., 2013).

To test these effects, we calculated the BF of each neuron with and without neuropil subtraction, and the results are presented in Figure 2.19. The results were examined separately for the three FRA classes of neurons, and only for neurons that were frequency-sensitive (two-way ANOVA, described above), both with and without neuropil correction. The presence of the neuropil signal substantially changed the BF of some neurons (Figure 2.19A). Removing the neuropil signal caused a smaller change in the BF of single-peaked neurons (BFs;  $0.48 \pm 0.05$  octaves; mean  $\pm$  Standard Error of the Mean (SEM)) than the BF of double-peaked neurons (BFd;  $0.76 \pm 0.07$  octaves; t-test:  $t = 3.36$ ,  $p = 8.4 \times 10^{-4}$ ), peak 2 of double-peaked neurons ( $1.42 \pm 0.09$  octaves;  $t = 9.97$ ,  $p = 2.3 \times 10^{-21}$ ), and the BF of complex neurons (BFc;  $1.08 \pm 0.11$  octaves;  $t = 5.94$ ,  $p = 6.2 \times 10^{-9}$ ). There was a trend for the local variance in BF within an imaging field to be higher for double-peaked and complex neurons when the neuropil was subtracted, but these trends were not statistically significant (Figure 2.19B). When the BFs of “neuropil contaminated” neurons were mapped onto the common template of A1, tonotopic organization was observed for single-peaked, double-peaked and complex cells (Figure 2.19C). Unlike in the neuropil-corrected signals (Figure 2.7A3, B3, C3), the second frequency peak (peak 2) of double-peaked neurons also showed tonotopic organization when the neurons contained neuropil contamination (Figure 2.19C3, D3).

## 2.2.9 In vivo electrophysiology

### Data acquisition

The animal preparation and anesthesia protocol was identical to the in vivo two-photon calcium imaging procedures described above. Recordings were carried out in the left AC. An Ag/AgCl external reference wire was inserted between the dura and the skull from the edge of craniotomy. After durotomy, the brain surface was covered with a solution of 1.25% agarose in 0.9% NaCl, and silicon oil was applied to the craniotomy regularly throughout recording.

A Neuropixels Phase 3a probe28 was inserted orthogonally to the brain surface through the entire depth of AC. Data were acquired at a 30kHz sampling rate using SpikeGLX software (<https://github.com/billkarsh/SpikeGLX>) and custom Matlab scripts.

### Sound presentation

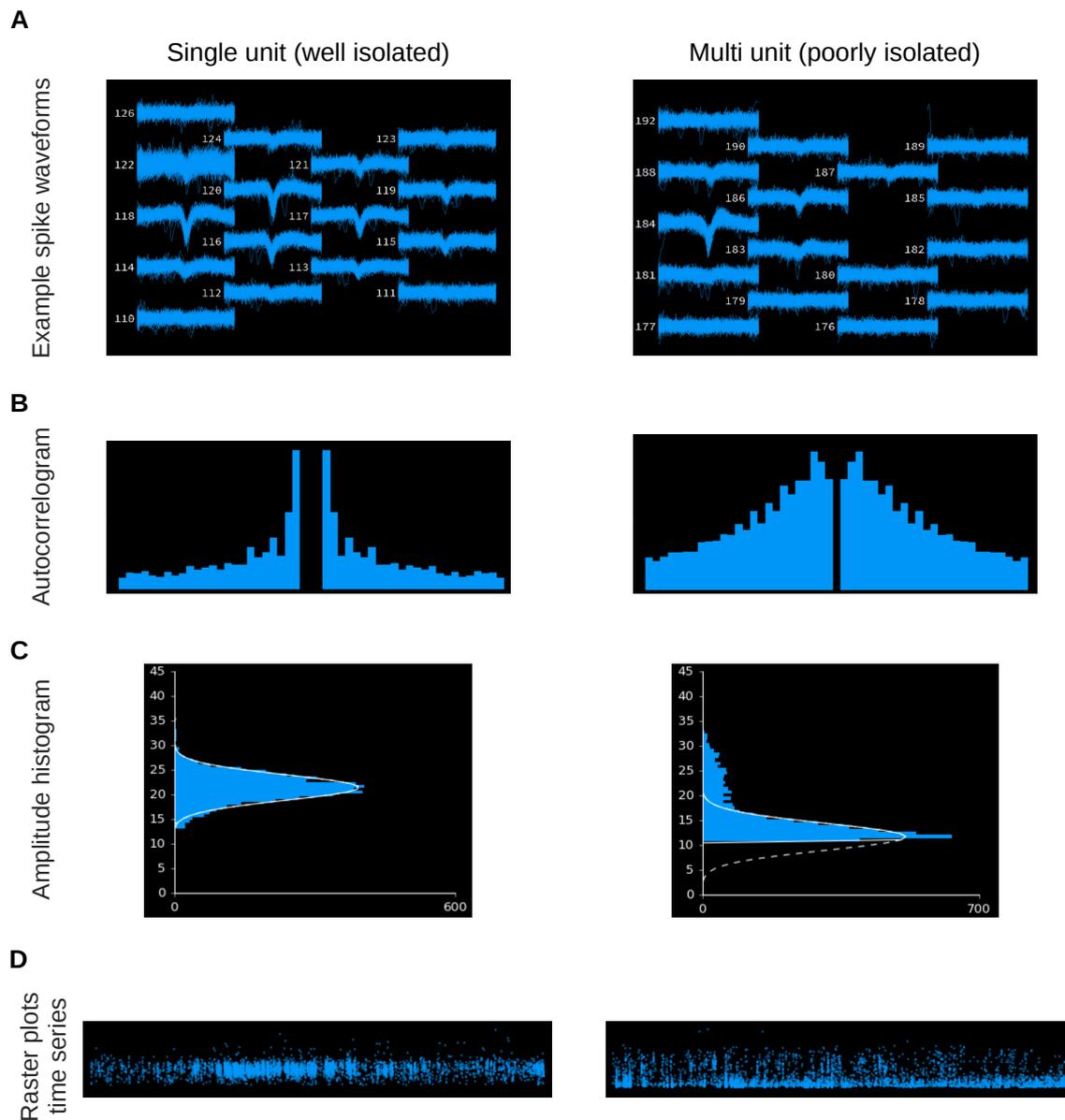
Electrophysiological recordings were made in a custom-built anechoic chamber. Stimuli were presented binaurally via Panasonic RP-HV094E-K earphone drivers, coupled to otoscope speculae inserted into each ear canal, and driven by a System 3 RP2.1 multiprocessor and headphone amplifier (Tucker-Davis Technologies). The speculae were sealed in place with Otoform (GmbH). Speaker calibration was performed as described above for imaging experiments. Pure tones (0.5-40kHz, 0.45 octave spacing, 110 ms duration, 5 ms cosine onset and offset ramps) were presented at 5 intensity levels (40-80dB SPL). Each frequency/intensity combination was presented for 20 repetitions, in pseudorandomized order, at a rate 1.37Hz.

### Spike sorting

The recorded signal was processed offline by first digitally highpass filtering at 150Hz. Common average referencing was performed to remove noise across electrode channels (Ludwig et al., 2009). Spiking activity was then detected and clustered using Kilosort2 software (<https://github.com/MouseLand/Kilosort2>) (Pachitariu et al., 2016). Responses from single neurons were manually curated using Phy (<https://github.com/cortex-lab/phy>), if they had stereotypical spike shapes with low variance (Figure 2.2A), and their autocorrelation spike histogram showed a clear refractory period (Figure 2.2B). Spikes from a given cluster were often measurable on 4-6 neighboring electrode channels, facilitating the isolation of single units (Figure 2.2A). We also assessed the spike amplitude distribution, looking for units with a more Gaussian shape (Figure 2.2C). In addition, we took into account the stability of the recording and chose units that showed consistent activity throughout the stimuli presentation period (Figure 2.2D). Only well isolated single units were included in subsequent analyses.

### Inverse Current Source Density analysis

In order to directly compare the properties of neurons recorded with Neuropixels probes to those measured with two-photon calcium imaging, we aimed to isolate single units from cortical layers 2/3 in our electrophysiological recordings. To achieve this, we identified the boundary between layer 1 and 2 based on  $\delta$ -source inverse Current



**Fig. 2.2 Example well isolate single unit and poorly isolated multi unit.** **A**, Example raw spike waveforms from a ferret A1 recording (single unit - left, multi unit - right). The numbers indicate the channel number on the probe, 300 randomly chosen waveforms are plotted on top of each other. Notice how the spike waveforms for the single unit are present on many channels and are highly reproducible, while the multi unit's waveforms are much less distinct, more variable, smaller in amplitude and present on only a few channels. **B**, Autocorrelograms for the same neurons as in **A**. Each time bin is 1ms. Notice the clear refractory period for the single unit, which is absent in the multi unit. **C**, Histograms of the spike amplitudes for the two types of units. Notice the clearly Gaussian distribution for the single unit, while the multi unit has a non-Gaussian histogram. **D**, Raster plots of the two units. Amplitude is on the y-axis, time on the x-axis and each dot represents a single spike. Notice the smaller spread of amplitudes and overall better stability in time for the single unit compared to the multi unit.

Source Density (iCSD) in each recording penetration, and analyzed all single units down to a depth of  $400\mu\text{m}$  from this boundary (Pettersen et al., 2006).

The Local Field Potential (LFP) signal was isolated from the signal on each recording channel by bandpass filtering from 2-300Hz, and notch filtering at 50Hz to remove potential electrical noise. The evoked LFP trace was defined as time-series of the LFP signal starting 50 ms before tone presentation, and ending 50 ms after sound offset. Each LFP channel was impedance-normalized by subtracting the root mean squared power of the first 50 ms of the trial-averaged trace from the entire trace. This trace was averaged across repeated presentations and intensity levels for each tone frequency, and the frequency eliciting the strongest evoked LFP deflection was determined to be the BF for the channel. The iCSD of the recording was then computed from the LFP trace in response to tones presented at BF ( $\pm 0.3$  octaves), for all channels aligned vertically along the electrode using a published Matlab package (Olsen, 2020). The layer 1 to 2 boundary is characterized as a switch in polarity of the LFP (Christianson et al., 2011), and a sharp transition from a current source (layer 1) to a current sink (layers 2/3) in the iCSD (Szymanski et al., 2009, 2011; Cooke et al., 2018). For all A1 penetrations, there was a clear reversal in LFP polarity at the boundary between layers 1 and 2 (Figure 2.13), as previously described (Christianson et al., 2011; Szymanski et al., 2011).

### Calculating the best frequency of single units

Single units were identified on channels from the upper border of Layers 2/3 (based on current source density analysis) to depths  $400\mu\text{m}$  below this border. The evoked spike rate for each neuron on each trial was calculated as the sum of spikes from tone onset to offset. The proportion of frequency-sensitive neurons (two-way ANOVA;  $p < 0.05$ ), BF, and FRA classifications of these evoked spike responses were calculated as described above for calcium imaging data.

## 2.3 Results

We investigated how the frequency preferences of neurons in the superficial layers of ferret A1 are organized spatially, both locally (within  $\sim 0.25\text{mm}^2$ ) and along the entire tonotopic axis ( $\sim 3.5\text{mm}$ ), and how this organization compares to the local tonotopy in mice (Bandyopadhyay et al., 2010; Rothschild et al., 2010; Panniello et al., 2018). We

used an AAV vector to express either GCaMP6m (n=4) or GCaMP6f (n=4) in the right A1 of ferrets (Figure 2.3A; Figure 2.9A,B) and recorded neuronal calcium transients in 3640 neurons across 32 imaging fields (e.g. Figure 2.3A,B), while presenting pure tones (0.3-41kHz; 30–90dB SPL) to the contralateral ear. To ensure that our recordings were carried out in layers 2/3, we imaged neuronal activity at  $176 \pm 26.83\mu\text{m}$  (median  $\pm$  SD) below the cortical surface (Figure 2.10).

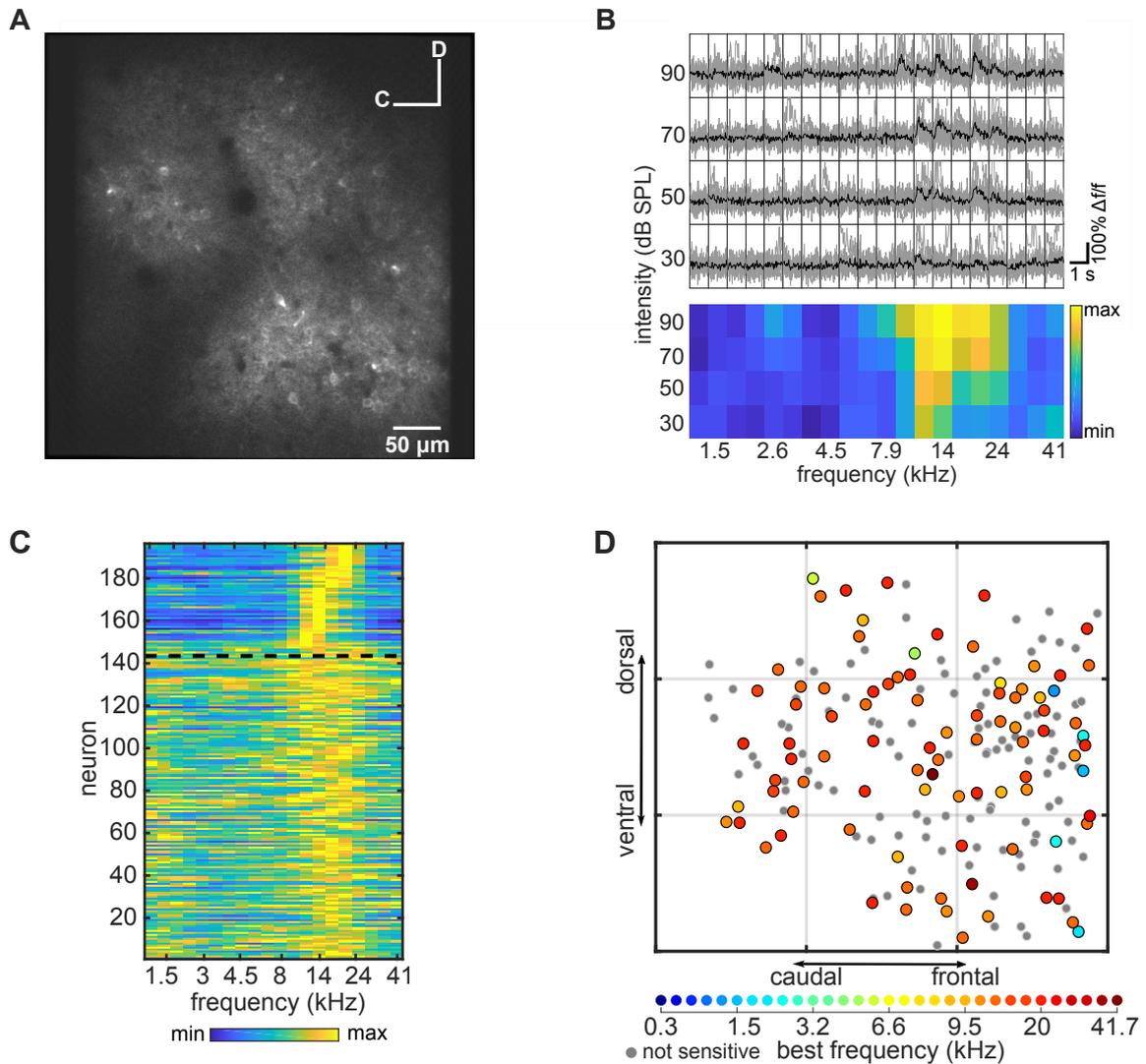
### 2.3.1 Frequency response organization at the local scale

Across all fields imaged in ferret A1, 693 of 3604 (19.23%) neurons were frequency-sensitive (two-way ANOVA;  $p < 0.05$ ). Neurons often showed a V-shaped Frequency Response Area (FRA; Figure 2.3B, Figure 2.4A), with a clearly defined BF. In the example field shown in Figure 2.3 (300 x 300 $\mu\text{m}$ ), 26.5% of neurons showed significant sensitivity to tone frequency or frequency/level combinations, and the average BF across all frequency-sensitive neurons (Figure 2.3C) was 14.7kHz ( $\pm 0.75\text{kHz}$ ; mean  $\pm$  SEM). As previously described for A1 of mice (Bandyopadhyay et al., 2010; Rothschild et al., 2010; Panniello et al., 2018), most neurons within this local region of ferret A1 were tuned to a similar BF, but there were some outliers preferring higher or lower BFs (Figure 2.3D). The BF of individual neurons in this field ranged from 2.2kHz to 19.9kHz, with 9% of BFs being more than one octave away from the mean.

The practice of assigning a BF is based on the assumption that neurons are tuned to a single preferred frequency, but some neurons can respond strongly to multiple frequencies (Bizley et al., 2005; Hackett et al., 2011; Sutter and Schreiner, 1991). We observed that the FRAs of frequency-sensitive neurons in our ferret A1 dataset could be classified into three broad types:

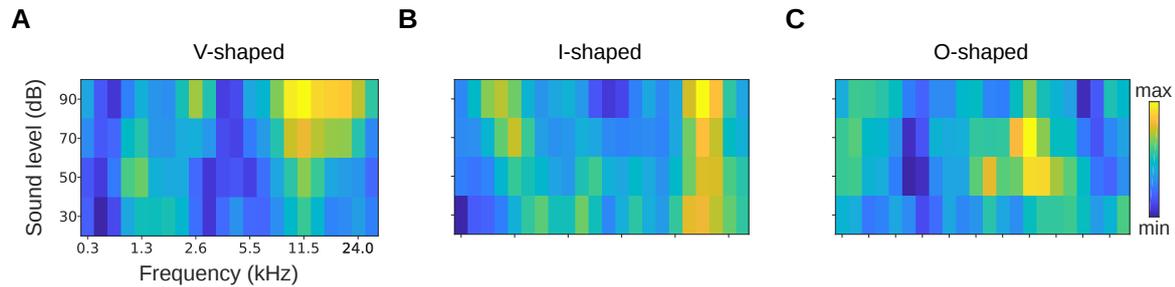
1. “Single-peaked neurons”, V-, I- or O-shaped FRAs, which all have a clear single BF (see Figure 2.4)
2. “Double-peaked neurons”, FRAs with response peaks at two distinct frequencies, where one peak is usually substantially stronger than the other
3. “Complex neurons”, with more speckled FRAs, often containing three response peaks and a poorly defined BF

Little is known about the functions of neurons with these different frequency response profiles, and their local spatial distribution has not been directly investigated, although



**Fig. 2.3 Imaging neuronal responses to pure tones in ferret A1.**

**A**, An example cortical field in A1 of ferret 1, imaged  $120\mu\text{m}$  below the pial surface. **B**, Responses of one neuron in (A) to pure tones presented at different frequencies and sound levels. Top panel: single trial (grey) and trial-averaged (black)  $\Delta F/F_0$  traces, measured for 1s from sound onset. Bottom panel: Frequency Response Area (FRA) of the same neuron. Color scale indicates the trial-averaged response of the neuron to tones presented at each frequency/level combination, calculated from the deconvolved fluorescence trace (see Methods). **C**, Level-averaged tuning curves of all neurons in A. Neurons above the dashed black line were significantly modulated by frequency (two-way ANOVA,  $p < 0.05$ ) and are sorted by their BF. Neurons below the line were not sensitive to sound frequency and are sorted by the p-value of the frequency predictor in the two-way ANOVA (bottom neurons have the largest p-value). **D**, Map of the neurons in A, colour-coded according to their BF (see colour scale below). Grey dots represent neurons that were not sensitive to tone frequency. Each grey square is  $100\mu\text{m}^2$ .



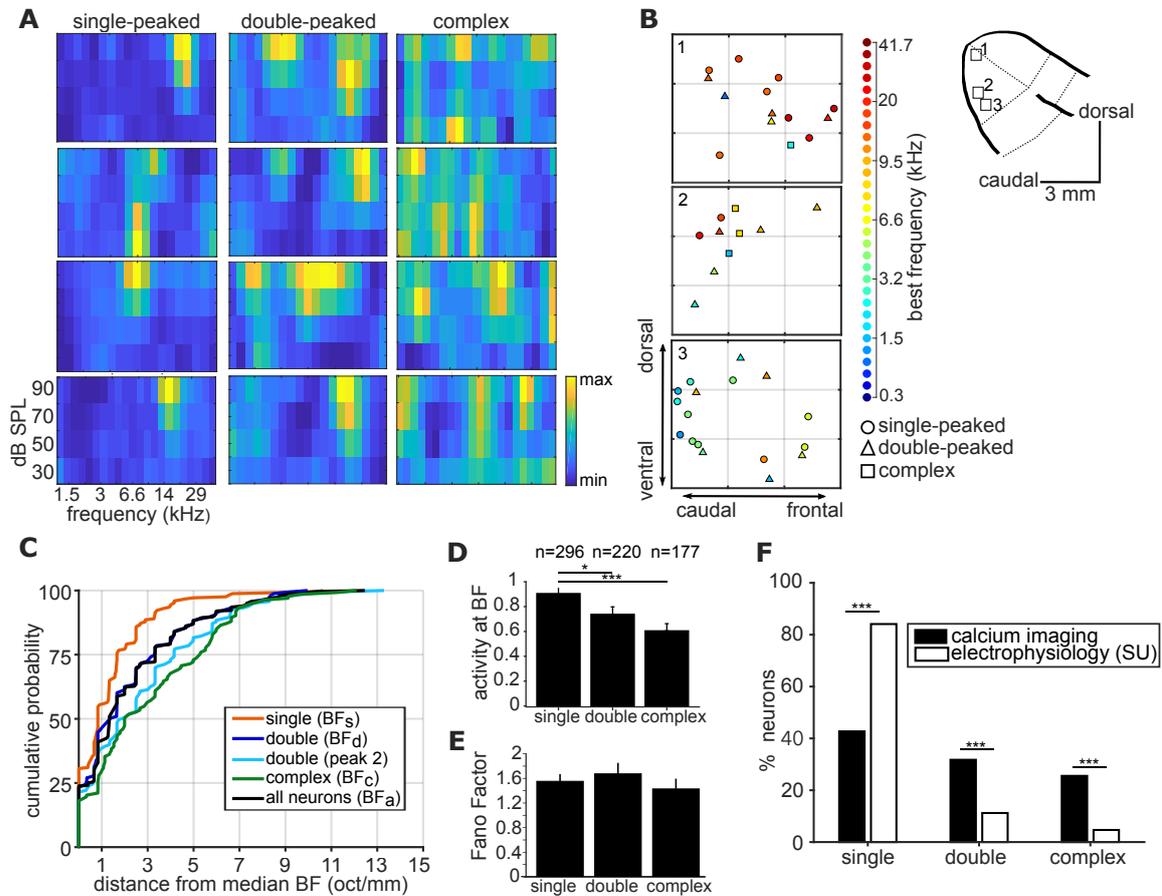
**Fig. 2.4 Example FRA shapes.**

Three common FRA shapes from one ferret. Color scale indicates the trial-averaged response of the neuron to tones presented at each frequency/level combination, calculated from the deconvolved fluorescence traces. **A**, An example V-shaped FRA. The tuning becomes broader with increasing sound level. This is the most common type of FRA. **B**, Example I-shaped FRA. Here the tuning width remains fairly constant with changing level and is localised around a given frequency, less common. **C**, An example O-shaped FRA. The tuning is non-monotonic with regards to sound level and looks like the letter “O” in the FRA plot. This type is much more rare.

neurons with multi-peaked FRAs have been reported in the A1 of marmosets (Kadia and Wang, 2003), cats (Sutter and Schreiner, 1991), rats (Turner et al., 2005) and mice (Winkowski and Kanold, 2013). We hypothesized that complexity in the shape of FRAs could help explain some of the reported local variability in cortical tonotopy, as it is less clear where double-peaked and complex FRAs should be located on a tonotopic map.

Each frequency-sensitive neuron was classified using an automated algorithm as having a single-peaked, double-peaked or complex FRA (see Methods; Figure 2.5A). In most imaging fields, we observed all three FRA classes, (as illustrated in Figure 2.5B), and the following analyses were restricted to imaging fields containing at least three neurons of each class. Best frequency was derived in the same manner for neurons across all three classes, as the peak of the frequency response curve.

Visual inspection of tonotopic maps (Figure 2.5B; Figure 2.11) suggested that the BFs of double-peaked ( $BF_d$ ) and complex neurons ( $BF_c$ ) may be more varied within a local field than those of single-peaked neurons ( $BF_s$ ). To quantitatively compare the local tonotopic organization of neurons with single-peaked, double-peaked and complex FRAs, we computed a BF variance metric. This metric was defined as the octave difference per millimeter between the BF of each neuron and the median BF of all other neurons in the imaging field.



**Fig. 2.5 Complexity of frequency receptive fields predicts local variability in ferret A1 tonotopy.**

**A**, FRA plots of four example neurons from ferret 6 for each of the three FRA classes. **B**, Three  $300 \times 300 \mu\text{m}$  spatial maps of frequency-sensitive neurons along the tonotopic axis of ferret 6. Their anatomical locations within the ferret's ectosylvian gyrus are represented in the schematic to the right. The colour of each neuron indicates its BF (see colour scale, right), and the shape corresponds to the classification of its FRA. **C**, Cumulative probability plots of the difference (in octaves) between each neuron's BF and the median BF of all neurons of the same FRA class in the same imaging field. Distributions for different FRA classes are plotted separately: single-peaked BF<sub>s</sub>; double-peaked BF<sub>d</sub>; double-peaked peak 2; complex BF<sub>c</sub>; and the BF<sub>a</sub> of all three FRAs together. **D**, Magnitude of the trial-averaged response at BF, calculated from deconvolved fluorescence traces in two-photon imaging experiments. Responses were averaged across all neurons in each FRA class, pooled across imaging fields and ferrets (mean  $\pm$  SEM). **E**, Fano Factor values (mean  $\pm$  SEM) for single-peaked, double-peaked, and complex neurons. **F**, Percentage of neurons classified into each FRA class from two-photon calcium imaging (black) and single unit microelectrode recordings (white). The results of t-tests are indicated above the bars in D and E, and Likelihood Ratio Tests in F (\* $p < 0.05$  and \*\*\* $p < 0.001$ ).

Best frequencies varied less within an imaging field among single-peaked neurons ( $1.34 \pm 0.10$  octaves; mean  $\pm$  SEM) than among double-peaked neurons ( $2.16 \pm 0.18$  octaves; t-test:  $t = -4.15$ ,  $p = 4.1 \times 10^{-5}$ ) and complex neurons ( $3.05 \pm 0.24$  octaves;  $t = 7.55$ ,  $p = 3.4 \times 10^{-13}$ ) (Figure 2.5C). Within double-peaked neurons, the frequency variability of the second peak (peak 2) and  $BF_d$  were comparable ( $t = 1.91$ ,  $p = 0.057$ ). Therefore, tonotopic organization within a local region of ferret A1 was less reliable among neurons with more complex frequency receptive fields.

We also investigated whether the three classes of neurons differed in their response strengths (Figure 2.5D). We found that the average deconvolved calcium response at the best frequency and level combination was stronger in single-peaked neurons compared to either double-peaked neurons (t-test:  $t = 2.22$ ,  $p = 0.027$ ) or neurons with complex FRAs ( $t = 4.02$ ,  $p = 6.9 \times 10^{-5}$ ). There was no significant difference in response strength between neurons with double-peaked and complex FRAs ( $t = 1.58$ ,  $p = 0.11$ ). The Fano Factor calculated at the best frequency and level did not significantly differ between neurons in the three FRA classes (single- and double-peaked:  $t = 0.60$ ,  $p = 0.55$ ; single-peaked and complex:  $t = 0.638$ ,  $p = 0.52$ ), indicating that responses at BF were equally reliable for neurons with single- and multi-peaked FRAs (Figure 2.5E). As expected from prior knowledge in the literature regarding the calcium indicators (Chen et al., 2013), when data from GCaMP6m and GCaMP6f injections were analyzed separately, we found a higher percentage of frequency-sensitive neurons in the GCaMP6m (32.29%), compared to the GCaMP6f dataset (10.81%). Importantly, both indicators showed similar effects of FRA class on local BF variance, response strength and response reliability (Figure 2.12), so our main findings are consistent across the two indicators, and data are pooled across all ferrets for our remaining analyses.

Extracellular recordings are known to be biased towards more active neurons, and Figure 2.5D suggests this may bias them towards single-peaked neurons. In addition, complex neurons may be expected to have more widespread dendritic branches, making them more prone to damage during electrode insertion. If either or both of these effects cause microelectrodes to oversample single-peaked neurons, this could explain the smoother tonotopic maps typically described using this technique. To investigate this possibility, we used high channel count microelectrodes to isolate the tone responses of single neurons in layers 2/3 of ferret A1 (Figure 2.13; see Methods), and the recorded FRAs were classified in the same way. The anaesthetic regime, surgery, and stimuli were similar to our imaging experiments. Although all three FRA classes were observed, the

microelectrode recordings yielded a higher proportion of neurons with single-peaked FRAs (Likelihood Ratio Test:  $\chi^2 = 68.45$ ,  $p = 1.1 \times 10^{-16}$ ) than the imaging experiments, and fewer with double-peaked ( $\chi^2 = 22.16$ ,  $p = 2.5 \times 10^{-6}$ ) and complex ( $\chi^2 = 30.06$ ,  $p = 4.2 \times 10^{-8}$ ) FRAs (Figure 2.5F).

Contrary to our two-photon imaging results, we did not find a significant difference between the spike rates of neurons with single- and double-peaked FRAs (t-test:  $t = 1.29$ ,  $p = 0.20$ ) or between those with single-peaked and complex FRAs ( $t = 0.05$ ,  $p = 0.96$ ) (Figure 2.14). This may result from a bias towards the most strongly responsive multi-peaked neurons, given the small number of double-peaked and complex neurons measured using microelectrodes. Together, our results suggest that biases towards sampling single-peaked neurons in microelectrode recording studies could lead to estimates of more ordered tonotopy than two-photon calcium imaging in the same cortical region.

The FRAs of single-peaked neurons can be well approximated by their BF. However, this is not the case for neurons with double-peaked and, in particular, complex FRAs. Consequently, an analysis based on BFs alone may be blind to an underlying organization of neurons with complex frequency receptive fields. To take into account a more complete representation of FRAs, we computed pairwise signal correlations. Correlations were calculated between pairs of frequency-sensitive neurons imaged simultaneously and belonging to the same FRA category: single-peaked, double-peaked and complex (Figure 2.6A), and were found to differ across the three classes (one-way ANOVA:  $F = 389.70$ ,  $p = 1.6 \times 10^{-163}$ ). In post hoc pairwise tests (Tukey's HSD), signal correlations within complex neurons ( $0.048 \pm 0.003$ ; mean  $\pm$  SEM) were significantly lower than those for single-peaked ( $0.207 \pm 0.002$ ;  $p = 9.6 \times 10^{-10}$ ) and double-peaked neurons ( $0.193 \pm 0.002$ ;  $p = 9.6 \times 10^{-10}$ ). Similarly, signal correlations were significantly higher within single-peaked neurons than double-peaked neurons ( $p = 4.4 \times 10^{-5}$ ). Furthermore, signal correlations decreased with cortical distance between neurons for single-peaked (Pearson's correlation:  $r = 0.21$ ,  $p = 5.3 \times 10^{-29}$ ; Figure 2.6E) and double-peaked cells ( $r = 0.17$ ,  $p = 1.5 \times 10^{-9}$ ; Figure 2.6F), but not for complex neurons ( $r = 0.06$ ;  $p = 0.12$ ; Figure 2.6G). These results further confirm that there is a greater degree of tonotopy for cells with simpler FRAs within a local cortical region.

To investigate whether single- and double-peaked neurons share different local networks from those with complex FRAs, we calculated pairwise noise correlations, which are thought to reflect connectivity and common inputs between neurons (Hofer

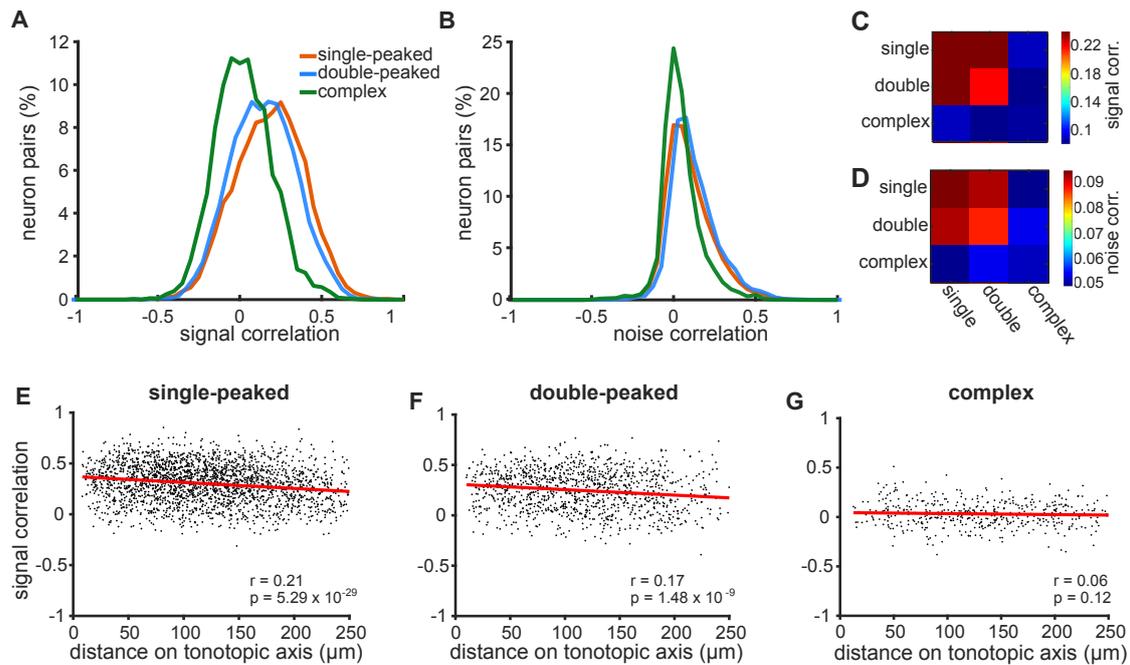
et al., 2011) (Figure 2.6B). Indeed, the strength of noise correlations differed across the three FRA classes (one-way ANOVA:  $F = 109.50$ ,  $p = 9.4 \times 10^{-48}$ ). Noise correlations were significantly higher for single-peaked ( $0.138 \pm 0.016$ ; mean  $\pm$  SEM) and double-peaked ( $0.136 \pm 0.004$ ) neurons than those with complex FRAs ( $0.059 \pm 0.004$ ) (Tukey's HSD tests; single vs complex:  $p = 9.6 \times 10^{-10}$ ; double vs complex:  $p = 9.6 \times 10^{-10}$ ). Noise correlations did not differ between single-peaked and double-peaked neurons ( $p = 0.86$ ). For all FRA types, noise correlations significantly decreased for pairs of neurons that were located further apart within the imaging field (Figure 2.15A-C). When we examined the correlation structure between neurons with different FRA classifications, we found that single- and double-peaked neurons had higher signal and noise correlations with one another than with complex neurons (Figure 2.6C,D). This suggests that neurons with complex FRAs may differ in their connectivity within the local cortical network from those with simpler frequency tuning.

### 2.3.2 Frequency response organization at the global scale

To investigate tonotopic organization along the entire extent of ferret primary AC, all neurons imaged in our 8 ferrets were projected onto a common template of A1. These projections were aligned based on both an anatomical criterion (i.e. the gross anatomy of the ectosylvian gyrus), and a functional one (i.e. A1 boundaries derived from neuronal responses to tones) (Figure 2.16A,B; see Methods).

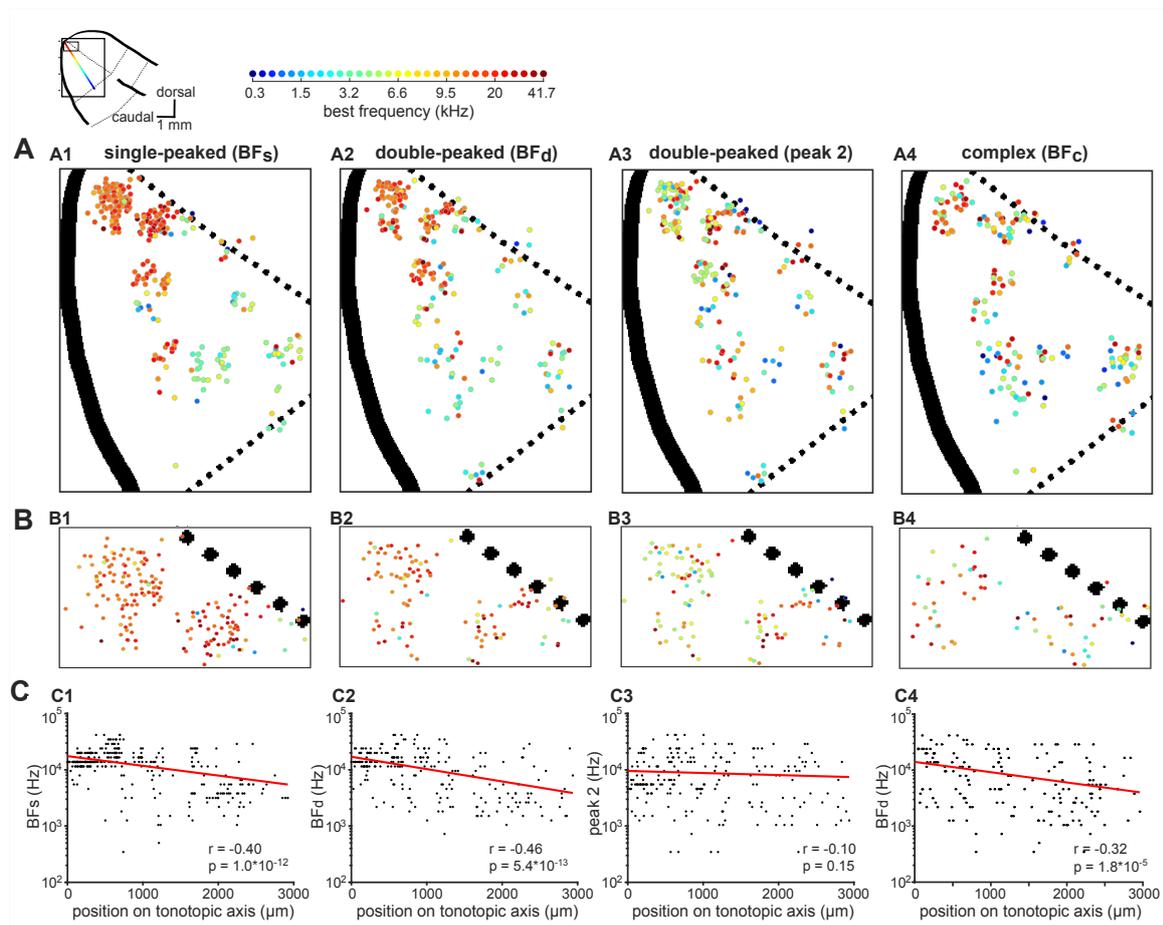
When single-peaked neurons were projected onto this global A1 template, the expected tonotopic gradient was clearly evident (Figure 2.7A1), with high frequencies mapped onto the dorsal tip of the ectosylvian gyrus (Figure 2.7B1) and lower frequencies toward the ventral border with secondary posterior fields (Kelly et al., 1986a; Bizley et al., 2005). The large-scale spatial organization of frequency preferences along the tonotopic gradient was quantified as a correlation between  $BF_s$  and the neuron's position along the tonotopic axis (Pearson's correlation:  $r = -0.40$ ,  $p = 1.1 \times 10^{-12}$ ; Figure 2.7C1).

When examining only neurons with double-peaked FRAs ( $BF_d$ ), global tonotopy was clearly visible (Figure 2.7A2, B2), and this tonotopic gradient was again statistically significant ( $r = 0.46$ ,  $p = 5.7 \times 10^{-13}$ ; Figure 2.7C2). However, the BFs of double-peaked neurons varied more around their fitted tonotopic gradient (red line, Figure 2.7C2) than



**Fig. 2.6 Correlations in neural activity are weaker and less spatially ordered for neurons with more complex FRAs.**

**A**, Distributions of signal correlations in the pure tone responses of pairs of simultaneously imaged neurons with single-peaked (orange), double-peaked (blue) or complex (green) FRAs. **B**, Distributions of noise correlations for the same neural pairs shown in **A**. **C**, The colour scale (right) shows the average signal correlation for pairs of neurons across all 9 combinations of the three FRA classes. **D**, Noise correlations across FRA classes, presented as in **C**. **E**, Pairwise signal correlations for single-peaked neurons plotted as a function of the distance between the two neurons along the tonotopic axis. The best linear fit to the data is shown (red line), with Pearson's correlation coefficient ( $r$ ) and p-value ( $p$ ) in the bottom right of the plot. **F**, Same as **E**, but for double-peaked neurons. **G**, Same as in **E**, but for complex neurons.



**Fig. 2.7 Global tonotopic organization of frequency preferences in ferret A1.**

The anatomical locations of neurons imaged across eight ferrets were projected onto a template map of AC, shown in the top-left corner. Thick black solid lines indicate sulci, and black dotted lines indicate approximate borders between known cortical fields. The two boxes represent the location of A1 (large box; A) and the most dorsal region of A1 (small box; B). The coloured line illustrates the tonotopic gradient. **A**, The BFs of individual neurons are colour-coded (legend above) and mapped onto A1. The spatial distributions of  $BF_s$  (A1),  $BF_d$  (A2), peak 2 of double-peaked neurons (A3), and  $BF_c$  (A4) are plotted separately. **B**, Frequency preferences are mapped as in A for the dorsal tip of A1, where many neurons are occluded in A. **C**,  $BF_s$  (C1),  $BF_d$  (C2), peak 2 (C3), and  $BF_c$  (C4) of each neuron are plotted against the neuron's position along the tonotopic axis on the template A1. Red lines show the best single-term exponential fits to the data, and Pearson's correlations ( $r$ ) with their  $p$ -values ( $p$ ) are also shown.

those of single-peaked neurons (t-test:  $t = 2.43$ ,  $p = 0.02$ ), suggesting that neurons with double-peaked FRAs have a less smooth tonotopic organization.

The BFs of neurons with complex FRAs were also organized tonotopically on the global A1 template ( $r = 0.32$ ,  $p = 1.8 \times 10^{-5}$ ; Figure 2.7A4, B4), but were more variable around the tonotopic gradient (Figure 2.7C4) than the BFs of either single-peaked ( $t = 6.83$ ,  $p = 2.7 \times 10^{-11}$ ) or double-peaked ( $t = 4.52$ ,  $p = 8.2 \times 10^{-6}$ ) neurons.

In contrast to BF, other aspects of the neurons' frequency responses were not found to be systematically ordered along the tonotopic gradient. The frequency of the second-strongest peak of double-peaked FRAs (peak 2) did not change systematically along this axis ( $r = -0.096$ ,  $p = 0.15$ ; Figure 2.7A3, B3, C3), nor did the differences between the two peaks of double-peaked neurons ( $r = 0.13$ ,  $p = 0.046$ ; Figure 2.17A-C).

These results indicate that, in ferrets, the BFs of A1 neurons are tonotopically organized at the global scale, but neurons with increasingly complex FRAs show more variance along the global tonotopic axis. In order to validate the method used for deconvolving calcium traces, we repeated our analyses on non-deconvolved  $\Delta F/F_0$  traces (Figure 2.18). The local BF variability was similar when computed from non-deconvolved (Figure 2.18A) and deconvolved traces (Figure 2.5C). In keeping with the data from the deconvolved signals, we found that the average amplitude of the calcium transient at BF was significantly higher in single-peaked neurons compared to double-peaked (t-test:  $t = 3.59$ ,  $p = 3.5 \times 10^{-4}$ ) and complex neurons ( $t = 5.10$ ,  $p = 5.0 \times 10^{-7}$ ) (Figure 2.18B; compare to Figure 2.5D). In addition, the Fano Factor calculated at BF from the non-deconvolved traces did not differ significantly between the three FRA classes (single- and double-peaked:  $t = 0.25$ ,  $p = 0.80$ ; single-peaked and complex:  $t = 1.030$ ,  $p = 0.30$ ; Figure 2.18C; compare to Figure 2.5E). At the global scale, the tonotopic organization was evident when BF was calculated from non-deconvolved traces (Figure 2.18D; compare to Figure 2.7B), and BF variability along the tonotopic axis was similar in the deconvolved and non-deconvolved datasets (two-way ANOVA:  $F = 1.24$ ,  $p = 0.27$ ; Figure 2.18E; see also Figure 2.7C).

Our analysis of both the local and global organization of frequency responses was carried out after removing the neuropil signal from each individual neuronal body. This procedure was necessary, as neuropil contamination can change the BF of individual neurons, particularly for those with complex frequency responses (see Methods; Figure 2.19A,B), as also highlighted in a recent study of mouse tonotopy (Romero et al., 2019). For example, we found that neuropil signal introduced the

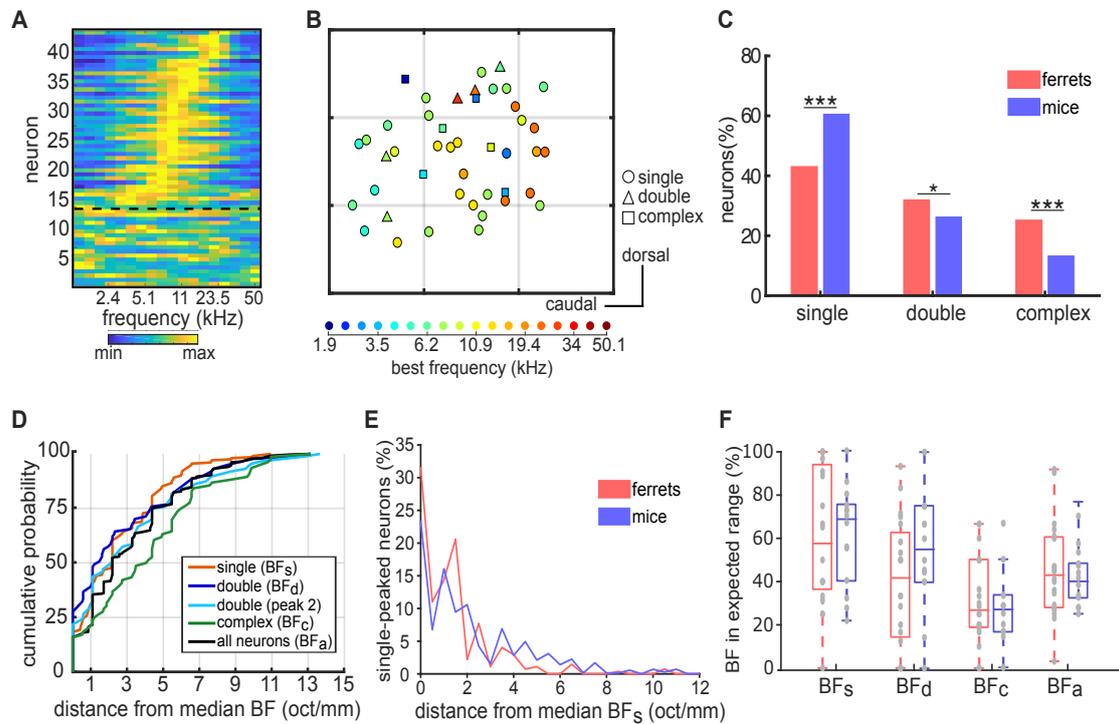
appearance of a tonotopic order to an otherwise spatially disorganized signal – namely, the second peaks of double-peaked neurons (Figure 2.19D3).

### 2.3.3 Comparison of tonotopic organization in ferrets and mice

The contrast between local heterogeneity and global tonotopic organization of BFs, sometimes referred to as “fractured” or “salt and pepper” organization, has been previously reported in the primary AC of the mouse (Bandyopadhyay et al., 2010; Rothschild et al., 2010; Panniello et al., 2018). The results above show that a qualitatively similar organization exists in the ferret A1, but that much of the heterogeneity at both the local and global scales arises from neurons with complex frequency receptive fields. To directly compare the local organization of frequency preferences between mice and ferrets, we imaged neuronal responses in the primary AC of 11 mice under the same experimental conditions.

The BFs of neurons and their spatial organization within an example imaging field in the mouse are shown in Figures 2.8A,B, respectively. In keeping with previous studies (Bandyopadhyay et al., 2010; Rothschild et al., 2010), these plots suggest that the spatial organization of BFs in the mouse primary AC is also locally heterogeneous. Across all imaging fields, 854 of 1964 (43.75%) neurons were frequency-sensitive in mice (two-way ANOVA;  $p < 0.05$ ), while only 693 of 3604 (19.23%) neurons were frequency-sensitive using the same statistical criterion in ferrets.

Classification of neurons based on their FRAs confirmed that single-peaked, double-peaked, and complex neurons also exist in the mouse (Figure 2.8B; Figure 2.20A-B), as they do in the ferret (Figure 2.5B). However, the relative proportions of these three FRA classes differ between the two species (Figure 2.8C). Specifically, a higher percentage of frequency-sensitive neurons in the mouse (60.54%) showed “simpler” frequency receptive fields with a single peak at BF, than in the ferret (42.71%; Likelihood Ratio Test:  $\chi^2 = 48.96$ ,  $p = 2.6 \times 10^{-12}$ ). Conversely, ferrets had more double-peaked (31.75% vs 26.23%;  $\chi^2 = 5.67$ ,  $p = 0.017$ ) and complex FRAs (24.96% vs. 13.23%;  $\chi^2 = 34.83$ ,  $p = 3.6 \times 10^{-9}$ ) than mice. Single-peaked and double-peaked neurons in mice responded more strongly at BF than complex neurons (single- and double-peaked:  $t = -0.73$ ,  $p = 0.46$ ; single-peaked and complex:  $t = 3.20$ ,  $p = 0.0010$ ; double-peaked and complex:  $t = 3.43$ ,  $p = 1.9 \times 10^{-4}$ ; Figure 2.20C), and trial-to-trial reliability in complex neurons



**Fig. 2.8 Similar tonotopic organization in the primary auditory cortex of ferrets and mice.**

**A**, Frequency Response Profile of all neurons imaged in one imaging field in one mouse. Data are presented as in Figure 2.3C. **B**, Map of the anatomical locations of the neurons in **A**, plotted as in Figure 2.5B. Neurons are color-coded according to their BF, and their shape corresponds to their FRA class. Each grey grid corresponds to a cortical area of  $100 \times 100 \mu\text{m}$ . **C**, Proportion of frequency-sensitive neurons from each FRA class for ferrets (red) and mice (blue). The results of Likelihood Ratio tests are indicated above (\* $p < 0.05$  and \*\*\* $p < 0.001$ ). **D**, Cumulative probability plots of the difference (octaves/mm) between the BF of each neuron and the median BF of all neurons in the same imaging field. As in Figure 2.5C, distributions were calculated separately for  $\text{BF}_s$  (orange),  $\text{BF}_d$  (dark blue), peak 2 of double-peaked neurons (light blue),  $\text{BF}_c$  (green), and  $\text{BF}_a$  (black). **E**, Distributions of the distance in octaves per mm between the BFs of each single-peaked neuron in an imaging field and the median BF of these neurons. Data are shown for ferrets (red) and mice (blue). **F**, Box plots showing the percentage of neurons having a BF within the expected frequency range, given the species, size and average BF of each imaging field. Percentages were calculated separately for  $\text{BF}_s$ ,  $\text{BF}_d$ ,  $\text{BF}_c$  and  $\text{BF}_a$ , and percentages for individual imaging fields are shown as grey dots. Boxplots for ferrets (red) and mice (blue) show the median, upper quartile, and lower quartile of each dataset.

was better than that of single-peaked neurons (single- and double-peaked:  $t = -0.99$ ,  $p = 0.32$ ; single-peaked and complex:  $t = 2.37$ ,  $p = 0.018$ ; Figure 2.20D).

Neurons with single-peaked FRAs showed broader frequency tuning in mice ( $0.91 \pm 0.03$  octaves) compared to ferrets ( $0.74 \pm 0.03$  octaves; t-test:  $t = 3.66$ ,  $p = 2.7 \times 10^{-4}$ ). This could partially account for the larger proportion of single-peaked neurons in mice, if the dominant peak occupies a greater proportion of the maximum bandwidth. However, there were no significant species differences in tuning bandwidth for double-peaked ( $t = 0.76$ ,  $p = 0.44$ ) or complex neurons ( $t = 0.72$ ,  $p = 0.47$ ; Figure 2.21A). Furthermore, larger bandwidths of single-peaked neurons were not associated with larger deviations in local BF in ferrets (Pearson's correlation:  $r = 0.020$ ,  $p = 0.73$ ; Figure 2.21B), and were weakly correlated with smaller BF deviations in mice ( $r = -0.11$ ,  $p = 0.030$ ; Figure 2.21C). In both species, bandwidth and BF variability were not significantly correlated in either double-peaked or complex neurons ( $p > 0.05$ ). Therefore, the sharpness of frequency tuning does not explain local BF variability.

Could the presence of double-peaked and complex FRAs help explain the local variance in cortical tonotopy in mice, as we observed in ferrets? To answer this question, the BF variance within an imaging field was calculated separately for single-peaked, double-peaked and complex neurons (Figure 2.8D), in every imaging field containing at least three neurons of each FRA class. Unlike in ferrets, the BF variability of double-peaked neurons ( $2.73 \pm 0.33$  oct/mm; mean  $\pm$  SEM) was not significantly higher than that of single-peaked neurons ( $2.57 \pm 0.16$  oct/mm; t-test:  $t = 0.46$ ,  $p = 0.64$ ) in mice, and the BF variability of the two peaks of double-peaked neurons did not differ significantly ( $t = 1.15$ ,  $p = 0.25$ ). In line with our ferret data, however, the BFs of neurons with complex FRAs ( $4.14 \pm 0.38$  oct/mm) were more variable than those of single-peaked ( $t = 4.33$ ,  $p = 2.0 \times 10^{-5}$ ) and double-peaked ( $t = 2.78$ ,  $p = 0.0060$ ) neurons, suggesting that the frequency preferences of neurons with more complex FRAs are less tonotopically organized within the local microcircuit. To ensure that the smaller number of multi-peaked neurons in mice did not bias our estimates of BF variance, we repeated this analysis using the same number of neurons in each imaging field. Furthermore, we bootstrapped our results across different subgroups of neurons in each field. The same patterns of BF variance were found across single-peaked, double-peaked and complex cells in these control analyses (Figure 2.22B,C,E and F).

As many neurons in sensory cortex can be unresponsive or poorly tuned to pure tone frequency, it is common practice to map only neurons that are significantly modulated

by sound frequency (Rothschild et al., 2010; Tischbirek et al., 2019). However, this inclusion criterion may favour the emergence of smoother BF distributions in mapping studies. Therefore, we repeated our BF analyses on all imaged neurons. The inclusion of more noisy neurons in this analysis predictably resulted in a larger proportion of complex FRAs overall. Importantly, mice again showed substantially more single-peaked neurons than ferrets (Figure 2.23A), as in our analysis of frequency-sensitive neurons (Figure 2.8C). Furthermore, more scatter in the spatial distribution of BF was observed with increased FRA complexity at both local (Figure 2.23B,C) and global (Figure 2.23D-G) levels of tonotopic organization, as we found for frequency-sensitive neurons (Figure 2.5C; Figure 2.7C; Figure 2.8D).

A comparison of the tonotopic organization between mice and ferrets is complicated by the anatomical and hearing range differences between the two species. The A1 tonotopic gradient is shorter in the mouse,  $\sim 1\text{mm}$  (Guo et al., 2012; Stiebler et al., 1997), than the ferret,  $\sim 3.5\text{ mm}$  (Kelly et al., 1986a; Bizley et al., 2005), and the audible hearing range is typically 2-60kHz in C57BL/6 mice (Heffner and Heffner, 2007; Ison et al., 2007) and 0.3-44kHz in ferrets (Kelly et al., 1986b). To illustrate the importance of these species differences, Figure 2.8E compares histograms of the  $\text{BF}_s$  variance in ferrets and mice, expressed in units of octaves per mm. The  $\text{BF}_s$  of single-peaked neurons were more variable within a millimetre of A1 in mice ( $2.18 \pm 0.08$  octave/mm) than in ferrets ( $1.33 \pm 0.09$ ; t-test:  $t = 5.62$ ,  $p = 2.5 \times 10^{-8}$ ). This may be taken to suggest that  $\text{BF}_s$  is more locally variable in mouse primary AC than in ferrets, but this metric does not account for the species differences in hearing range and tonotopic axis length.

To account for this difference, we computed the percentage of neurons that had a BF within the expected frequency range in each species, for a given imaging field. The expected BF range was calculated based on the diameter of the imaging field, the median BF across all neurons in the field, and an expected tonotopic gradient of 4.91 oct/mm in mice and 2.93 oct/mm in ferrets. To simplify the calculation, this measure assumes a constant log frequency gradient across the entire tonotopic axis. The percentage of neurons with a BF within the expected range provided a measure of the variance in tonotopic organization that could be directly compared between species, and was computed separately for neurons with: 1) single-peaked FRAs; 2) double-peaked FRAs; 3) complex FRAs; and 4) all three FRA types combined (Figure 2.8F). The resulting values varied across the three different FRA classes (two-way ANOVA:  $F = 8.41$ ,  $p = 4.3$

$\times 10^{-4}$ ), but not between species ( $F = 3.72$ ,  $p = 0.057$ ), and there was no significant interaction between FRA class and species ( $F = 1.66$ ,  $p = 0.19$ ). Furthermore, post hoc pairwise Tukey's HSD tests found no significant differences between BF variability in mice and ferrets for single-peaked ( $p = 0.97$ ), double-peaked ( $p = 0.12$ ), or complex neurons ( $p = 1.0$ ).

Taken together, these results suggest that mice and ferrets have equivalent local heterogeneity in frequency preferences within each FRA class of neurons in layers 2/3 of primary AC. More complex frequency tuning is associated with more variance in the frequency preferences of neighboring neurons in both species.

## 2.4 Discussion

We used two-photon calcium imaging to study the representation of sound frequency in populations of neighboring neurons in mouse and ferret primary AC. Previous studies in mice have described global A1 tonotopy with local variability of frequency preferences (Bandyopadhyay et al., 2010; Rothschild et al., 2010; Panniello et al., 2018; Romero et al., 2019; Winkowski and Kanold, 2013; Tischbirek et al., 2019), and here we show that ferret A1 shares the same spatial organization. Furthermore, we demonstrate that the complexity of frequency receptive fields can account for much of the local heterogeneity in both mice and ferrets. Thus, the locally heterogeneous sensory maps described in primary cortical areas may be a consequence of complex stimulus feature extraction.

### 2.4.1 Primary auditory cortex is tonotopically organized at the global scale

Electrophysiological and widefield imaging studies have described a tonotopic gradient in the primary AC of all mammalian species studied, including humans (Romani et al., 1982), monkeys (Merzenich and Brugge, 1973), cats (Sally and Kelly, 1988), ferrets (Kelly et al., 1986a), gerbils (Steffen et al., 1988), rats (Sally and Kelly, 1988) and mice (Stiebler et al., 1997). This global tonotopic organization has been confirmed in two-photon calcium imaging studies of primary AC in mice (Bandyopadhyay et al., 2010; Rothschild et al., 2010; Romero et al., 2019; Winkowski and Kanold, 2013; Tischbirek et al., 2019; Issa et al., 2014) and marmosets (Zeng et al., 2019). Here, we performed

the first two-photon calcium imaging of AC in carnivores, and show that this result generalizes to ferrets. The stereotyped position of ferret A1 within reliable anatomical landmarks allowed us to combine data from multiple animals onto a template of A1, as in previous electrophysiological studies (Bizley et al., 2009). This map showed a clear dorsal-to-ventral tonotopic gradient for neurons with single-peaked, double-peaked and, to a lesser extent, complex frequency receptive fields. In contrast, we observed no systematic spatial arrangement of the second frequency peak, or the difference between frequency peaks in double-peaked neurons. Together, these findings support the well-established idea that global tonotopy is the main spatial organizational principle in primary AC, and is observed within all neurons that are frequency sensitive.

### 2.4.2 Tonotopy is well preserved locally for neurons with simple, but not complex, frequency receptive fields

An advantage of two-photon calcium imaging over other techniques is its ability to sample the activity of large numbers of individual neurons within a local cortical region, while knowing their relative spatial locations. In both ferrets and mice, neurons with single-peaked FRAs were organized homogeneously, with almost 90% of them in a given imaging field presenting BFs within the expected frequency range, assuming a tonotopic gradient of  $<2.9$  oct/mm in mice and  $4.9$  oct/mm in ferrets. Signal correlation analysis further confirmed local tonotopic organization, as proximate neurons were more similarly tuned than distant ones within an imaging field ( $\sim 250\mu\text{m}^2$ ). Double-peaked neurons were less ordered, with  $\sim 60\%$  of their BFs within the expected range and greater local variability around the global tonotopic gradient, compared to single-peaked neurons. The BFs of neurons with complex frequency responses were even more spatially varied than double-peaked neurons. In contrast to single- and double-peaked neurons, only 30% of complex neurons in a given imaging field were within the expected BF range, and their signal correlations were lower and less spatially dependent. Previous studies have identified multi-peaked neurons in the A1 of mice (Winkowski and Kanold, 2013), rats (Turner et al., 2005), cats (Sutter and Schreiner, 1991) and marmosets (Kadia and Wang, 2003), and Romero et al. (2019) have recently shown that BFs are more locally variable among neurons with poorly-defined frequency tuning. Here, we further show that BFs are also more variable among nearby neurons with complex frequency receptive fields, even when their trial-to-trial responses at BF are equally reliable (as shown by our Fano Factor analysis). The association between frequency receptive field

complexity and tonotopy was remarkably similar for mice and ferrets, suggesting that this may be a common feature of the mammalian cortex. The complexity of receptive fields may also explain the observed local variability of orientation tuning in V1 (Bonin et al., 2011; Andermann et al., 2013) and whisker preferences in barrel cortex (Kerr et al., 2007; Sato et al., 2007).

The departure from strict tonotopy in A1, compared to earlier processing stations, such as the central nucleus of the inferior colliculus (Barnstedt et al., 2015), may in fact be an unavoidable result of complex spectrotemporal processing. We found that noise correlations were lower within local populations of complex neurons than within single- and double-peaked neurons. This may be due to neurons with complex frequency responses making more horizontal connections with distant cortical columns, or receiving inputs from layer 4 neurons in different areas of the tonotopic map. In V1, synaptic connections are more probable among V1 neurons with similar orientation tuning (Cossell et al., 2015). From our experiments, we might expect A1 neurons with simple frequency tuning to have stronger local connections to similarly tuned neurons, supporting well-defined tonotopy. Those with complex frequency responses, on the other hand, may receive inputs from neurons in more distant regions of the tonotopic gradient.

### 2.4.3 Similarities and differences between mice and ferrets

Our results show that an equivalent tonotopic organization of well-tuned neurons exists in ferrets and mice, despite the fact that the A1 of ferrets represents  $\sim 3$  oct/mm, while that of mice represents  $\sim 5$  oct/mm. Thus, local heterogeneity in the tonotopic map is unlikely to be due to limitations in brain size, or other potential factors that are inherent to rodents.

In contrast to this view, Zeng et al. (2019) recently reported that BFs of A1 neurons are more locally variable in rats than in marmosets. It is possible that the marmoset cortex contains a smoother tonotopic arrangement than other species studied, and this would be consistent with the high number of single-peaked neurons presented in their study (Zeng et al., 2019). Alternatively, the reported difference between marmoset and rat A1 tonotopy might be accounted for if differences in the size of A1, complexity of frequency tuning, and hearing ranges of these two species are taken into consideration.

Issa et al. (2014) reported more homogenous tonotopy in mouse A1 than other two-photon imaging studies (Bandyopadhyay et al., 2010; Rothschild et al., 2010; Panniello et al., 2018; Romero et al., 2019; Tischbirek et al., 2019), although the local variability in BF was still higher than that of marmosets (Zeng et al., 2019). Two distinguishing methodological features of Zeng et al. (2019) and Issa et al. (2014) are: (1) Issa et al. (2014) imaged tone responses in awake animals, and (2) neither study applied neuropil corrections. Both two-photon calcium imaging (Romero et al., 2019; Tischbirek et al., 2019) and single unit electrophysiological studies (Guo et al., 2012) have demonstrated that there is little change in a neuron's BF between passively listening and anaesthetized states, so this alone is unlikely to account for the smooth tonotopy observed by Issa et al. (2014). Neuropil signals contaminating neural responses can lead to smoother tonotopic maps (Romero et al., 2019), on the other hand, as the signals in axons and dendrites are averaged across the local population. In fact, our comparison of data with and without neuropil correction (Figure 2.14) demonstrated that neuropil contamination can result in the second, weaker peak of double-peaked neurons appearing to be tonotopically organized, while this is not spatially ordered in the neuropil-corrected signal. Finally, Issa et al. (2014) used the genetically encoded calcium indicator GCaMP3, which has a lower calcium sensitivity than GCaMP6. This may have led to the detection of only the strongest variations in calcium influx, typically seen here in single-peaked neurons.

Our study highlighted two main differences between mice and ferrets. First, in both species, the majority of neurons were not sensitive to any of the tone frequencies tested, but fewer A1 neurons exhibited frequency sensitivity in ferrets (19%) than in mice (44%). While these numbers may seem low, most superficial A1 neurons do not even respond to pure tones, especially in anaesthetized animals (Panniello et al., 2018; Tischbirek et al., 2019). Therefore, it is important to keep in mind that tonotopic maps do not reflect the spatial organization of response properties of most auditory cortical neurons. Second, a greater proportion of frequency-sensitive neurons showed double-peaked or complex FRAs in ferrets than in mice. In contrast, the proportions of multi-peaked neurons in our microelectrode recordings in ferret A1 are equivalent to those reported in marmoset A1 recordings (Kadia and Wang, 2003). These results may indicate that neurons in ferret and marmoset A1 are better suited to integrate information across frequency bands or encode more complex spectrotemporal features. For example, neurons with multiple frequency peaks would be useful for identifying sound sources based on their spectral timbre (Bizley et al., 2009), or may provide harmonic templates for pitch perception (Feng and Wang, 2017). Further physiological

and behavioural studies are required to test these potential species differences in auditory cortical function.

While layers 2 and 3 are often regarded as a single processing unit, neurons in these two laminae of A1 differ in their morphology, connectivity and function (Atencio and Schreiner, 2010; Oviedo et al., 2010; Meng et al., 2017). In our study, most ferret imaging areas were likely located in layer 2 (based on their depth), while calcium transients in mice were imaged in both layers 2 and 3. This may contribute to the species differences in our results. For example, Meng et al. (2017) reported that neurons in layer 2 have broader frequency tuning than those in layer 3. This potential laminar difference in imaging location could partially explain why neurons imaged in our mice dataset had broader frequency tuning than our ferret data.

#### 2.4.4 Comparison to previous electrophysiological studies

With some exceptions (Bizley et al., 2005), electrophysiological studies have typically reported smoother tonotopic gradients than two-photon imaging studies (Hackett et al., 2011; Stiebler et al., 1997). Our results offer an explanation for this inconsistency across methodologies. Compared to neurons with more complex FRAs, the BFSs of single-peaked neurons are more precisely mapped at both the local and global scale, and they also have stronger responses, as measured through their calcium dynamics (Figure 2.5D). Because extracellular microelectrode recordings are more likely to detect and cluster neurons with higher firing rates (Shoham et al., 2006), such studies may oversample neurons with single-peaked FRAs and therefore with more precise tonotopy. In support of this hypothesis, we found that >80% of frequency-sensitive neurons that we recorded from superficial A1 using microelectrodes had single-peaked FRAs: approximately double the proportion of single-peaked FRAs identified using two-photon calcium imaging in the same species. An electrophysiological study of marmoset AC also found that 20% of neurons showed multi-peaked frequency tuning (Kadia and Wang, 2003).

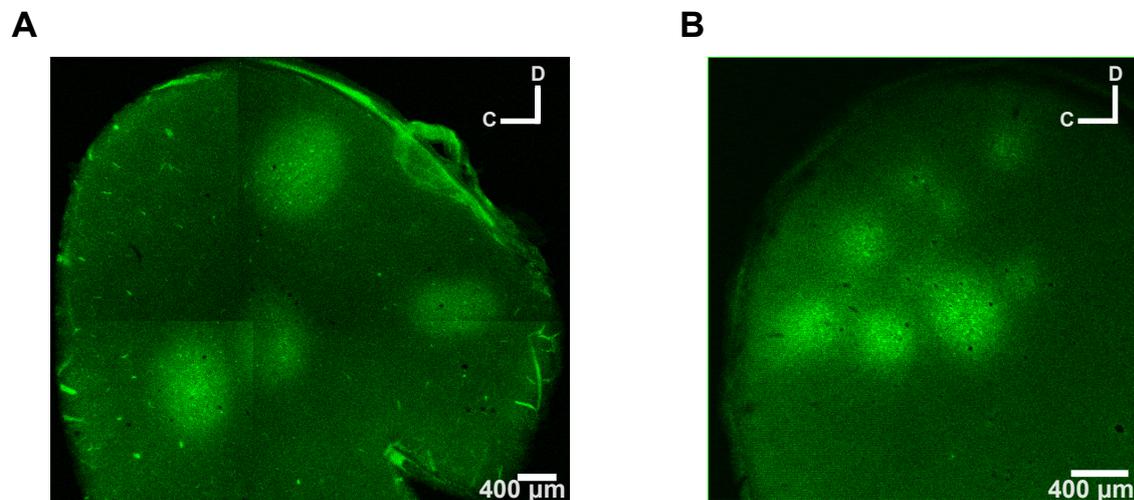
Electrophysiological studies in mice have also reported that the thalamorecipient layers of A1 are more tonotopically ordered than more superficial layers (Hackett et al., 2011; Guo et al., 2012; Stiebler et al., 1997). However, two-photon imaging studies have provided conflicting evidence about this (Winkowski and Kanold, 2013; Tischbirek et al., 2019). We suggest that the more ordered tonotopy in layers 4/5 may arise from a

larger proportion of single-peaked neurons in this part of the cortex. This is consistent with the electrophysiological data of [Guo et al. \(2012\)](#), who report that more neurons with irregular frequency receptive fields exist in the superficial layers of mouse A1. Here, we show that tonotopic variability in layers 2/3 is not simply due to neurons having less reliable responses, but is also associated with neurons that respond to a combination of distinct frequency bands.

### 2.4.5 Conclusions

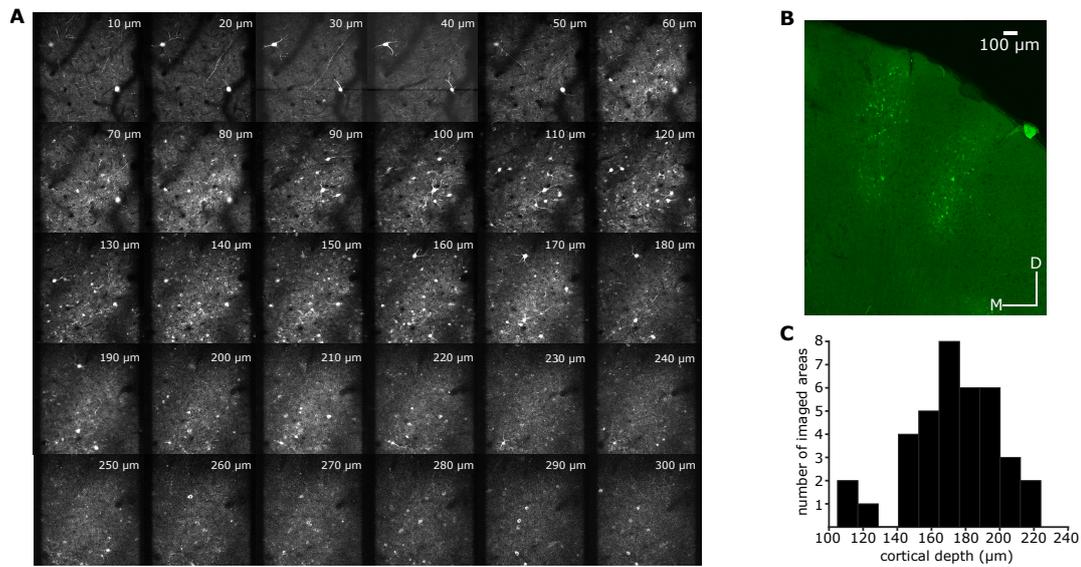
Applying two-photon imaging to the study of neuronal activity in the ferret AC for the first time, we found that A1 neurons tuned to a single best frequency are tonotopically organized at both the local and the global scales. The presence of neurons with more complex FRAs disrupts this strict tonotopic order, increasing the local heterogeneity of neuronal frequency preferences, just as it does in mice. Cells with complex sensory receptive fields are likely to be important for extracting information from natural environments, and are more common in ferrets than mice. Future imaging studies are required to better understand the functional properties and connectivity of these subpopulations of neurons.

## 2.5 Supplementary figures



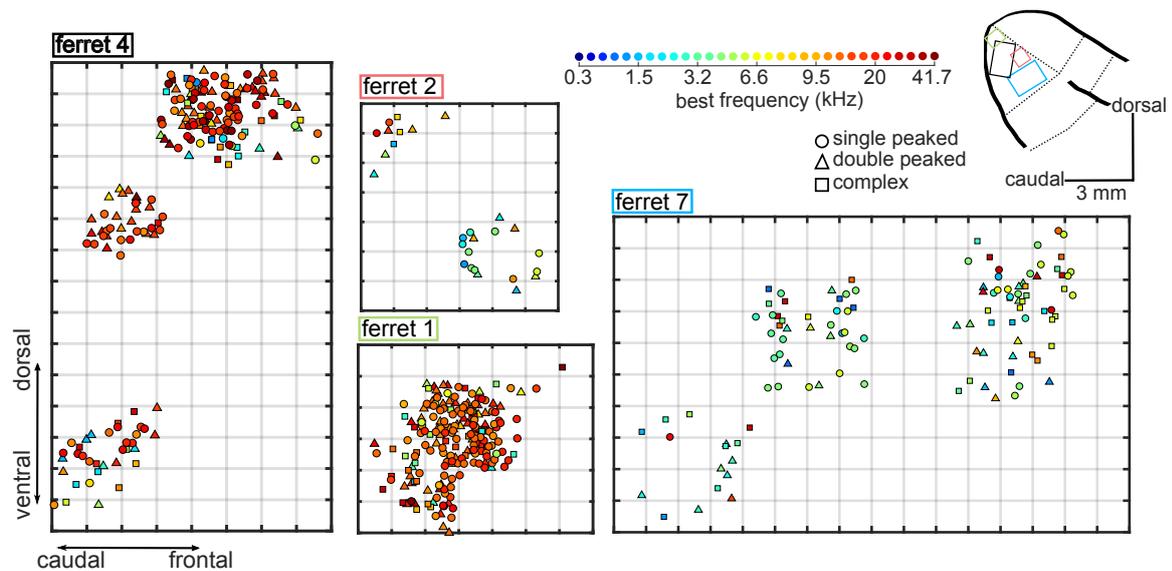
**Fig. 2.9 GCaMP6 expression in ferret auditory cortex.**

Tangential sections of the AC are shown to visualize the locations and spread of GCaMP6 injections. Sections were  $50\mu\text{m}$  thick and were acquired at  $200\mu\text{m}$  below the cortical surface. **A**, The AC of ferret 3 was injected with a 1:1 solution of AAV1.Syn.GCaMP6f.P2A-nls-dTomato.WPRE.SV40 and AAV1.mDlx.GFP-GCaMP6f-Fishell-2.WPRE.SV40 in four cortical locations. The typical spread of these injections was  $\sim 800\mu\text{m}$  in diameter per site. **B**, AC shown as in **A**, but for ferret 7. Here, a 1:1 solution of AAV1.Syn.GCaMP6m.WPRE.SV40 and PBS was injected at seven injection sites, and the typical spread was  $\sim 400\mu\text{m}$  in diameter per site.



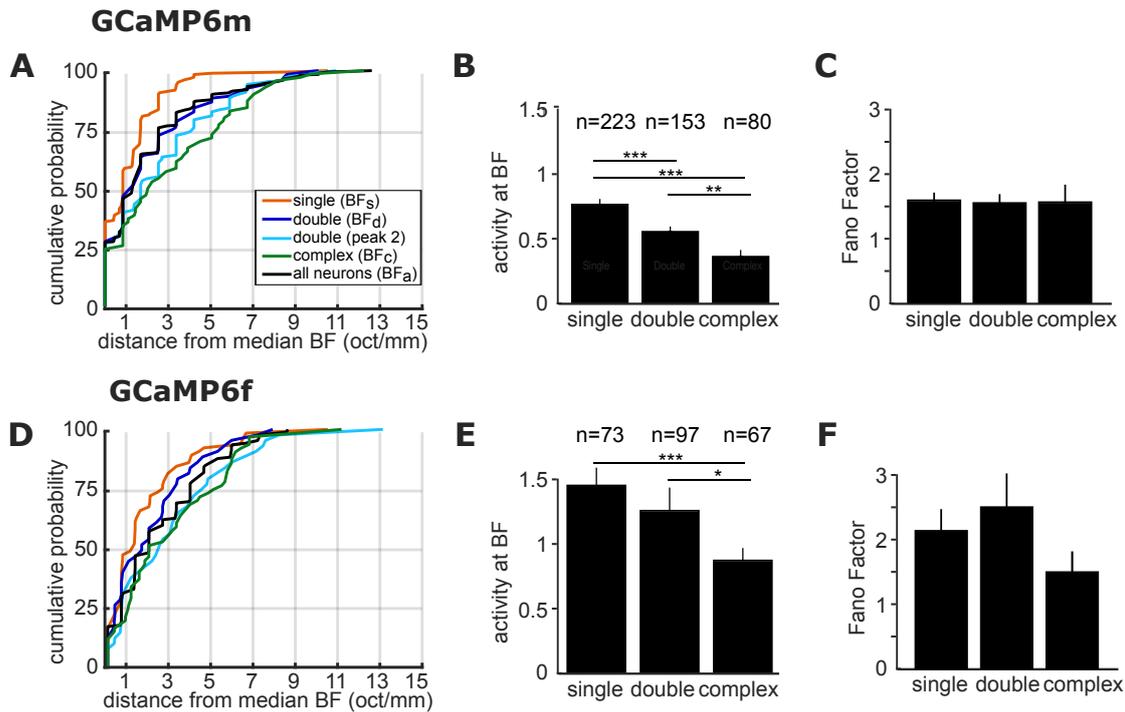
**Fig. 2.10 Imaging in layer 2/3 of ferret A1.**

**A**, A series of z-projections acquired *in vivo* using two-photon imaging in ferret 7. The projections have the same fronto-caudal and dorso-ventral axis as the imaged field shown in Figure 2.3. Projections were selected to show 10  $\mu\text{m}$  increments in cortical depth. While imaging, the most superficial neuronal bodies were visible  $\sim 60 \mu\text{m}$  below the cortical surface. **B**, A coronal section of the A1 of ferret 4 showing the extent of GCaMP6 labeling throughout the cortical depth. **C**, Distribution of the number of fields imaged at each cortical depth.



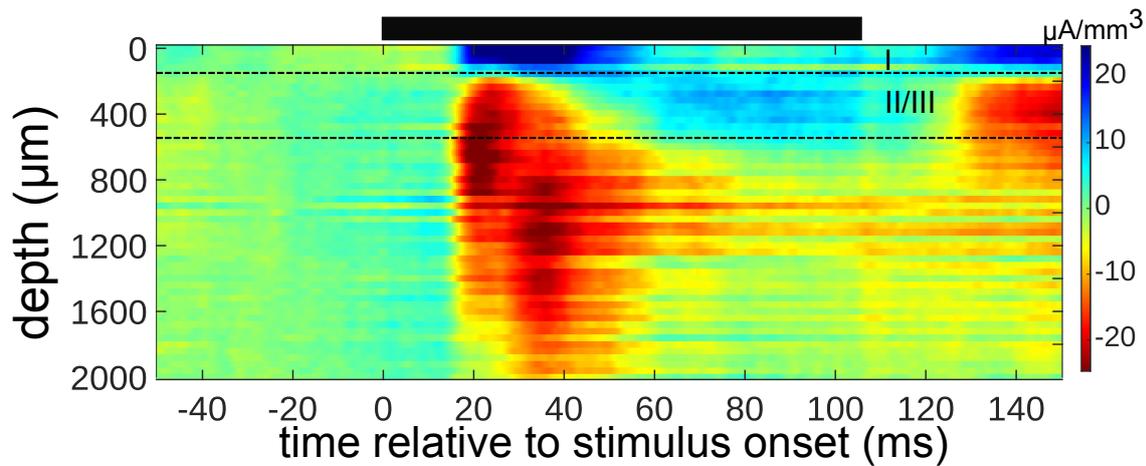
**Fig. 2.11** Spatial maps of best frequency in A1 of 4 ferrets.

Each plot is a map of frequency-responsive neurons imaged in individual ferrets (ferrets 1, 2, 4 and 7). Neurons are plotted along the rostro-caudal and dorso-ventral axes. The color of each neuron indicates its BF, while the shape corresponds to the classification of its FRA (see colour scale, top center), as in Figure 2.5B. Each grey grid in the maps corresponds to a cortical area of  $100 \times 100 \mu\text{m}$ . The locations of the four maps within A1 are represented on a common map of ferret AC as coloured rectangles (top right, colour matches ferrets' numbers).



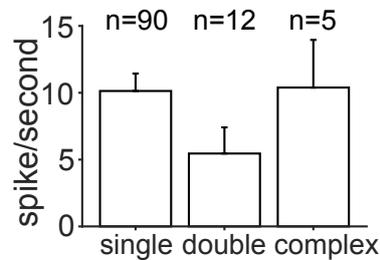
**Fig. 2.12 BF variance, response strength and response reliability are comparable in GCaMP6m and GCaMP6f ferret datasets.**

Top row **A**, **B**, **C**: results from imaging experiments using GCaMP6m as calcium indicator. Bottom row **D**, **E**, **F**: same analyses as in **A**, **B**, **C** above, but using data from GCaMP6f experiments. **A**, **D**, Cumulative probability plots of the difference (in octaves) between each neuron's BF and the median BF of all neurons of the same FRA class in the same imaging field. Distributions for different FRA classes are plotted separately as colored lines, as in Figure 2C (see legend in **A**). **B**, **E**, Magnitude of the trial-averaged response at BF, calculated from deconvolved fluorescence traces. Responses were averaged across all neurons in each FRA class, and pooled across imaging fields and ferrets (mean  $\pm$  SEM). **C**, **F**, Fano Factor values (mean  $\pm$  SEM) for single-peaked, double-peaked, and complex neurons. The results of t-tests are indicated above the bars in **B** and **E** (\*p < 0.05, \*\*p < 0.01 and \*\*\*p < 0.001).



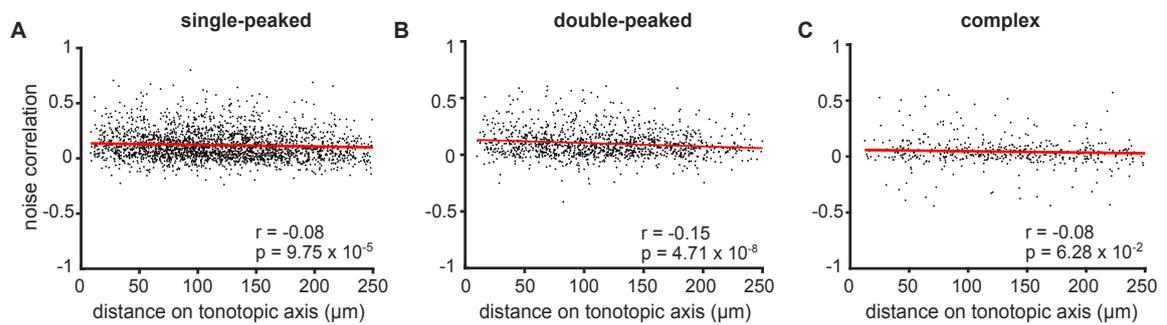
**Fig. 2.13** Determining cortical layers from an iCSD map of electrophysiological responses.

The iCSD shown was constructed from local field potential recordings of responses to tones presented near BF ( $\pm 1/3$  octave) in one microelectrode penetration. The black bar above the plot indicates stimulus presentation time. The depth in  $\mu\text{m}$  relative to the cortical surface is plotted on the y-axis. The boundary between layers 1 and 2 (upper black dotted line) was inferred based on the reversal in the iCSD profile, and the boundary between layers 3 and 4 (lower black dotted line) was estimated to be  $400\mu\text{m}$  below the beginning of layer 2.



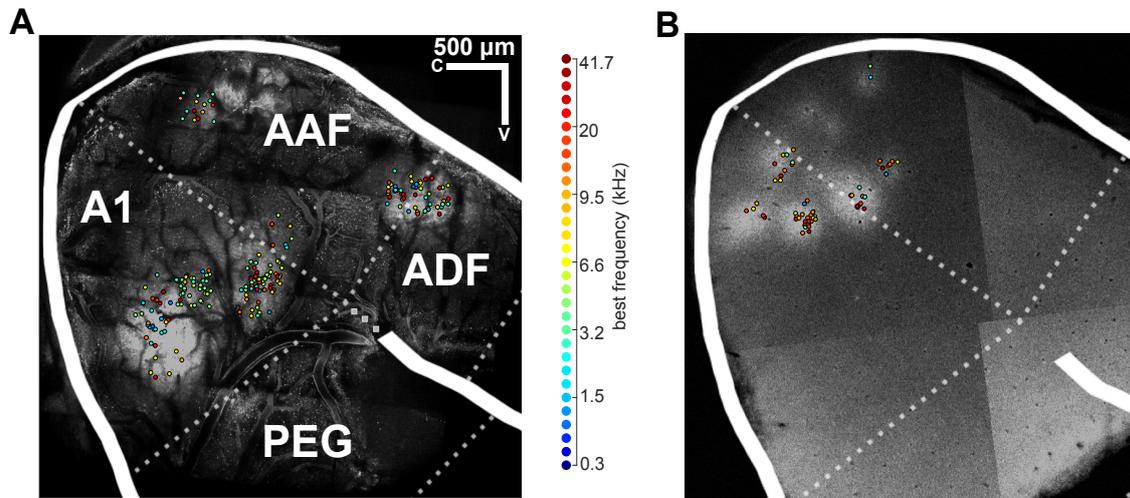
**Fig. 2.14** Firing rates per FRA type from Electrophysiology.

Magnitude of the trial-averaged response at the best frequency and level combination, calculated from spike rates measured with microelectrodes. Responses were averaged across all neurons in each FRA class, pooled across penetrations and ferrets (mean  $\pm$  SEM). The number of neurons per group is given above each bar.



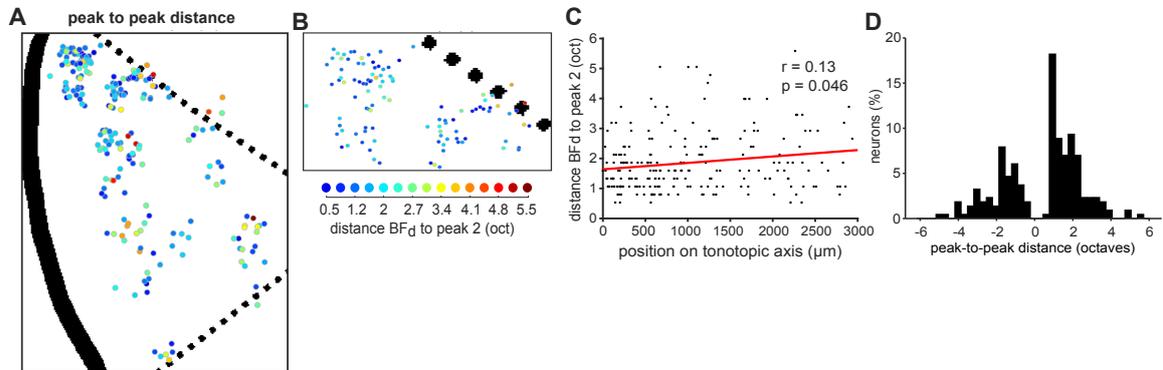
**Fig. 2.15 Noise correlations in neurons with single-peaked, double-peaked and complex FRAs.**

**A**, Noise correlations for pairs of single-peaked neurons imaged simultaneously, as a function of the distance between the two neurons along the tonotopic axis. **B**, Same as **A** for double-peaked neurons. **C**, Same as **A** for complex neurons. Red lines indicate the best linear fit to the data. Pearson's correlation coefficient ( $r$ ) and  $p$ -value ( $p$ ) are also shown.

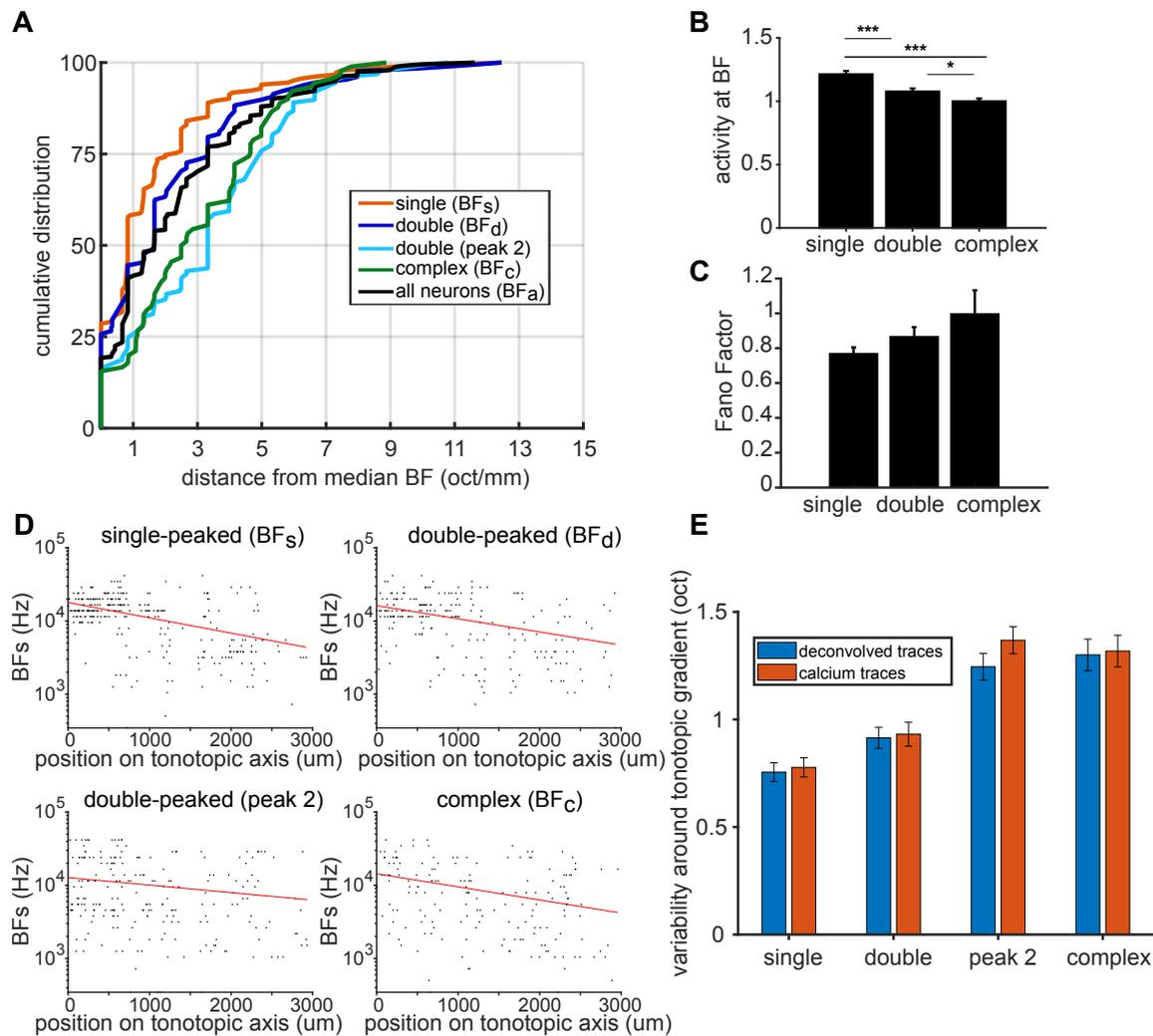


**Fig. 2.16 Mapping data from individual ferrets onto a common template of auditory cortex.**

**A**, An image of the brain surface of ferret 3, acquired *in vivo* using two-photon calcium imaging. Individual frequency-sensitive neurons imaged across all fields are plotted on this map according to their anatomical position, where their colour indicates their BF (colour scale, right). The common template of AC is superimposed, where white solid lines indicate sulci and white dotted lines indicate approximate borders between known cortical fields (names in white text). The brain image and template were aligned according to the positions of sulci, and frequency responses of imaged neurons. Only neurons in A1 were included in our analyses. **B**, A post mortem section of the AC of ferret 7, which was cut parallel to the two-photon imaging plane (i.e. parallel to the surface of AC) and imaged with a confocal microscope. The BF of imaged neurons and the cortical template are superimposed, as in A. A1 = primary AC; AAF = anterior auditory field; PEG = posterior ectosylvian gyrus (containing the posterior pseudosylvian and posterior suprasylvian fields); ADF = anterior dorsal field.

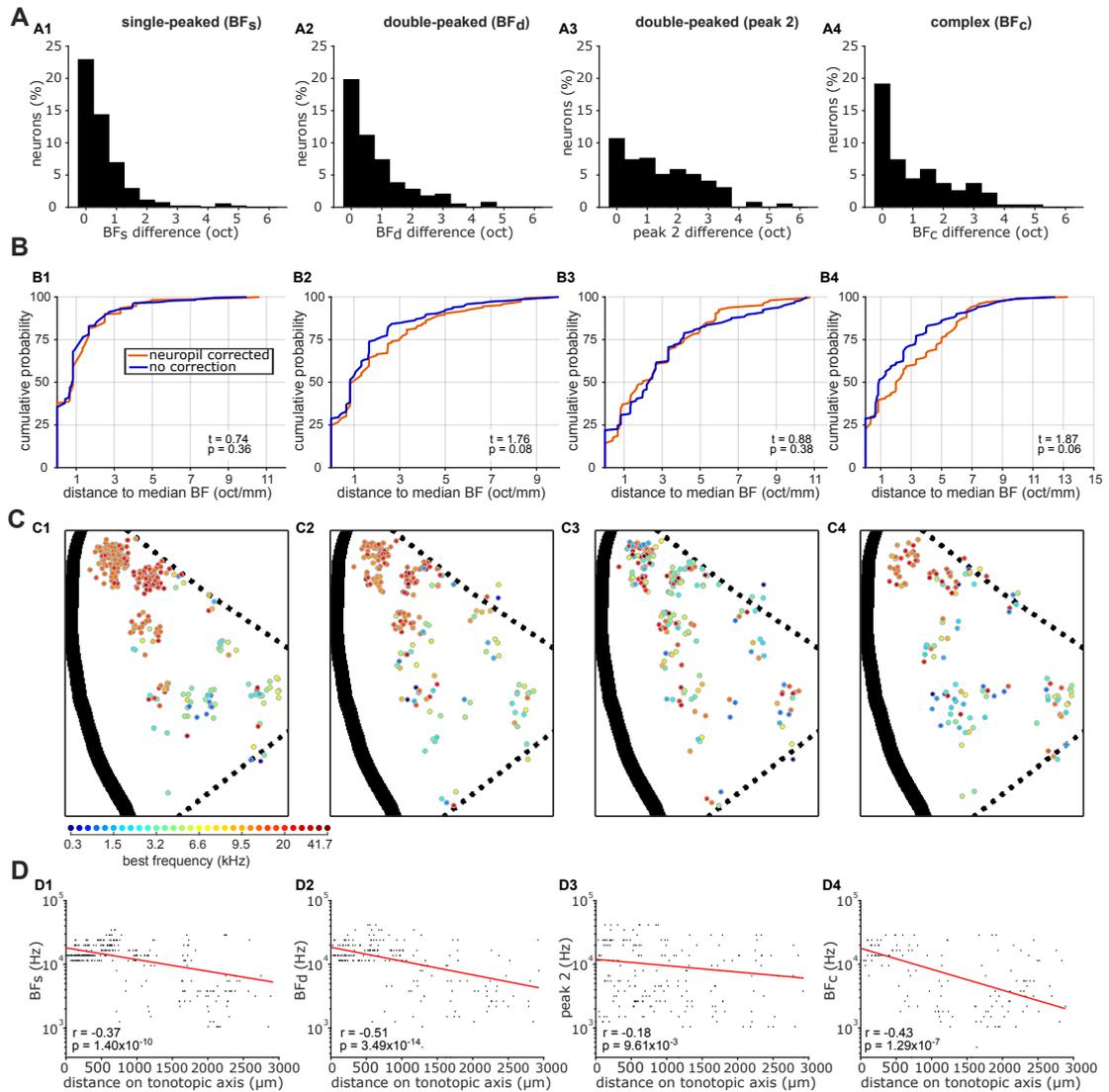


**Fig. 2.17 Mapping the peak-to-peak distances in double-peaked neurons.** **A, B,** Double-peaked neurons imaged across all ferrets are mapped onto the template A1, as in Figures 2.7A,B. Here, the color scale (below B) indicates the difference, in octaves, between  $BF_d$  (i.e. the frequency eliciting the strongest response) and peak 2 (i.e. the frequency eliciting the second-strongest response) for each neuron. **C,** The octave difference between  $BF_d$  and peak 2 is plotted for each double-peaked neuron, as a function of its position along the tonotopic axis (as in Figure 2.7C). The red line represents the best linear fit to the data, and the coefficient ( $r$ ) and  $p$ -value ( $p$ ) of Pearson's correlation are shown. **D,** Distribution of the difference (in octaves) between  $BF_d$  and peak 2, across all double-peaked neurons.  $BF_d$  and peak 2 were on average  $1.74 \pm 0.07$  (mean  $\pm$  SEM) octaves apart, and their octave distances did not differ when  $BF_d$  was lower or higher than peak 2 (t-test:  $t = -0.17$ ,  $p = 0.87$ ).



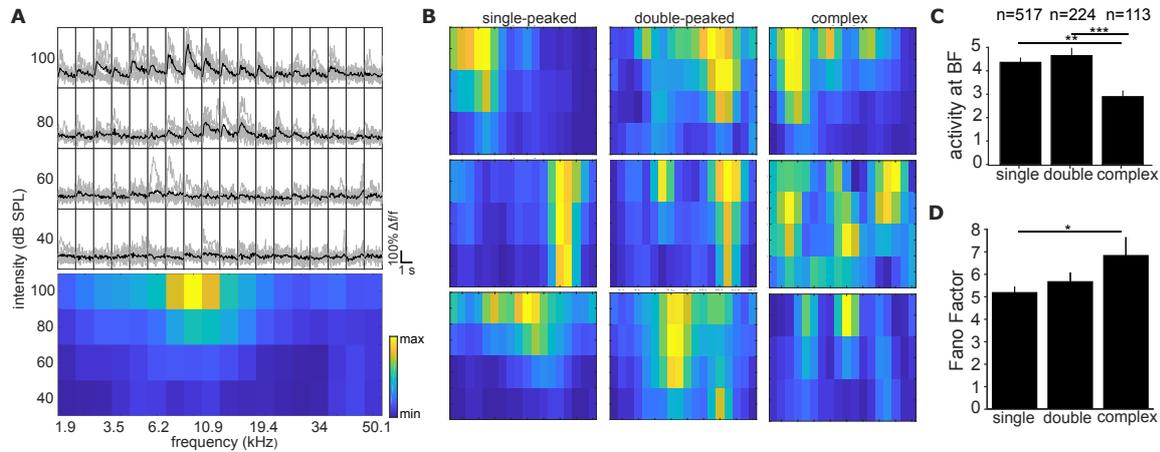
**Fig. 2.18 Comparison between deconvolved and non-deconvolved traces.**

**A**, Cumulative probability plots of the difference (in octaves) between each neuron's BF and the median BF of all neurons in the same FRA class and imaging field. Distributions for different FRA classes are plotted separately: single-peaked BF<sub>s</sub>; double-peaked BF<sub>d</sub>; double-peaked peak 2; complex BF<sub>c</sub>; and the BF<sub>a</sub> of all three FRAs together. **B**, Magnitude of the trial-averaged response at BF, calculated from non-deconvolved fluorescence traces. Responses were averaged across all neurons in each FRA class, pooled across imaging fields and ferrets (mean  $\pm$  SEM). Significant results of t-tests are shown above the bars: \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ . **C**, Fano Factor values calculated at BF (mean  $\pm$  SEM) for single-peaked, double-peaked, and complex neurons in non-deconvolved fluorescence traces. T-tests show no significant differences between groups ( $p > 0.05$ ). **D**, BF<sub>s</sub>, BF<sub>d</sub>, peak 2, and BF<sub>c</sub> of each neuron are plotted against the neuron's position along the tonotopic axis on the template A1. Red lines show the best single-term exponential fits to the data. **E**, Average octave distance of each BF from the fit (red curves in D and in Figure 2.7C in the main text) for deconvolved traces (blue) and non-deconvolved fluorescence traces (red) (mean  $\pm$  SEM). T-tests show no significant differences between results for each FRA type ( $p > 0.05$ ).

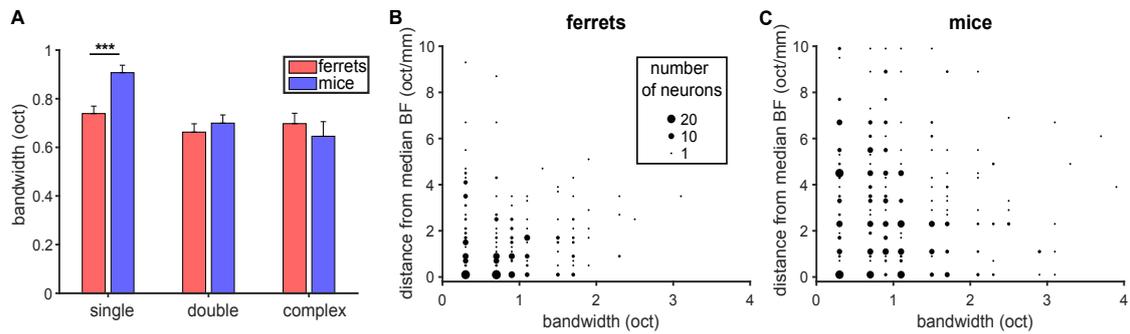


**Fig. 2.19 Effects of neuropil contamination on local and global tonotopic organization.**

**A**, Distributions of the difference (in octaves) between  $BF_s$  (A1),  $BF_d$  (A2), peak 2 (A3) and  $BF_c$  (A4) calculated with and without neuropil signal subtraction. **B**, Cumulative probability plots of the difference (in octaves) between each neuron's BF and the median BF of all frequency sensitive neurons in the same imaging field. Distributions for BFs obtained with (red) and without (blue) neuropil correction are plotted separately. The results of t-tests between the two distributions are given in the bottom right of each plot. **C**, Frequency preferences are mapped onto a template A1, as in Figure 4A, but here without neuropil correction. **D**, Scatterplots show the frequency preferences of neurons as a function of their position along the tonotopic axis of the maps in C (as in Figure 4C), for neuropil-uncorrected data.

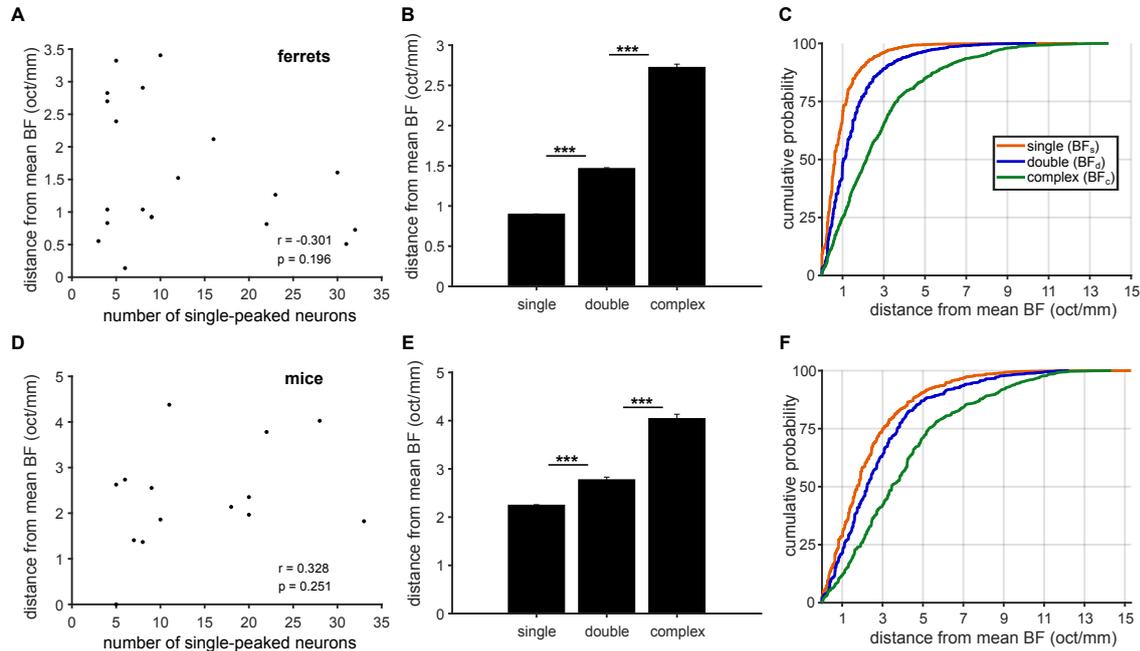


**Fig. 2.20 All 3 FRA types are present in the mouse primary auditory cortex.** **A**, Responses of one neuron to pure tones presented at different frequencies and sound levels. Top panel: single trial (grey) and trial-averaged (black)  $\Delta F/F_0$  traces, measured for 1.5s from sound onset. Bottom panel: Frequency Response Area (FRA) of the same neuron. The colour scale indicates the trial-averaged response of the neuron to tones presented at each frequency/level combination, calculated from the deconvolved fluorescence trace (see Methods). **B**, FRA plots of three example neurons for each of the three FRA classes, imaged in one mouse. **C**, Magnitude of the trial-averaged response at BF, calculated from deconvolved fluorescence traces, as in Figure 2.5D. Responses were averaged across all neurons in each FRA class, pooled across imaging fields and mice (mean  $\pm$  SEM). **D**, Fano Factor values (mean  $\pm$  SEM) for single-peaked, double-peaked, and complex neurons. The results of t-tests are indicated above the bars in C and D (\* $p < 0.05$  and \*\*\* $p < 0.001$ ).



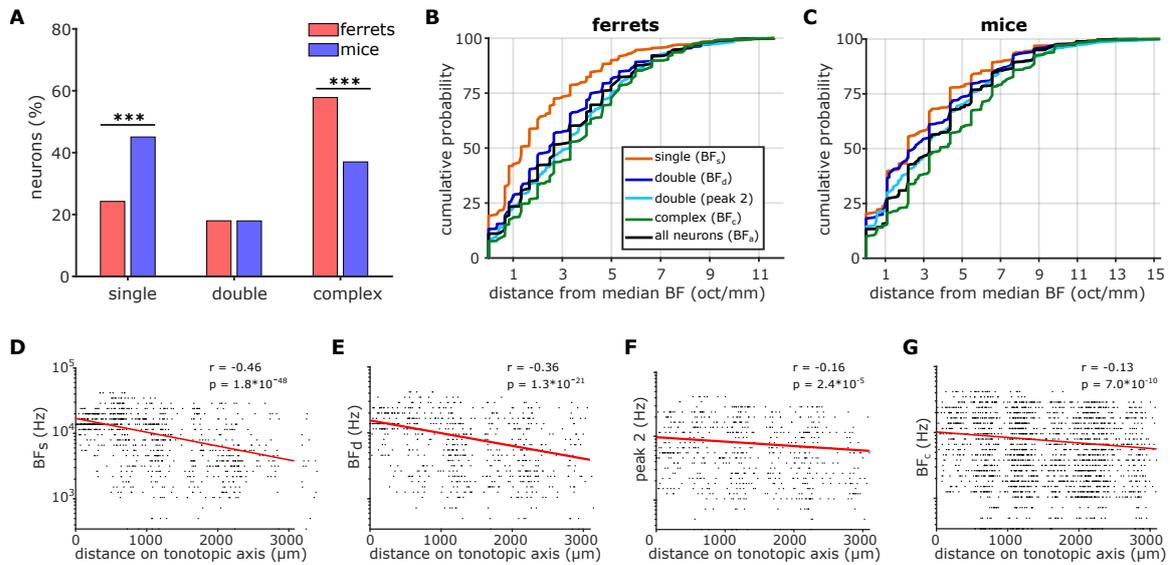
**Fig. 2.21 FRA bandwidth does not predict local tuning heterogeneity.**

**A**, Average (mean  $\pm$  SEM) frequency tuning bandwidth around BF for single-peaked, double-peaked and complex neurons in ferrets and mice (t-test; \*\*\*p < 0.001). **B**, **C** The scatterplots show the frequency tuning bandwidth as a function of the distance of the BF for each neuron from the median BF value in each imaging field in ferrets (**B**) and mice (**C**). Data are shown for single-peaked neurons only. Dot size is proportional to the number of neurons (see legend).



**Fig. 2.22** The BF variance within an imaging field does not depend on the number of neurons in the field.

**A**, BF variance for single-peaked neurons in ferrets is plotted as a function of the number of these neurons within an imaging field, for all imaging fields containing at least 3 single-peaked, 3 double-peaked, and 3 complex neurons. The statistics of a Pearson's correlation of these two values are included in the figure. **B**, For each ferret imaging field, we estimated the BF variance for each neuron class using only 3 neurons, and this calculation was repeated for every possible combination of 3 neurons. The bar plot shows the average (mean  $\pm$  SEM) BF variance across all iterations for each neuron type (t-test; \*\*\* $p < 0.001$ ). **C**, Cumulative probability of the bootstrapped BF variance within a ferret imaging field, for single-peaked (orange), double-peaked (blue) and complex (green) neurons. **D-F**, Results plotted as in A, B, C above (respectively), but for mouse data.



**Fig. 2.23 Effects of FRA complexity on BF organization hold across all imaged neurons.**

**A**, Proportion of neurons in each FRA class are shown for ferrets (red) and mice (blue), classified as in Figure 2.8C. The results of Likelihood Ratio tests are indicated above (\*\*\*)  $p < 0.001$ . **B**, Cumulative probability plots of the difference (octaves/mm) between the BF of each neuron imaged in ferrets and the median BF of all neurons in the same imaging field. Distributions were calculated separately for  $BF_s$  (orange),  $BF_d$  (dark blue), peak 2 of double-peaked neurons (light blue),  $BF_c$  (green), and  $BF_a$  (black). **C**, Results plotted as in B, but for neurons imaged in mice. **D-G**, Global tonotopic gradients in ferrets, presented as in Figure 2.7C.  $BF_s$  (D),  $BF_d$  (E), peak 2 (F), and  $BF_c$  (G) of individual neurons (black dots) are plotted against the neuron's position along the tonotopic axis. Red lines show the best single-term exponential fits to the data, and Pearson's correlations ( $r$ ) and their  $p$ -values ( $p$ ) are given in the top right corner of each plot.



# Chapter 3

## Functional properties of inhibitory neurons in auditory cortex

### 3.1 Introduction

“The functional superiority of the human brain is intimately bound up with the prodigious abundance and the unusual wealth of forms of the so-called neurons with short axons”

---

*Ramón y Cajal*

In this chapter, we continue our exploration of the tuning properties of auditory cortex (AC) in the ferret, with emphasis on the similarities and differences between the two main classes of neurons, the principal EXC (pyramidal neurons), and the local INH<sup>1</sup>.

Cortical inhibitory interneurons are comprised of a vastly diverse population, with cells differing both morphologically and physiologically. Whereas a vast number of INH sub-types can be recognized, the INH have been grouped into three predominant

---

<sup>1</sup>During the time of Cajal this was not yet known, but we now understand the “neurons with short axons” to be primarily inhibitory neurons.

classes based on molecular markers: the calcium binding protein parvalbumin (PV)<sup>2</sup>, the neuropeptide somatostatin (SOM), and the neurotransmitter 5-hydroxytryptamine 3a receptors (5HT), which include the vasoactive intestinal polypeptide-positive interneurons (VIP) (Lee et al., 2010; Rudy et al., 2011; Zeisel et al., 2015; Tasic et al., 2016; Kepecs and Fishell, 2014; Tremblay et al., 2016). It has only recently become possible to study in depth the role and properties of different types of INH in AC owing to the development of molecular, genetic and other technological advancements, most notably optogenetics (Boyden et al., 2005; Roux et al., 2014), allowing the targeting of genetically defined neuronal populations (Russell, 2011; Dimidschstein et al., 2016). Consequently, many of the recent studies that examine the properties and effects of interneuron subtypes on auditory processing have been performed in mice, where the genetic targeting tools are much more readily available in transgenic models (Moore and Wehr, 2013; Li et al., 2015; Mesik et al., 2015; Maor et al., 2016; Liang et al., 2019).

Before exploring the interactions between EXC and INH, the most straightforward analysis is to describe their frequency tuning properties. Reports from the literature measuring the tuning widths of individual PV, SOM, Vasoactive Intestinal Peptide neurons (VIP) and EXC so far did not yield clear distinctions between them. One study found that PV in mouse AC are well-tuned for frequency and the tuning widths did not differ from EXC (Moore and Wehr, 2013). Another study found that PV exhibited broader tonal receptive fields with lower intensity thresholds and stronger tone-evoked spike responses compared with SOM and EXC, while SOM and EXC did not differ from one another in these regards (Li et al., 2015). A third study looking at VIP and SOM found that both of them had broader frequency tuning comparing to EXC, but similar to each other (Mesik et al., 2015). Interestingly, many of the VIP exhibited strong intensity tuning characterized by “O-shaped” frequency response areas (FRAs). An in vivo patch-clamp study in anaesthetized rats that did not discriminate cell type found that the excitatory and inhibitory currents observed in response to different frequency-level combinations were very similar and there was no difference in the tuning width between the two types of currents in individual cells (Wehr and Zador, 2003). One potential caveat is that these classes of neurons are themselves comprised of multiple cell types as mentioned earlier. Indeed, a recent review of SOM estimated that there are up to 100 subtypes of SOM (Yavorska and Wehr, 2016). These differences

---

<sup>2</sup>Note that we will use GAD, PV, SOM, 5HT, VIP font for the protein markers and GAD, PV, SOM, 5HT, VIP font to indicate the types of neurons, to distinguish between proteins and neuronal types.

may be exacerbated across studies by the possible biases toward specific subclasses with different recording techniques. As we saw in the previous chapter, the biased sampling of different types of neurons in electrophysiology and calcium imaging could explain previous conflicting results in the literature.

Another important aspect of function across neuron classes is the relative timing of sound-evoked responses relative to other neurons. In one study, PV showed faster sound onset responses to pure tones than EXC neurons, whereas SOM exhibited slower response onsets than EXC neurons (Moore and Wehr, 2013; Li et al., 2015). Mesik et al. (2015) also found that PV tended to respond faster than EXC, while VIP and SOM had a delayed onset. When dissecting the individual current contributions, Wehr and Zador (2003) found that inhibition and excitation occurred in a precise and stereotyped temporal sequence: an initial barrage of excitatory input was rapidly quenched by inhibition, truncating the spiking response within a few milliseconds. Inhibition thus reduced the jitter of cortical tonal responses and increased the temporal precision.

How do the EXC and the different classes of INH work together to shape frequency tuning? In the AC, activating PV optogenetically resulted in narrowing of the receptive fields of EXC and an increase in the strength of their tone-evoked responses (Aizenberg et al., 2015; Hamilton et al., 2013), whereas PV suppression led to opposite effects in EXC (Aizenberg et al., 2015). Suppressing or activating SOM increased or decreased the firing rate of EXC, respectively (Seybold et al., 2015; Phillips and Hasenstaub, 2016). But the effect on EXC activity was more often multiplicative when SOM activity was reduced as compared with PV, whose suppression provided both multiplicative and linear shifts in excitatory neuronal responses to tones. Activation of PV or SOM also produced a mixture of multiplicative and linear shifts in EXC responses to tones. Particularly interesting is the study by Phillips and Hasenstaub (2016) which demonstrated asymmetric effects of optogenetic activation vs inactivation, where SOM inactivation increased response gain, while PV inactivation weakened tuning and decreased information transfer, implying that these neurons support delineable computational functions. But activating SOM and PV revealed no such differences, highlighting the difficulty in making causal claims and assigning function based on interventional approaches such as optogenetics. Manipulating VIP has shown a differential effect, consistent with disinhibition of EXC: activation of VIP caused an increased firing rate in EXC in response to pure tone stimuli (Pi et al., 2013). More specifically, activating cholinergic inputs to VIP broadened the tuning of EXC to tones of different frequencies by decreasing responses to the preferred

frequency and increasing responses to the less preferred stimuli (Nelson and Mooney, 2016).

Thus, if detailed knowledge about the properties and functions of INH in AC of mice is incomplete, it is almost completely absent in higher mammals such as carnivores, non-human primates (NHP) and humans. Although rodent research is invaluable in elucidating neuronal circuit mechanisms and the properties of genetically defined neuronal subpopulations, it does not always reveal the whole story regarding cortical function. Past research in visual cortex has shown us that there can often be substantial differences in the properties and functions of neurons in rodent and non-rodent higher mammals. For example, instead of pooling activity from local excitatory neurons as has been shown in rodents (Hofer et al., 2011; Kerlin et al., 2010; Scholl et al., 2015), Wilson et al. (2017) found that INH participate in functionally specific networks with co-tuned EXC at both local and long-range interaction distances. Therefore, as we saw in the previous chapter, it can be highly illuminating to examine the properties of neurons and neural circuits across species.

In this chapter, we set out to explore the stimulus response properties of EXC and INH in anaesthetized ferrets, and if possible to divide the inhibitory neurons further into subclasses based on known molecular markers (namely, PV and SOM). Given the lack of transgenic ferrets lines, it was necessary to first develop reliable electrophysiological and 2-photon calcium imaging procedures and analysis pipelines to separate the different types of neurons. Below, I describe the results of this process, going into detail behind the experimental and data analysis procedures involved. This is then followed by a direct comparison of the tuning properties of EXC and INH in ferret AC and some analyses regarding their interactions.

## 3.2 Materials and methods

### 3.2.1 Animals

All animal procedures were approved by the local ethical review committee of the University of Oxford and performed under license from the UK Home Office. Two adult female ferrets (*Mustela putorius furo*; Marshall BioResources, UK) were used in the two-photon imaging experiments, and six adult ferrets (two male) were used in the electrophysiology experiments. These are the same animals described in Chapter 2.

### 3.2.2 Viral vector injections

The viral vector injection procedure is described in subsection 2.2.2 of Chapter 2 (see “Ferret surgery”). In the present Chapter, we used only the ferrets that had the injection of a 1:1 solution of AAV1.Syn.GCaMP6f.WPRE.SV40 (Addgene) and AAV1.mDlx.GFP-GCaMP6f-Fishell-2.WPRE.SV40 (Penn Vector Core).

### 3.2.3 In vivo two-photon calcium imaging

The imaging procedure is described in Chapter 2, subsection 2.2.3 (see “Ferret surgery” and “Two-photon imaging”).

### 3.2.4 Sound presentation

The sound presentation is described in Chapter 2, subsection 2.2.4. The sound stimuli used are the pure tones described for ferrets 6 and 7 in Table 2.1.

### 3.2.5 Histology

At the end of each imaging session, experimental animals were overdosed (Euthatal, 1ml pentobarbital sodium, i.p.) and perfused transcardially, first with 0.01M PBS and heparin (20 units/ml), and then with 4% paraformaldehyde in PBS. Brains were removed and placed in 4% paraformaldehyde for two hours, after which they were stored in PBS with 0.01% sodium azide. Ferret brains were cryoprotected 2-14 days after perfusion in a 1:3 solution of sucrose and PBS for 24 hours. Sections (60 $\mu$ m thickness) were cut parallel to the in vivo cranial window surface using a freezing sliding microtome (Leitz Wetzlar). This cutting orientation helped us to identify the same neurons that were previously imaged in vivo. Sections were washed once in PBS for 10 mins, after which an antigen retrieval solution (“Antigen Retrieval Reagent-Universal”, R&D Systems) was applied for 3 mins at a temperature of  $\sim$ 95°C. Sections were then washed in PBS 3 times for 10 mins. After that they were incubated in blocking buffer (5% Donkey Serum Albumin and 0.5% Triton X-100) for 2 hours at room temperature. Sections were then incubated for 2 hours at room temperature and then at  $\sim$ 4°C for 3 days with blocking buffer containing primary antibodies (Goat anti-parvalbumin at 1:2500, Swant PVG-214; Rat anti-somatostatin at 1:125, Millipore MAB 354; and

Mouse anti-GAD67 at 1:1000, Millipore MAB5406). Sections were stained with all three primary antibodies present simultaneously. Sections were then washed in PBS 3 times for 10 mins. They were next incubated for 2 hours at room temperature, followed by  $\sim 4^{\circ}\text{C}$  overnight in blocking buffer containing 1:500 secondary antibodies (Alexa 647 donkey anti-mouse, Molecular Probes, A31571; Alexa 568 donkey anti-rat, Abcam AB175475; and Alexa 405 donkey anti-goat, Abcam AB175665), all mixed together. Finally, sections were washed 3 times in PBS for 10 mins and were mounted onto microscope slides using Vectashield mounting medium (Vector Laboratories Ltd., USA). A coverslip was placed on the slide for imaging and sealed with clear nail polish.

**Table 3.1 Key histological resources**

Reagents	Dilution	Source	Identifier
Goat anti-PV	1:2,500	Swant	Swant Cat# PVG-214 RRID: AB_2313848
Rat anti-SOM	1:125	Millipore	Millipore Cat# MAB354 RRID: AB_2255365
Mouse anti-GAD67	1:1,000	Millipore	Millipore Cat# MAB5406 RRID: AB_2278725
Alexa 647 donkey anti-mouse	1:500	Molecular Probes	Molecular Probes Cat# A-31571 also A31571 RRID: AB_162542
Alexa 568 donkey anti-rat	1:500	Abcam	Abcam Cat# AB175475; RRID: AB_2636887
Alexa 405 donkey anti-goat	1:500	Abcam	Abcam Cat# AB175665; RRID: AB_2636888
Antigen Retrieval Reagent-Universal	1:10	R&D Systems	Cat# CTS015

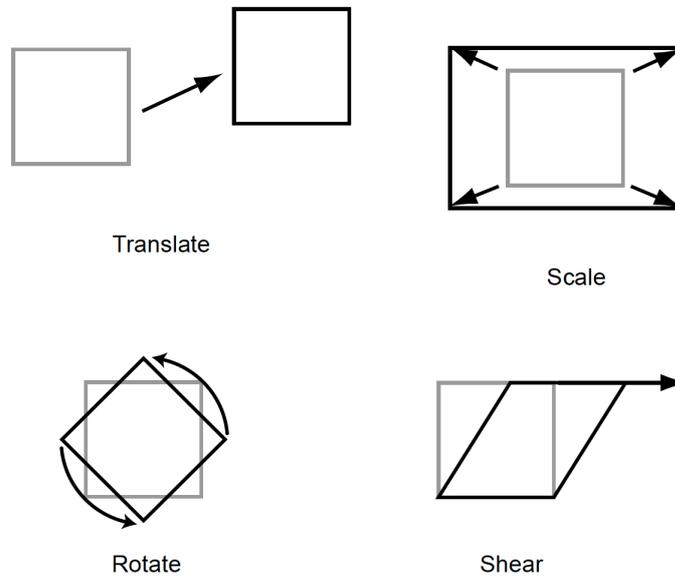
### 3.2.6 Confocal imaging

Brain sections were imaged using an inverted Olympus FV3000 six laser line spectral confocal microscope fitted with high sensitivity gallium arsenide phosphide (GaAsP detectors) and a 4x, 0.16 NA UplanSApo objective using 405, 488, 561, and 633 nm laser lines with emission channels optimized for each fluorophore (Alexa 405, GCaMP6f, Alexa 568, and Alexa 647) while minimizing cross-talk between channels. The confocal pinhole was set to one airy unit to optimize optical sectioning with emission collection. Images were collected in resonant scanning mode at 512x512 pixels (pixel size  $6.21\mu\text{m}$ ) and 16x averaging. Tile scans were stitched using Olympus FluoView software.

### 3.2.7 Image registration

Images of the brain surface were acquired across the craniotomy at the beginning of the in vivo two-photon imaging session. These images were then tiled together and stitched to reconstruct the surface of AC and locate the imaged fields using a custom Matlab script. The tiled AC was then aligned with boundaries of the sulci in the auditory cortical template (see Chapter 2, Figure 2.16). After the imaging experiment was complete, this stitched image was used together with the acquired in vivo imaging planes and z-stacks of the recording sites to align them to confocal z-stacks. We determined the approximate location of the injection site using GCaMP6f fluorescence and then used blood vessel patterns and cellular morphology to identify the imaging site within the stitched in vivo image and confocal image of the brain surface. We matched at least three points in the confocal z-stack on the GCaMP6f channel to points in the in vivo imaging plane to obtain a three-dimensional affine transformation matrix. Here we will briefly explain this process and the principles behind it.

Because of the requirements of matching our confocal 3D z-stack to the in vivo 2-photon imaging plane, we need to apply a set of transformations to the confocal z-stack such that it matches the imaging plane as closely as possible. The set of operations providing for all such transformations, are known as the affine transforms. The affine transformations include translations and all linear transformations, like scaling, rotation, and shear (see Figure 3.1). For any given reconstruction case, it might be necessary to apply a number of these different transformations in order to achieve satisfactory results. Each affine transformation can be compactly summarised by a matrix which then can be applied to the set of 3D points we are interested in transforming:  $\mathbf{v}' = M\mathbf{v}$ . Since the matrix form is so handy for building up complex transforms from simpler ones, it would be very useful to be able to represent all of the affine transforms by matrices. The problem is that translation is not a linear transform. The way out of this dilemma is to turn the 3D problem into a 4D problem, but in homogeneous coordinates. This simply means that a vector in 3D  $\mathbf{v} = (x, y, z)$  has to be transformed to a 4D one by the addition of a 4th coordinate,  $\mathbf{v} = (x, y, z, 1)$ . By convention, we call this 4th coordinate the w coordinate, to distinguish it from the usual 3D z coordinate. We also extend our 3D affine matrices to 4D homogeneous form by appending an extra row and column, thus allowing us to preserve linearity. Thus,



**Fig. 3.1 Types of affine transformations.**

Here we represent the four basic types of affine transformations: translation, scaling, rotation and shear. Note that here we use 2D transformations for ease of representation but the same principles will apply in 3D. Translation moves a set of points a fixed distance in x, y and z. Scaling scales a set of points up or down in the x, y and z directions. Rotation rotates a set of points about one of the 3 cardinal axes: x-rotation, y-rotation or z-rotation. A shear offsets a set of points a distance proportional to their x, y and z coordinates. Each affine transformation is defined by a transformation matrix (see text for more details) and because of the linearity property, a sequence of transformations can be represented by a single matrix which is the product of the individual matrices. This figure was adapted from the Computer Science [lecture course](#) at Clemson University, South Carolina.

armed with these transformations we arrive at the following transformation matrices of 3D space for our 4 cases of translation, scaling, shear and rotation (about x, y and z):

$$T = \begin{bmatrix} 1 & 0 & 0 & \Delta x \\ 0 & 1 & 0 & \Delta y \\ 0 & 0 & 1 & \Delta z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3.1)$$

$$\mathbf{S} = \begin{bmatrix} s_x & 0 & 0 & 0 \\ 0 & s_y & 0 & 0 \\ 0 & 0 & s_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3.2)$$

$$\mathbf{H} = \begin{bmatrix} 1 & h_{xy} & h_{xz} & 0 \\ h_{yx} & 1 & h_{yz} & 0 \\ h_{zx} & h_{zy} & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3.3)$$

$$\mathbf{R}_x = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\Theta_x & -\sin\Theta_x & 0 \\ 0 & \sin\Theta_x & \cos\Theta_x & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3.4)$$

$$\mathbf{R}_y = \begin{bmatrix} \cos\Theta_y & 0 & \sin\Theta_y & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\Theta_y & 0 & \cos\Theta_y & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3.5)$$

$$\mathbf{R}_z = \begin{bmatrix} \cos\Theta_z & -\sin\Theta_z & 0 & 0 \\ \sin\Theta_z & \cos\Theta_z & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3.6)$$

Thus, e.g. if we wanted to apply a scaling (S), followed by a sheer (H), rotation in x ( $\mathbf{R}_x$ ) and translation (T) to a point in 3D space  $\mathbf{v}$ , we would describe this by the following equation:

$$\mathbf{v}' = \mathbf{TR}_x\mathbf{HSv} \quad (3.7)$$

And because of linearity and homogeneity, this can be compactly represented by a single affine composite matrix M:

$$\mathbf{v}' = \mathbf{Mv} \quad (3.8)$$

In this way, by identifying points of correspondence between the confocal z-stack and the in vivo imaging plane, we are able to arrive at a single affine transformation matrix  $M$ . This matrix was applied to the entire confocal z-stacks, including the channels for PV, SOM and GAD (Alexa 405, Alexa 568 and Alexa 647), in order to align them to the in vivo imaging fields. In some cases, the in vivo imaging plane spanned more than one physical confocal slice. In those cases, the parts of the in vivo plane that match two neighbouring confocal sections were registered separately and combined afterwards. Two-photon imaged cells were then manually identified in the corresponding confocal images, and assigned to cell classes based on their immunostaining.

### 3.2.8 Analysis of two-photon imaging data

Many parts of the data analysis were the same as those described in Chapter 2, subsection 2.2.8 (see “Isolating the response of single neurons from imaging fields”, “Calculating best frequency and classifying the frequency response”, “Quantification of local BF variability”, and “Signal and noise correlations”). Here, I will highlight those analysis that are unique to the present chapter.

#### Identifying responsive and selective neurons

To identify responsive, frequency selective, and level selective neurons, neuronal responses to sound presentation were quantified within a response time window starting at stimulus onset and lasting twice as long as the stimulus. The “evoked activity” on a given trial was defined as the average of the inferred spike probability trace within the response window. To be defined as sound “responsive”, a neuron needed to show a significant effect in at least one t-test comparing activity during stimulus presentation and the period of silence immediately preceding stimulus presentation, for tones presented at any one of the frequency-level combinations ( $\alpha = 0.01$ ). A two-way ANOVA, with tone frequency and sound level as predictors, was used to determine if the evoked activity was significantly modulated by sound frequency or intensity ( $\alpha = 0.05$ ). Neurons showing a significant main effect of frequency or a frequency/level interaction were defined as “frequency selective”. Similarly, being “level selective” required a significant main effect of level ( $\alpha = 0.05$ ). Finally, we included a less conservative inclusion criteria of “responsive or selective”. These neurons needed only to satisfy any one of the criteria for frequency selectivity (ANOVA), level selectivity (ANOVA), or sound responsiveness (t-test) described above. In theory, a neuron that is frequency or

level selective, will also necessarily be responsive to sound. However, not all neurons which passed the ANOVA for selectivity also passed the t-test for responsivity or vice versa, because of statistical noise. Given the small number of INH present, this more inclusive response category allowed us to exclude less neurons from our analyses, while making sure that the neurons were modulated by the sound stimuli in some way.

### 3.2.9 In vivo electrophysiology experiments

The in vivo electrophysiology experimental procedures are described in Chapter 2, subsection 2.2.9 (see “Data acquisition” and “Sound presentation”).

### 3.2.10 Analysis of electrophysiology data

Again some analyses performed here were the same as those described in Chapter 2 subsection 2.2.9 (see “Spike sorting” and “Inverse Current Source Density analysis”). Only analyses that are not described in Chapter 2 are explained here.

#### Calculating the Characteristic Frequency, Q10 and Q30

We followed a procedure previously described by Bizley et al. (2005) to estimate the Characteristic Frequency CF, Q10 and Q30 of each neural unit. Briefly, frequency response areas (FRAs) were constructed for each neural unit by summing activity that occurred during a response window (100ms, the length of each pure tone presentation, starting from tone onset) for all repetitions of each frequency–level combination. This spike matrix was smoothed with a nine-point running hanning window. This smoothing window has an equivalent rectangular area of four bins, and is therefore roughly equivalent to the use of a square smoothing window with an edge of two bins. A neural unit was deemed to have responded to a given frequency–level combination if the average spike rate within the FRA was higher than a criterion firing level. The criterion level was defined as the mean spontaneous firing rate (i.e. the average firing rate in a 100ms window immediately before tone onset) plus 20% of the peak firing rate (Sutter and Schreiner, 1991). This criterion level was used to construct a “threshold curve” (see Figure 3.13A, white line), from which other parameters, CF, Q10 and Q30, were extracted. CF was defined as the frequency that elicited a response at the lowest sound level. Where more than one frequency was effective at threshold, the logarithmically weighted mean of the frequencies was used as CF. The Q10 value is the CF divided by

the bandwidth calculated 10dB above the CF level. Bandwidth was calculated as the difference between the highest and lowest frequencies that pass the threshold criterion level at 10dB above the CF level (see Figure 3.13A, black line). The Q30 value is very similar to Q10, but here CF is divided by the bandwidth determined for tones 30dB louder than CF. These algorithm were performed automatically by MATLAB. All FRAs were individually examined and, in the vast majority of cases, this algorithm produced results that corresponded to the CFs estimated by visual inspection.

### Separation of putative excitatory and inhibitory neurons

For the separation of putative excitatory and inhibitory neurons, only single units<sup>3</sup> were used (see subsection 2.2.9, “spike sorting”). For each single unit, the average spike waveform was calculated from 300 randomly chosen recorded spikes from the raw data high-pass filtered at 150Hz and then this averaged waveform was interpolated 100-fold to produce a smoother estimate of the spike features. Five spike features were extracted from the resulting waveform, as summarised in Table 3.2 and Figure 3.2.

**Table 3.2 Spike features used for clustering**

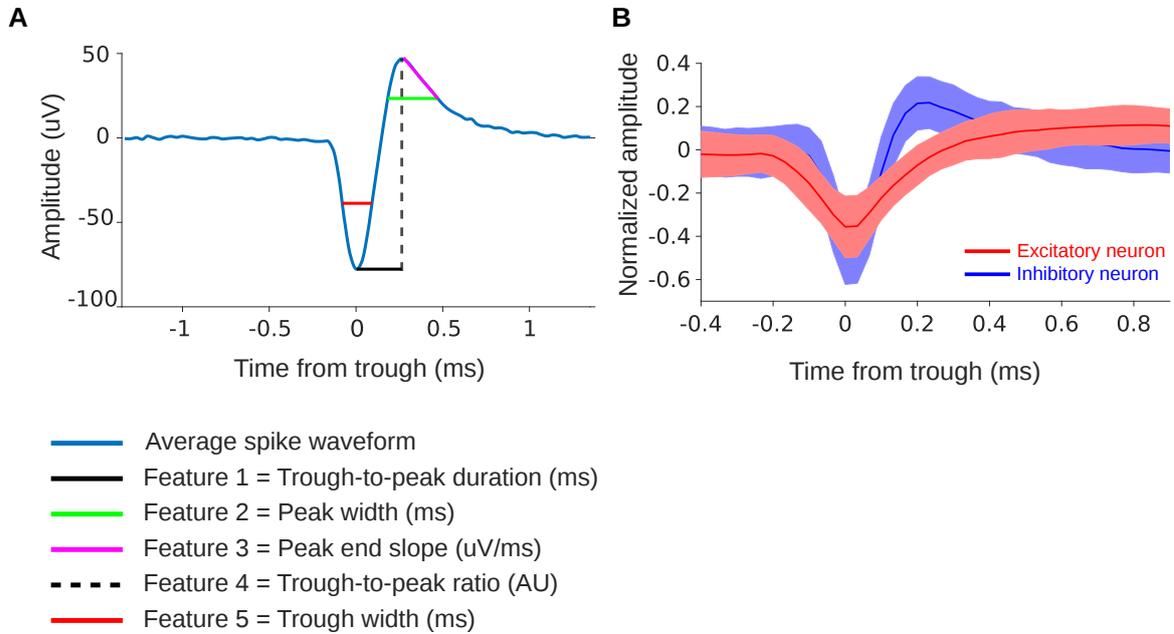
Spike Feature	Spike Property	Units
1	Trough-to-peak duration	ms
2	Peak width	ms
3	Peak end slope	$\mu\text{V}/\text{ms}$
4	Trough-to-peak ratio	AU
5	Trough width	ms

In order to assign the identity of putative EXC and INH, we ran a k-means clustering algorithm to cluster 2 groups of neurons based on pairs of spike features. To assess the quality of the clustering output for different pairs of features, we used the silhouette score measure (Rodriguez et al., 2019) to compute an overall mean clustering score for each feature pair across all clustered single units (see Figure 3.4).

To test for putative interactions (synaptic connections) between pairs of neurons, we analysed only neurons that were recorded together within the same animal and penetration. Firstly, to corroborate our classification approach we computed Crosscorrelogram (CCG)s for all pairs of neurons within a given penetration. We

---

<sup>3</sup>Note that we will use the terms “single units” and neurons interchangeably.



**Fig. 3.2 Separating excitatory and inhibitory neurons using extracellular spike waveform features.**

**A**, Example extracellular spike waveform, showcasing the 5 waveform features we used to separate putative excitatory and inhibitory neurons. These included: “Trough-to-peak duration” (Feature 1, black line), the time difference (ms) between the peak (hyperpolarization) and trough (depolarization) of the spike; “Peak width” (Feature 2, green line) which measures the width in ms at half the maximum value of the spike peak; “Peak end slope” (Feature 3, magenta line), the slope of the spike peak following the maximum positive voltage deflection ( $\mu\text{V}/\text{ms}$ ); “Trough-to-peak ratio” (Feature 4, black dotted line), the ratio of the amplitude of the trough to that of the peak; and “Trough width” (Feature 5, red line), the width in ms at half the minimum value of the spike trough. **B**, Example waveforms are shown from a putative excitatory neuron (regular spiking, broad waveform) in red and a putative inhibitory neuron (fast spiking, narrow waveform) in blue. For each neuron, 300 randomly chosen raw spike waveforms were obtained using extracellular recordings with microelectrodes in ferret AC. The solid line shows the mean waveform, and the shaded area shows the standard deviation. Waveforms are temporally aligned by their troughs (i.e. their minimum amplitude).

then used established methods in order to detect statistically significant temporal bias in a neuronal pair's relative spike timing indicative of putative monosynaptic connections (Fujisawa et al., 2008; Peyrache et al., 2012). This was implemented using previously published software available from the laboratory of György Buzsáki (<https://github.com/buzsakilab/buzcode>). Briefly, the spikes of every neuron were jittered by adding a random value (from a normal distribution with a 10-ms standard deviation and 0 mean) to the spike times. For each neuronal pair, 1,000 jittered spike trains are created, and the expected CCG (and 99.9% confidence interval) are estimated on 1-ms time bins under the null hypothesis of no monosynaptic effects between the two cells (Figure 3.6B, green dotted lines). For any given cell pair where at least one bin in the [1 ms, 4 ms] interval exceeded (putative excitatory, Figure 3.6B left) or fell below (putative inhibitory, Figure 3.6B right) the 99.9% confidence interval, the interaction is considered monosynaptic. A final examination of the cell pair CCG is carried out to remove noisy pairs, as this procedure does lead to some false positive results. For example, putative excitatory connections can result from oscillations that are sometimes induced by the anaesthesia, while putative inhibitory connections (dips in the CCG), albeit significant, may result from the refractory period of a narrow waveform neuron following excitation by a broad waveform neuron.

### 3.3 Results

#### 3.3.1 Separating putative excitatory and inhibitory neurons based on extracellular spike waveforms

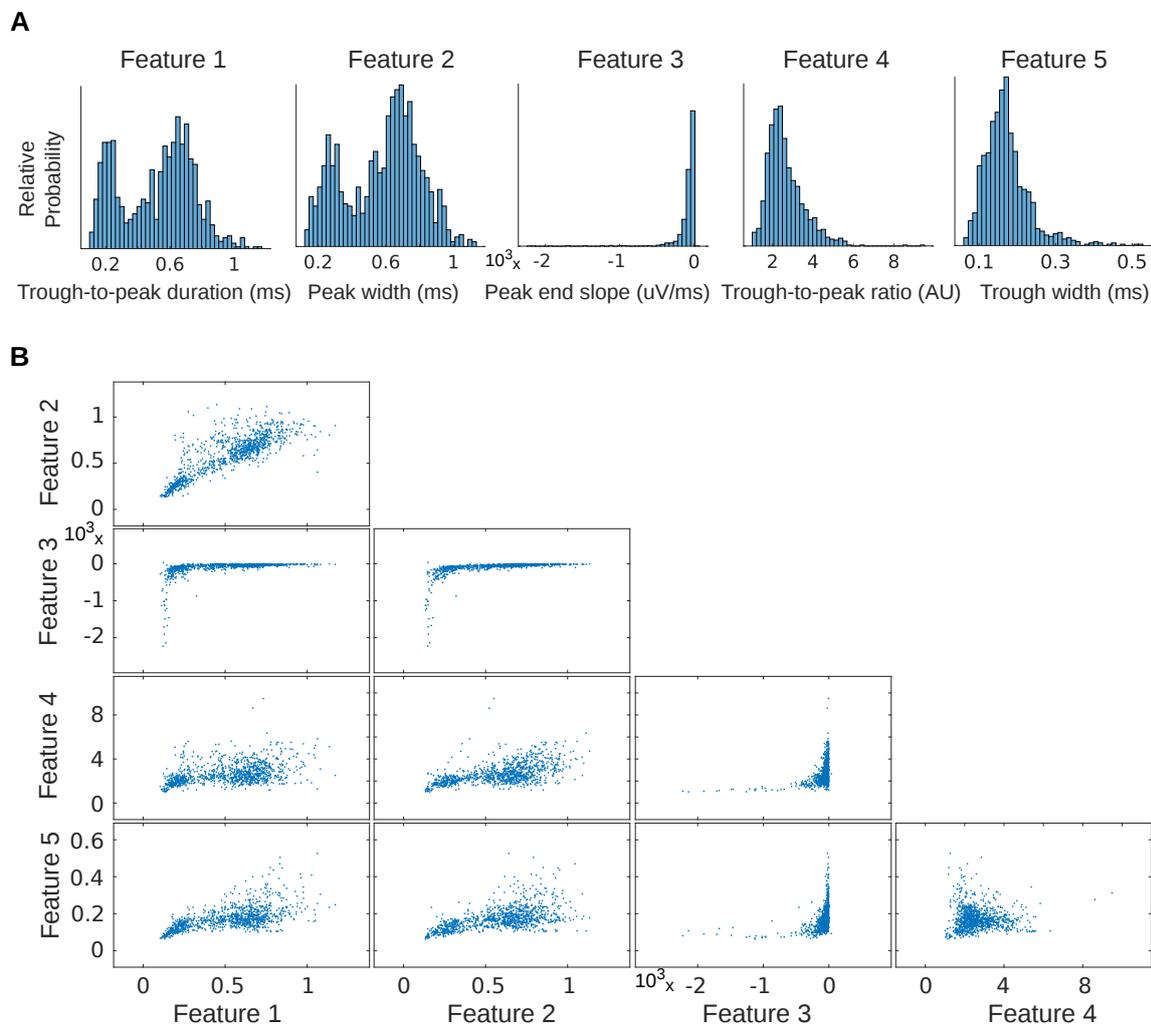
The first approach we took was based on microelectrode recordings using the newly developed Neuropixels probes (Jun et al., 2017; Steinmetz et al., 2020). The probe shank has 384 densely spaced electrodes (“pixels”) which allow for the simultaneous recording and isolation of tens up to hundreds of single units. In addition, due to the dense sampling and high acquisition rate it is possible to capture extracellular spike waveforms in great detail. This in turn facilitates the extraction of various spike features which can then be used to separate putative EXC and INH based on those spike features. From the existing literature, we know that there are differences in the spike waveform properties (Peyrache and Destexhe, 2019) of EXC and INH. Based on their electrophysiological properties, neurons can be broadly divided into regular-spiking

neurons with broader waveforms (putative EXC) and fast-spiking neurons with narrow waveforms (putative INH) (Barthó et al., 2004; Niell and Stryker, 2008; Peyrache et al., 2012; Insel and Barnes, 2015; Mendoza et al., 2016), as shown in Figure 3.2A. This is largely due to a briefer hyperpolarisation, measured by widths of the extracellular recorded peak at half of the amplitude (Kawaguchi, 1993) and more rapid peak rates of repolarisation, measured by the gradient of the slope returning to baseline (McCormick et al., 1985) for INH compared to EXC (Snyder et al., 2016). Our aim was to find the spike features that allowed us to most consistently separate the putative EXC and INH, and ideally further corroborate the usefulness of these features using some functional measure of connectivity.

We chose to examine five spike waveform features that have been commonly used in the literature to separate putative EXC and INH (Sirota et al., 2008; Peyrache et al., 2012; Stark et al., 2013; Medrihan et al., 2017; Yu et al., 2019). These features are illustrated in Figure 3.2A and summarised in table 3.2 (see Methods 3.2.10).

If the spike features differ substantially between putative excitatory and inhibitory neurons, one would expect them to have bimodal distributions, corresponding to the two types of neurons. To investigate this, we pooled data from 6 animals (16 penetrations; 1,106 single neurons). We limited our analysis to areas belonging to primary AC, the A1 and AAF cortical fields in ferrets (see Chapter 1, Figure 1.4B), as neurons in secondary areas can have substantially different stimulus response properties, as described in Chapter 1, subsection 1.3.2 (Bizley et al., 2005; Elgueda et al., 2019). The cortical area of each penetration was determined based on its anatomical location in the ferret ectosylvian gyrus, the local field potential response latency, and the FRA shapes of neurons recorded, as described in Chapter 2. Using current source density (CSD) analysis (see Chapter 2, subsection 2.2.9, “Inverse Current Source Density analysis”), we also limited our analysis to neurons recorded in layers 2/3 of AC, so as to better compare our microelectrode data to our calcium imaging data.

As we can see in Figure 3.3A the “Peak-to-trough duration” (Feature 1) and the “Peak Width” (Feature 2) both displayed clearly bimodal distributions. What is more, the second peak in the histograms of features 1 and 2 was higher and there are more neurons associated with longer Trough-to-peak durations and Peak widths. This is consistent with the existing literature, as there are more excitatory than inhibitory neurons in cortex, with a ratio of roughly 4:1 (Kepecs and Fishell, 2014; Tremblay et al., 2016). The other features (3-5), however, are largely unimodal in distribution, as



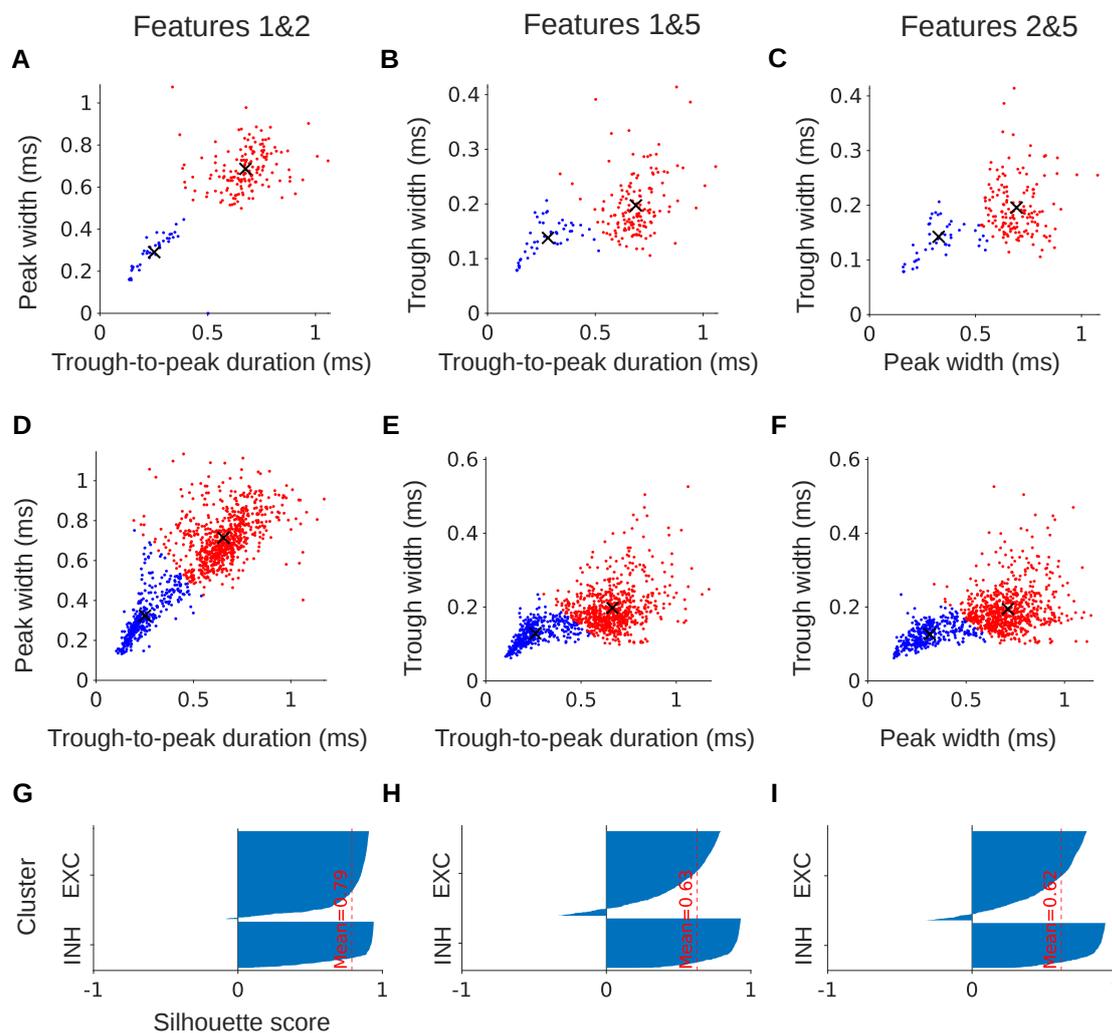
**Fig. 3.3 Distribution of spike features across neurons.**

**A**, Histograms of the 5 spike features extracted from the neuronal waveforms: Trough-to-peak duration (Feature 1); Peak width (Feature 2); Peak end slope (Feature 3); Trough-to-peak ratio (Feature 4); Trough width (Feature 5). The histograms were constructed from data pulled across penetrations and animals (16 penetrations from 6 ferrets;  $n=1,106$  neurons). Features 1 and 2 had a clear bimodal distribution. **B**, Scatter plots of the 10 possible pairs of the 5 spike features plotted above. Each dot represents the average spike waveform from a single neuron.

we can see in their histograms (Figure 3.3A). In order to explore how the spike features relate to one another, we plotted scattergrams of feature pairs as shown in Figure 3.3B. Out of the 10 possible pairs of the 5 spike features, the pairs involving features 1, 2 and 5 seemed to be the most promising. Features 1 and 2 displayed a clear bimodal distribution (Figure 3.3A), while feature 5 seem to visually provide a broad spread in conjunction with features 1 or 2 (Figure 3.3B). Therefore, our further analyses will examine the clustering of neurons based on these 3 pairs of features: features 1 and 2; features 1 and 5; and features 2 and 5.

In order to separate the single units into putative EXC and INH from their spike features, we wanted to use an automatic, unbiased way of classifying them into 2 groups and then visualise them on a 2-dimensional plot. We performed k-means clustering in individual penetrations using the three most promising spike features (1, 2 and 5), and the results of one example penetration are shown in Figure 3.4A-C. In this example penetration, it is visually evident that features 1 and 2 performed the best as they provided the widest separation of the putative two groups with little overlap. We also performed the k-means clustering on all of the data pooled across penetrations, as shown in Figure 3.4D-F. Here again, features 1 and 2 seemed to give the best separability and they also minimised the overlap between the two groups, although due to the high number of neurons this is less clear.

In order to quantitatively confirm our visual inspection of the ability of different spike features to separate the putative EXC and INH, we applied the silhouette score measure (Rodriguez et al., 2019). The silhouette score computes, for each neuron, a width depending on its membership inside a cluster relative to that between clusters. The silhouette value ranges from  $-1$  to  $1$ . A high silhouette value indicates that a given point is well matched to its own cluster, and poorly matched to other clusters. If most points have a high silhouette value, then the clustering solution is appropriate. If many points have a low or negative silhouette value, then the clustering solution does not provide an adequate separation. As we can see from Figure 3.4G-I, most points have a large positive silhouette score for features 1 and 2 (mean = 0.79), while features 1 and 5 and 2 and 5 have smaller scores on average (mean scores of 0.63 and 0.62, respectively). Taken together, these data indicate that Trough-to-peak duration and the Peak width are the spike features which allow for an excellent separation of extracellularly recorded neurons into two distinct clusters. It is tempting to speculate

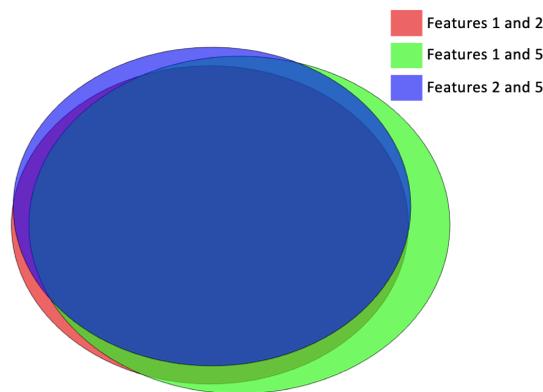


**Fig. 3.4 K-means clustering provides good separation of neuronal groups based on their spike features.**

**A-C**, Scatter plots in which colours indicate the 2 clusters of neurons assigned by k-means clustering in a single penetration, based on three different spike features: **A**, features 1 and 2 (Trough-to-peak duration and Peak width); **B**, features 1 and 5 (Trough-to-peak duration and Trough width); **C**, features 2 and 5 (Peak width and Trough width). Each dot represents the average spike waveform from a single neuron. The red dots are neurons which we assigned to the putative excitatory cluster, while the blue dots are those which we assigned to the putative inhibitory cluster based on the k-means algorithm and our expectation of their respective spike properties. The black crosses represent the centroids of the two clusters. **D-F**, K-means clustering using pooled data from all animals and penetrations (16 penetrations, 6 animals, 1,106 neurons) using the same features as those shown in A-C, and plotted as in A-C above. **G-I**, Silhouette scores to assess the clustering quality for the 3 different feature pairs used in D-F above. The silhouette score for every neuron in each of the two clusters (EXC - putative excitatory, INH - putative inhibitory) is shown, ordered from highest to lowest score. The mean value across all neurons for each feature pair is indicated by the red dotted line. The mean silhouette scores are: features 1 and 2 = 0.79; features 1 and 5 = 0.63; features 2 and 5 = 0.62.

that these two clusters correspond to putative EXC and INH, but further validation is necessary to confirm this with high certainty.

Even though features 1 and 2 gave the best separation of neurons, there was an overall good agreement in the neuron classifications across the three feature pairs. In total, 91% of the neurons were classified into the same clusters (putative EXC and INH) across all three feature pairs (Figure 3.5). This gives us confidence that the separation between neuronal classes is fairly robust.



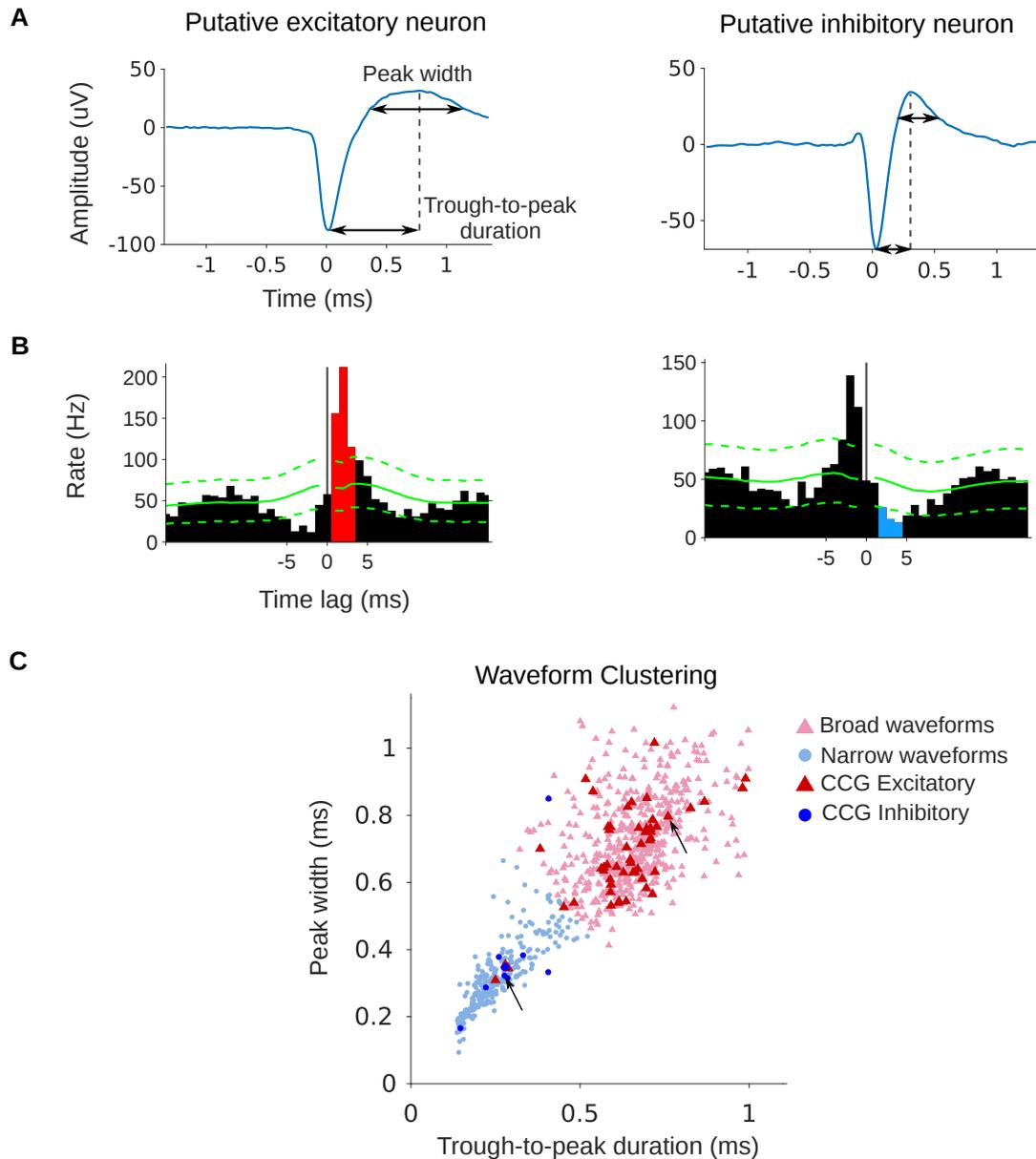
**Fig. 3.5 High degree of overlap in the clustering of neurons.**

Venn diagram showing the overlap in clustering using the 3 different feature pairs: features 1 and 2 (Trough-to-peak duration and Peak width) in red; features 1 and 5 (Trough-to-peak duration and Trough width) in green; features 2 and 5 (Peak width and Trough width) in blue. After performing k-means clustering using these features (Figure 3.4D-F), the amount of overlap in assigning neurons to the putative excitatory (EXC) or inhibitory (INH) clusters was determined. There was good agreement in the classification across the three feature pairs: 91% of neurons were classified into the same cluster across all three feature pairs.

We would also like to validate that our clustering approach is related to excitatory and inhibitory classes of neurons using some sort of functional connectivity measure or a reasonable proxy for it. We would expect that given the dense sampling of the Neuropixel probes and the large number of single units in a given penetration, it might be possible to detect interactions between EXC and INH. In the cortical network, EXC would tend to stimulate and therefore increase the firing rates of other neurons, while INH would usually decrease firing in EXC.

A simple yet powerful approach to measuring these interactions is to use cross-correlograms (CCGs). The CCG method tries to detect monosynaptic connections from

the spike data using statistical approaches (Fujisawa et al., 2008; Peyrache et al., 2012). A CCG measures the correlation between the spikes of a pair of neurons at various time lags. One neuron is taken to be the “reference neuron”, while the spiking of the other neuron is taken with respect to this reference. Roughly, the CCG tells us: How likely is the second neuron to spike relative to the first neuron at various time points, before, during and after the spiking of the first neuron? The logic behind the CCG analysis is that if two cells are connected monosynaptically, one would expect to see an increase in the case of excitatory connection and decrease in the case of an inhibitory connection in the firing rate of the second neuron with respect to the reference neuron in the causal direction (positive time lags) within a short period of time (1-4ms) consistent with a monosynaptic connection delay (see Figure 3.6 and subsection 3.2.10 for more details). Using this approach in our data, we found that 50 out of 53 putative EXC and 10 of 11 putative INH were categorized into the same neuronal type based on CCG and the above k-means clustering analysis (Figure 3.6C). These neurons were selected based on the CCG approach showing a significant interaction with each other. This agreement between the two methods supported our classification of individual neurons based on the clustering approach.



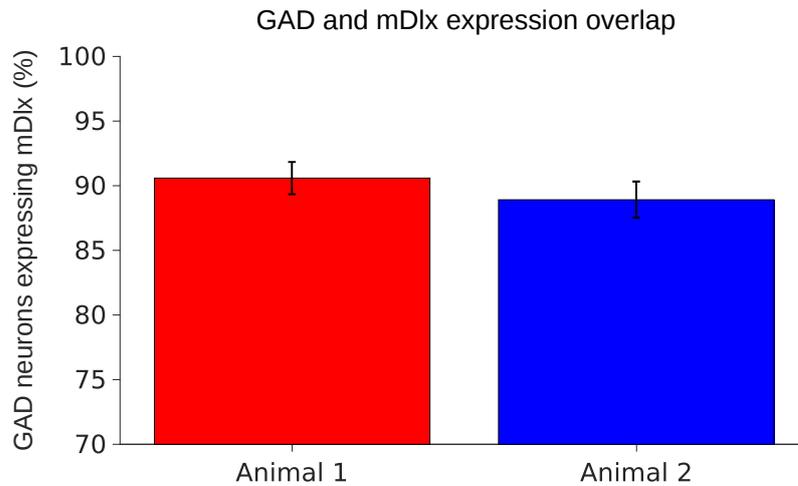
**Fig. 3.6 Functional confirmation of the spike feature clustering.**

**A**, Example spike waveforms from two neurons, putative excitatory (left) and putative inhibitory (right). The plots represent averages of 300 randomly chosen waveforms. The Trough-to-peak duration (feature 1) and Peak width (feature 2) features used for clustering are indicated. **B**, Cross-correlograms (CCGs) for the the same neurons as in **A** above, which corroborate their identity as putative excitatory and inhibitory neurons. The CCGs are binned at 1ms, green dotted lines show the 99.9% confidence interval from jittered spike trains, bins which where higher or lower than these bounds in the 0-4ms time window are indicated in red or blue, respectively. A neuron was identified as excitatory or inhibitory if it had at least one significantly higher or lower bin (respectively) in the time window 0-4ms, consistent with monosynaptic transmission delay. **C**, K-means clustering across all penetrations and animals (same data as Figure 3.4D) based on features 1 and 2. Putative excitatory neurons identified by the clustering are presented as light red triangles, inhibitory ones by light blue circles. The excitatory neuron labels based on the CCG analysis are represented by dark red triangles, the inhibitory ones by dark blue circles. The location of the two neurons from **A** and **B** are indicated by the black arrows.

### 3.3.2 Separating excitatory and subclasses of inhibitory neurons using imaging approaches

In addition to being able to separate neurons into the broad categories of excitatory and inhibitory types, it can be even more informative to explore the properties of different subclasses of INH. Often the different genetic, molecular, morphological and structural differences of neurons lead to interesting and meaningful functional differences and the kind of computations that they can perform (Tremblay et al., 2016). For example, PV are believed to target mostly the soma, proximal dendrites and the axon initial segment of other neurons while SOM converge mostly on the distal dendrites of their postsynaptic targets. This has implications for the type of inhibition they can provide, as PV are thought to implement divisive normalization, whereas SOM might perform relatively uniform subtraction of responses in their targets, leading to complementary effects on neuronal responses (Wilson et al., 2012; Sturgill and Isaacson, 2015; Natan et al., 2017; Bos et al., 2020). Therefore, we set out to separate EXC from Gamma-Aminobutyric acid neurons (GAD) in addition to PV and SOM inhibitory subclasses in the ferret primary AC as well as recording, analysing and comparing their responses to pure tones of different frequencies.

To begin with, we performed 2-photon calcium imaging in the ferret AC while presenting pure tones of different frequencies and sound intensities, as described in Chapter 2 (see subsections 2.2.3 and 2.2.4). Following previous research in ferret visual cortex (Wilson et al., 2017) and our own experience, we knew that the widely used AAV1.hSyn.GCaMP6f.WPRE.SV40 2-photon imaging construct would primarily target the EXC of AC. Although in theory INH should also possess the Human synapsin 1 (hSyn) gene promoter (Kügler et al., 2003), in practice expression of the GCaMP6f calcium indicator is much less efficient in INH as pointed out by Wilson et al. (2017). The efficiency and relative rate of transfection can vary between EXC and INH based on the concentration of the viral solution (Haggerty et al., 2020). Therefore, we used the AAV1.hSyn.GCaMP6f construct along with the AAV1.mDlx.GFP-GCaMP6f-Fishell-2 construct simultaneously. The mDlx promoter targets INH specifically and thus ensures efficient expression of the calcium indicator in these neurons (Dimidschstein et al., 2016). The Dlx transcription factors are classically thought of as controlling the migration of GABA interneurons. The role of Dlx in the adult neocortex is less well studied, although the contribution of Dlx1 to subtype specific interneuron survival has been previously reported (Cobos et al., 2005). It has been reported that the Dlx enhancer



**Fig. 3.7** Overlap in expression between GAD and mDlx-GCaMP.

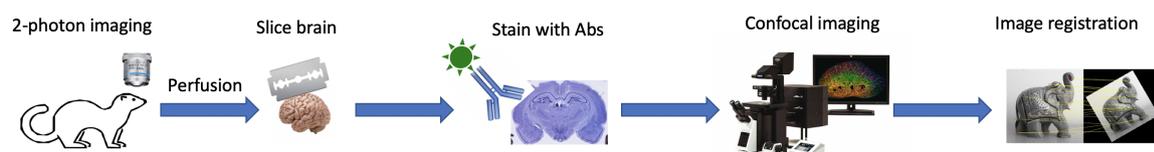
We wanted to quantify the proportion of GAD-positive neurons that also expressed mDlx-GCaMP. We counted those in the two animals used in this chapter from slices that were stained with anti-GAD67 antibody. The bars represent the % of GAD-positive neurons that also expressed mDlx-GCaMP, error bars show  $\pm$ SEM. The data for animal 1 was from 6 imaged areas, and that for animal 2 from 3 areas.

element captures the majority of interneurons in rodents but less is known about the ferret (Dimidschstein et al., 2016). Previous research has shown that the majority of mDlx-expressing neurons also express GAD (Wilson et al., 2017), which tells us about the specificity of the mDlx construct. However, Wilson et al. (2017) did not quantify the efficiency of expression i.e. what proportion of the GAD-expressing neurons are transfected by the mDlx construct and express the GCaMP protein. In order to do this, we counted the proportion of GAD-positive neurons that express the mDlx construct (Figure 3.7). In the two animals used in this chapter, we found a high degree of overlap (Animal 1: mean % = 90.6, n = 6 areas; Animal 2: mean % = 88.9, n = 3 areas). Thus, the mDlx construct should transfect efficiently and specifically the majority of INH.

Given that both the EXC and INH will express the same GFP indicator under GCaMP6f, we required a way to distinguish them from one another in addition to separating out the PV and SOM subclasses. We opted for a strategy using histological techniques, antibody staining and image registration (Figure 3.8). This approach has been previously applied successfully in the mouse (Khan et al., 2018) and ferret (Wilson et al., 2017) visual cortices. However, this was new to our lab. The approach consists of fixing the imaged brain immediately after calcium imaging, slicing the brain into

thin sections ( $60\mu\text{m}$ ), staining with antibodies for the GAD, PV and SOM proteins, imaging the resulting stained sections and then registering the confocal images back to the original 2-photon images in order to assign the neuronal identities. Figure 3.8 below summarises the experimental and data analysis steps involved. Below I describe in more detail the histology, antibody staining procedure and optimisation and the image registration.

In order to identify the different types of neurons (EXC, GAD, PV and SOM) we used an antibody staining approach. Before we can do the staining, it is necessary to perfuse the brain with a fixative (PFA) and slice it using a freezing microtome. Given the lateral location of the ferret AC (ectosylvian gyrus), the standard ways of preparing brain sections (coronal, sagittal or horizontal) would provide confocal images which are in a different orientation from the in vivo calcium images. Therefore, I developed a way to rest the brain using an embedding medium (Tissue-Tek) in an orientation that allowed me to cut brain sections on the microtome which are parallel to the imaging plane. This aided the subsequent image registration with images acquired in vivo. Another factor to consider is the section thickness. Sections that are thinner make it easier for the antibodies to penetrate the brain tissue and stain effectively. On the other hand, if the sections are too thin they are more fragile and the subsequent image registration step is difficult as the in vivo imaging plane will be distributed across more

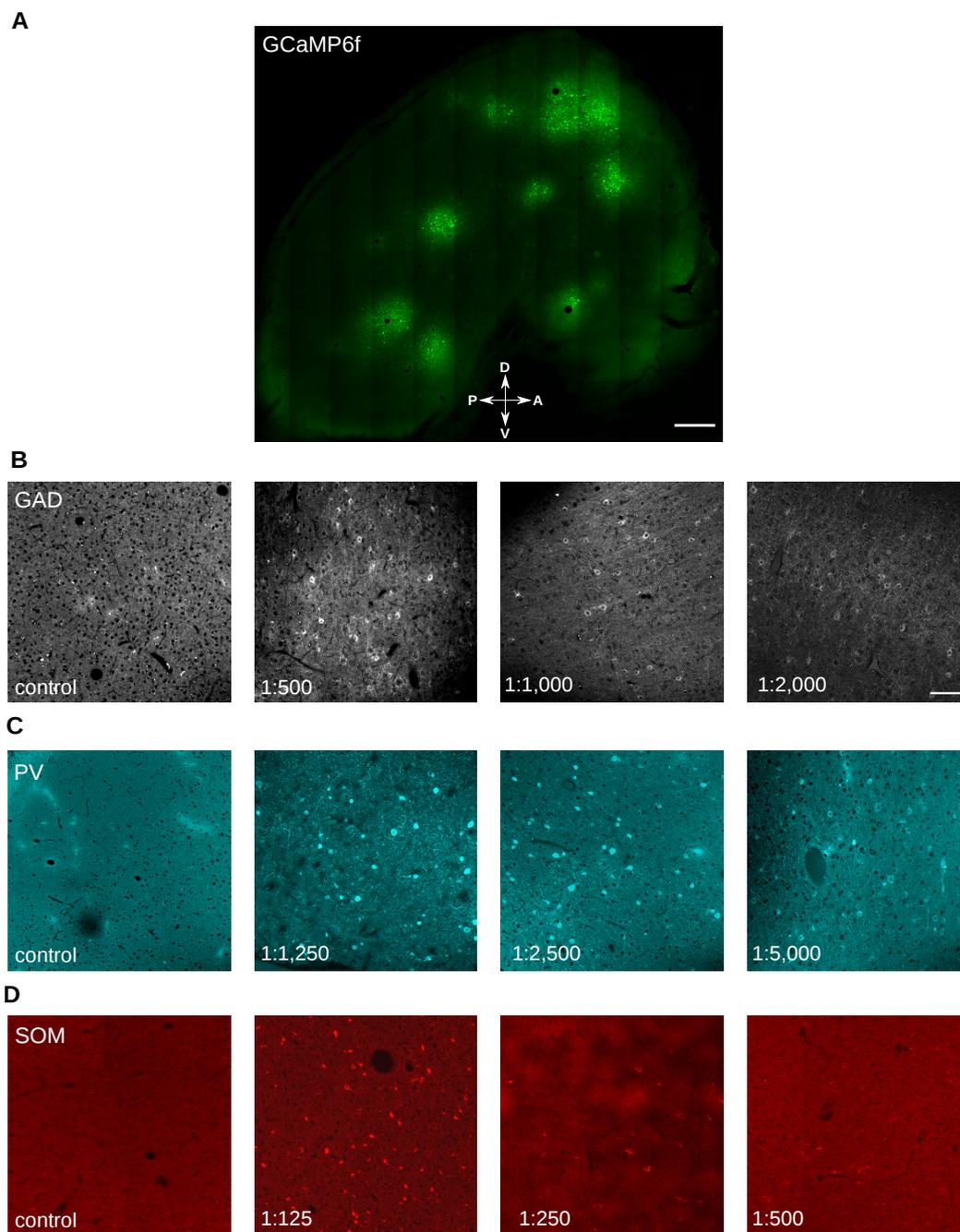


**Fig. 3.8 General imaging workflow.**

In order to separate excitatory pyramidal neurons from inhibitory GAD and assess their functional properties, we used the following pipeline. First, we performed in vivo 2-photon  $\text{Ca}^{2+}$  imaging in the AC of anaesthetized ferrets to record their neuronal responses to pure tones. Next, after perfusion with PFA,  $60\mu\text{m}$  thick brain sections were cut on a freezing microtome. The resulting sections were stained with primary antibodies for GAD, parvalbumin (PV) and somatostatin (SOM), followed by staining with secondary fluorescent antibodies. The stained sections were then imaged using a confocal microscope. Using an affine transformation approach, the confocal images were then registered with the in vivo 2-photon images which allowed for the identification of the recorded neurons as excitatory, GAD, PV or SOM-expressing.

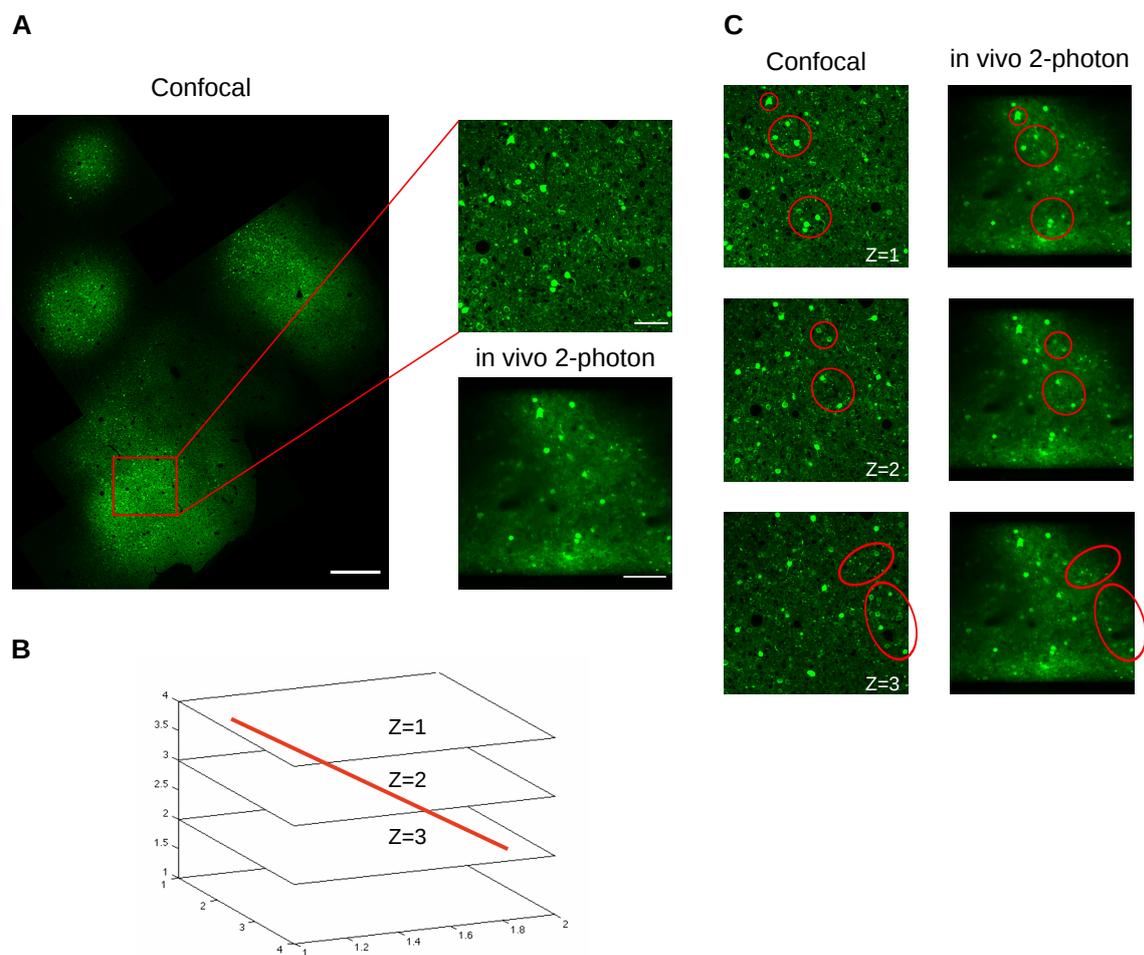
confocal sections (see Figure 3.10B). We found that a thickness of  $60\mu\text{m}$  worked best, after trying  $50\mu\text{m}$  and  $80\mu\text{m}$  thickness too. The  $50\mu\text{m}$  proved too thin separating the 2-photon in vivo imaging plane between two many physical sections, while  $80\mu\text{m}$  did not allow for adequate antibody penetration.

Next, we optimised the antibody staining. There are many parameters that could be varied in the staining procedure, however some of the parameters might have a bigger impact and influence more the quality of labelling. The primary antibody concentration, staining time and temperature are of critical importance. In addition, the optimal conditions for one antibody might not be ideal for another so it was necessary to find a reasonable compromise. I tried out different primary antibody dilutions on brain sections that were not critical to our experiments, and found that the optimal concentrations were: 1:1,1000 for GAD (Figure 3.9B); 1:2,500 for PV (Figure 3.9C); 1:125 for SOM (Figure 3.9D). The SOM staining proved particularly elusive as the SOM protein is primarily localized in vesicles and the endoplasmic reticulum (Baraban and Tallent, 2004), making its epitopes less accessible to the antibodies. Eventually, we found that the SOM protein was properly revealed only after an antigen retrieval procedure followed by a 72-hour-long staining at  $4^{\circ}\text{C}$ . Using a higher concentration of the detergent (0.5% Triton X-100), which serves to dissolve membranes, also improved the staining. The same conditions also worked for the GAD and PV stain, so we chose those for our protocol (see subsection 3.2.5 and Table 3.1 for more details).



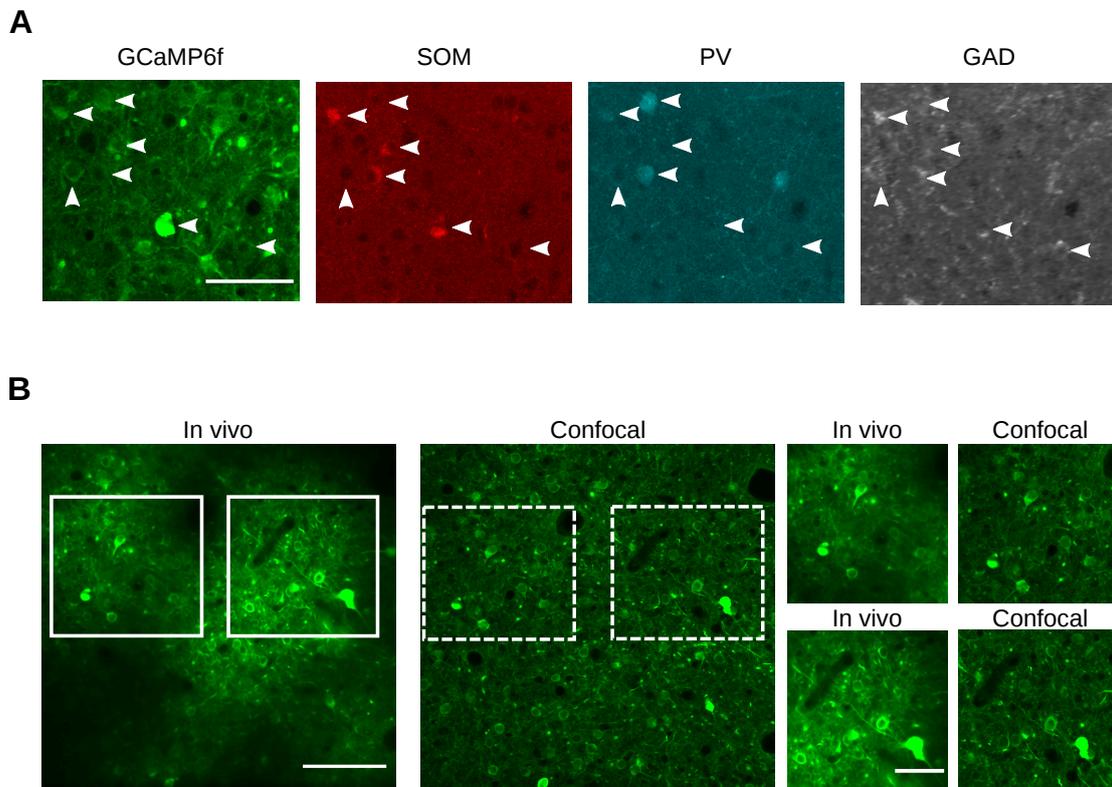
**Fig. 3.9 Optimising antibody staining for different inhibitory subclasses.** **A**, Confocal image of a brain section cut tangentially to the surface of the middle ectosylvian gyrus following the procedure described in Figure 3.8. This animal was expressing GCaMP6f in neurons following a viral vector injection. Scale bar =  $500\mu\text{m}$ . **B-D**, Antibody staining for glutamate decarboxylase (GAD, gray), parvalbumin (PV, cyan) and somatostatin (SOM, red) proteins. Each row represents example staining for one type of protein. The first column is a control stain with the secondary antibody present but not the primary to assess the background. Columns 2-4 in each panel represent increasing dilutions of the primary antibody to find the best signal-to-noise ratio. For GAD this was 1:1,000, for PV 1:2,500 and for SOM 1:125. Scale bar =  $100\mu\text{m}$ .

Following the antibody staining and confocal imaging is the step of image registration (Figure 3.8 and Figure 3.10). Here we applied an approach very similar to those described by Wilson et al. (2017) and Khan et al. (2018). In essence, we tried to match the in vivo 2-photon imaging plane to the confocal sections. This allowed us to identify the cells imaged in vivo with the histological markers (GAD, PV or SOM). We selected landmarks in the in vivo and confocal images (e.g. large and brightly fluorescing cells, blood vessels or other obvious imaged patterns), which were easily located in both images (Figure 3.10C). As there were discrete injection sites (Figure 3.9A, Figure 3.10A) and each of them was large enough to fit roughly a single 2-photon field of view ( $0.5 \times 0.5 \text{ mm}^2$ ), we know approximately where each imaged area is relative to the confocal image as shown in Figure 3.10A. The main difficulty lies in the fact that the 2-photon imaging plane can be slanted relative to the confocal imaging planes (Figure 3.10B). Each of the ferret brain sections is about  $60 \mu\text{m}$  thick so due to light scattering and the limit of the point-spread function, it is necessary to take several optical sections (Z-stacks) in order to capture the neurons adequately. Thus, each in vivo 2-photon image may be spread between several confocal planes, as we can see in Figure 3.10C where the 2-photon plane has a left-to-right tilt with increasing Z-stack depth. The problem is further exacerbated by the fact that there might be non-linear distortions in the tissue due to the pressure from the glass coverslip in vivo, or the perfusion, slicing, freezing and antigen retrieval steps (e.g. boiling at  $\sim 100^\circ\text{C}$ , see subsection 3.2.5). Therefore, to solve this problem we applied a 3D affine transformation from landmark points to transform the confocal Z-stack such that it matches the 2-photon imaging plane as closely as possible (Biswal and Banerjee, 2010). After selecting a few dozen points in the GCaMP green channel, a custom written software calculates the optimal transformation matrix, which is then applied to the channel in question and all the other confocal Z-stack channels (GAD, PV and SOM). As the 4 different fluorescent channels were collected together they require the same affine transformation. An example successful transformation of an in vivo imaged area along with the 4 different fluorescent channels is shown in Figure 3.11. There is a good degree of matching of cell bodies and even finer details such as dendritic spines and axonal boutons (Figure 3.11B, rightmost panel). In a final step, we manually assigned each in vivo imaged cell to one or more of the four neural subtype categories based on the immunohistology: EXC, GAD, PV and SOM.



**Fig. 3.10 Registration of confocal and in vivo 2-photon images.**

**A**, Example confocal image of the AC of a ferret expressing GCaMP6f (left), scale bar =  $500\mu\text{m}$ . The inset (top right) shows a region in more detail (scale bar =  $100\mu\text{m}$ ), corresponding to the area that was imaged in vivo (bottom right, scale bar =  $100\mu\text{m}$ ). **B**, Schematic illustrating the relationship between the histological confocal z-stack and the in vivo 2-photon imaging plane for the example shown in A. Ideally, the in vivo imaging plane would match only one of the confocal z-planes. However, due to the imperfect slicing process and potential distortions of the brain in vivo caused by the glass coverslip, the in vivo imaging plane is not parallel to the confocal z-stack. This means that the in vivo image (red line) intersects several of the confocal sections (z-planes). **C**, The left column of the panel represents the 3 confocal z-planes matching the single in vivo imaging plane in the right column (repeated 3 times to aid visual comparison). Example matching cells in the three planes ( $Z = 1,2,3$ ) are highlighted by the red ellipses. Notice the left-to-right tilt of the in vivo imaging plane relative to the confocal stack similar to the schematic in B. Scale is the same as in A.



**Fig. 3.11** Viral expression of GCaMP6f via mDlx enhancer yields specific labeling of GABAergic neurons in ferret auditory cortex.

**A**, Labeling of GABAergic neurons in ferret AC with AAV2/1.mDlx.GCaMP6f expression; (left to right) Confocal images of GCaMP expression (green), immunostaining for somatostatin (SOM, red), parvalbumin (PV, cyan), and glutamate decarboxylase (GAD, grayscale); The four panels match the top image in the rightmost panel of **B**; scale bar =  $50\mu\text{m}$ ; **B**, Identification of corresponding neuronal cell bodies in 2-photon in vivo images and confocal images of tissue sections. Left image shows the 2-photon in vivo field of view (scale bar  $100\mu\text{m}$ ), middle image shows corresponding confocal field of view, and the right four images show insets, illustrating a high percentage of corresponding cell bodies. The confocal image in the middle panel was first registered to the in vivo image on the left via an affine transformation as described in Figure 3.10 and subsection 3.2.7 from the Methods. Scale bar =  $50\mu\text{m}$ .

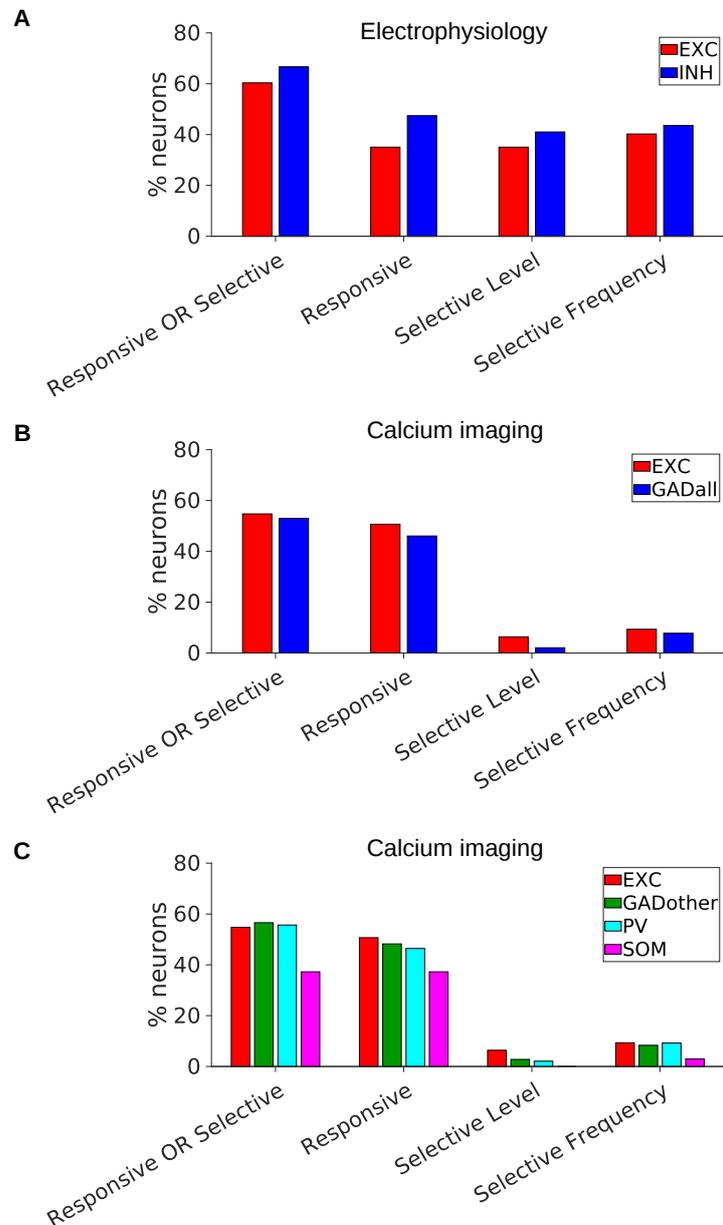
After successfully separating putative EXC and INH in our microelectrode recordings, and EXC and different subclasses of INH with imaging approaches, we explored their functional and stimulus response properties.

### 3.3.3 Response properties of excitatory and inhibitory neurons

Given that the differences in frequency tuning and other properties of EXC and INH in AC are still largely unresolved (Moore and Wehr, 2013; Li et al., 2015; Mesik et al., 2015; Maor et al., 2016; Liang et al., 2019; Blackwell and Geffen, 2017) we wanted to explore and compare them directly. Additionally, most previous studies that address this question have examined neural tuning properties in rodents, and the results may differ in carnivores, such as the ferret, and other higher mammals. To address these questions we performed a number of analyses following our successful separation of EXC and INH described above (sections 3.3.1 and 3.3.2).

First, we set out to quantify the proportion of neurons that were responsive to tones or selective for their frequency within putative EXC and INH in our electrophysiological data (Figure 3.12A), and within EXC and subclasses of INH in our imaging data (Figure 3.12B,C). We defined 4 response categories: 1. Responsive or selective; 2. Responsive; 3. Level selective; 4. Frequency selective. See subsection 3.2.8 from Materials and Methods on details of how these were defined. Briefly, a neuron needed to show a significant effect in at least one t-test comparing activity during stimulus presentation and the period of silence immediately preceding stimulus presentation to be defined as “Responsive”. Neurons showing a significant main effect of frequency or a frequency/level interaction in a two-way ANOVA, with tone frequency and sound level as predictors, were defined as “Frequency selective”. Similarly, being “Level selective” required a significant main effect of level. Finally, a “Responsive or selective” neuron needed to satisfy any of the criteria for frequency selectivity, level selectivity or sound responsiveness. We have to make an important clarification here: In theory, a neuron that is frequency or level selective, will also necessarily be responsive to sound. However, not all neurons which passed the ANOVA for selectivity also passed the t-test for responsiveness or vice versa, because of statistical noise. Given the small number of INH present, this more inclusive response category allowed us to exclude less neurons from our analyses, while making sure that the neurons were modulated by the sound stimuli in some way.

We applied these criteria to the putative EXC and INH identified in microelectrode recordings (see section 3.3.1). We observed that similar proportions of excitatory and inhibitory neurons fell into each of the four tuning categories, as shown in Figure 3.12A.



**Fig. 3.12 Similar proportions of excitatory and inhibitory neurons are responsive to tones.**

**A**, Excitatory (EXC, red) and inhibitory neurons (INH, blue) were assigned by k-means clustering (see Figures 3.4 and 3.6) from single unit microelectrode recordings. Here, these neurons are further classified based on whether they were significantly responsive to tones (t-test,  $p < 0.01$ ), and selective for the level and frequency of tones (2-way ANOVA,  $p < 0.05$ ) (see 3.2.8). **B**, Neurons from our calcium imaging experiments are categorized into cell types. Those expressing the glutamate decarboxylase (GAD, blue) marker were labeled as inhibitory (GADall) and the rest as excitatory (EXC, red). The proportion of these neurons that were responsive or selective for pure tones are plotted as in A. **C**, The INH from the calcium imaging experiments are further divided into subclasses based on antibody staining (see Figures 3.9 and 3.11): parvalbumin (PV, cyan), somatostatin (SOM, magenta), neurons that express GAD but not PV or SOM (GADother, green). The proportion of neurons in these 4 classes that were responsive or selective for pure tones are plotted as in A and B. Likelihood Ratio Tests did not reveal any significant differences between the neuron classes in any of the response categories plotted in A, B or C ( $p > 0.05$ ).

Likelihood Ratio Tests showed no significant difference in the proportions of “Responsive or Selective” neurons between the two neuron classes ( $\chi^2 = 0.92$ ,  $p = 0.33$ ). Similarly, EXC and INH did not differ significantly in their proportions of Responsive neurons ( $\chi^2 = 3.43$ ,  $p = 0.06$ ), Level Selective neurons ( $\chi^2 = 0.81$ ,  $p = 0.36$ ), or Frequency Selective neurons ( $\chi^2 = 0.25$ ,  $p = 0.61$ ). We also explored the proportions of neurons that fall into the four response categories in our imaging data, where we discriminated between GAD-expressing inhibitory and GAD-negative excitatory neurons (see section 3.3.2), shown in Figure 3.12B. In agreement with the microelectrode results, Likelihood Ratio Tests showed no significant difference in the proportions of “Responsive or Selective” ( $\chi^2 = 0.16$ ,  $p = 0.69$ ), “Responsive” ( $\chi^2 = 0.63$ ,  $p = 0.43$ ), “Level Selective” ( $\chi^2 = 3.20$ ,  $p = 0.07$ ) or “Frequency Selective” ( $\chi^2 = 0.34$ ,  $p = 0.56$ ) neurons. One of the advantages of the imaging over the electrophysiology data is that in addition to separating the neurons into excitatory and inhibitory groups, we can also probe the properties of subclasses of inhibitory neurons. We repeated the analysis above but this time making comparisons between the subclasses of INH and the EXC. Given the large number of comparisons (24), we summarise the results in Table 3.3. In all of the comparisons, we did not observe significant differences in the proportions of the different classes of neurons belonging to the 4 different categories.

**Table 3.3 Comparison of responsivity and selectivity of 4 neuron subtypes in calcium imaging data.**

The first column shows the pair of neuron types that are being compared in each row: EXC, excitatory; PV, parvalbumin; SOM, somatostatin; GADother, GAD-expressing but not PV- or SOM-positive. Columns 2-5 show the results of the Likelihood Ratio Tests comparing the proportions of neurons in each response category. The chi-squared ( $\chi^2$ ) and p-value are indicated. In the Level Selective category there were not enough SOM to compute  $\chi^2$ , indicated by “-”. None of the comparisons were significant at  $\alpha=0.05$  ( $\alpha=0.0083$  with Bonferroni correction).

Interaction	Responsive OR Selective	Responsive	Selective Level	Selective Frequency
EXC-GADother	$\chi^2=0.01$ , $p=0.91$	$\chi^2=0.11$ , $p=0.74$	$\chi^2=0.95$ , $p=0.33$	$\chi^2=0.07$ , $p=0.79$
EXC-PV	$\chi^2=0.02$ , $p=0.89$	$\chi^2=0.11$ , $p=0.74$	$\chi^2=0.82$ , $p=0.36$	$\chi^2=0.03$ , $p=0.87$
EXC-SOM	$\chi^2=2.98$ , $p=0.08$	$\chi^2=1.70$ , $p=0.19$	-	$\chi^2=1.24$ , $p=0.27$
GADother-PV	$\chi^2=0.02$ , $p=0.98$	$\chi^2=0.01$ , $p=0.99$	$\chi^2=0.01$ , $p=0.97$	$\chi^2=0.03$ , $p=0.95$
GADother-SOM	$\chi^2=1.98$ , $p=0.16$	$\chi^2=0.63$ , $p=0.43$	-	$\chi^2=0.54$ , $p=0.46$
PV-SOM	$\chi^2=1.99$ , $p=0.16$	$\chi^2=0.60$ , $p=0.44$	-	$\chi^2=0.61$ , $p=0.43$

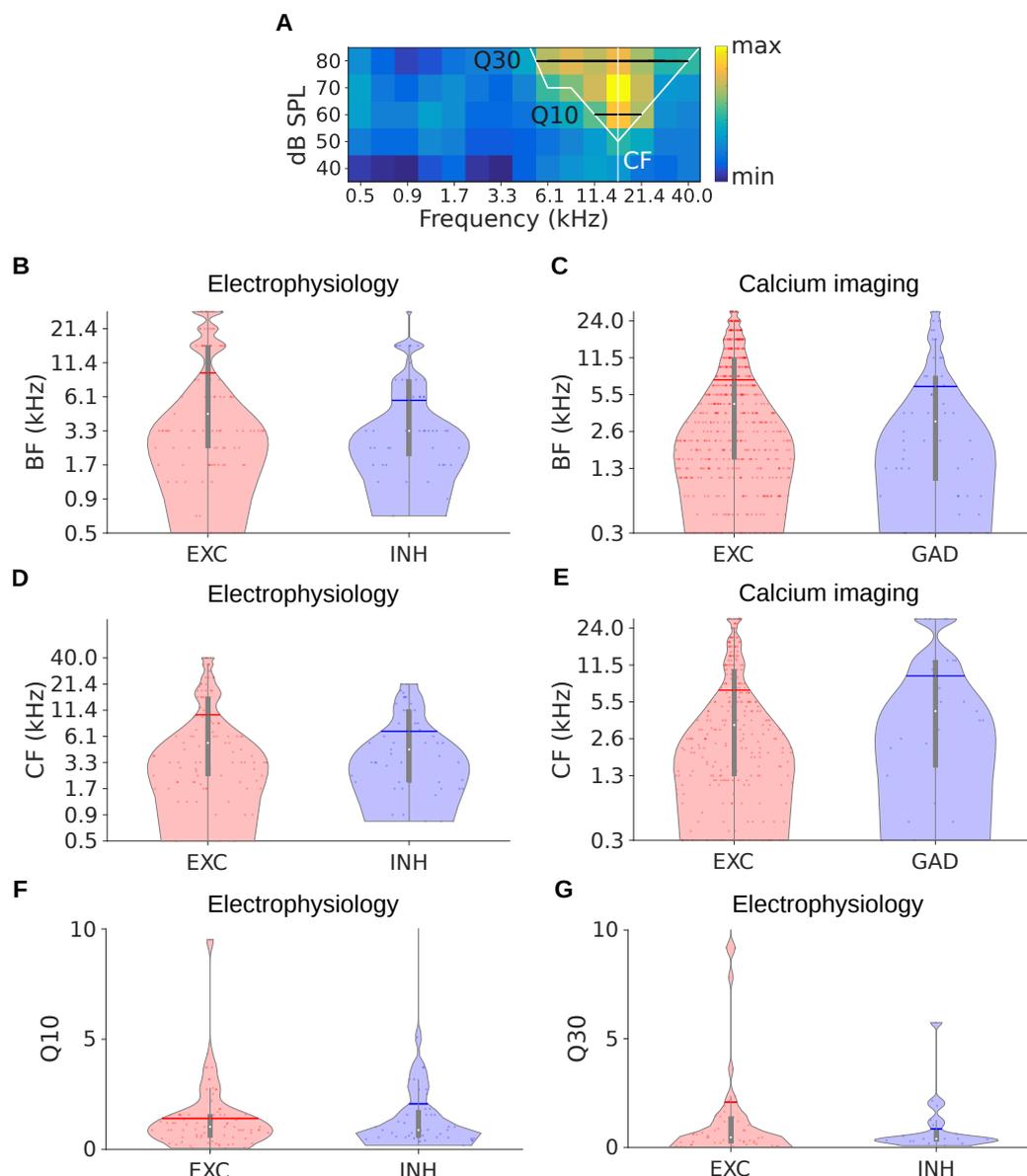
Thus, in both the microelectrode and imaging data we did not observe differences in the proportions of EXC and INH that were responsive to tones, or that were selective to their frequency or level. Additionally, further dividing the INH in subclasses in the imaging data also did not reveal differences in these measures. Given that INH comprise ~15-20% of all neurons, the fact that we had only 2 animals where the whole imaging pipeline yielded satisfactory results and our focus on primary areas, we obtained imaging data from 92 INH in total. Taking into account this small number, we opted for a more relaxed inclusion criterion for our further analyses. In the remaining results in this chapter (Figures 3.13, 3.14, 3.15, 3.16, 3.17) we include neurons in the “Responsive or Selective” group. This provided a good compromise between maximizing the number of INH in the analyses and excluding neurons which do not respond to our sound stimuli.

In order to investigate the tuning properties of neurons further we constructed Frequency Response Area (FRA) plots for each neuron (Figure 3.13A), as described in Chapter 2. Using the FRA matrix as a starting point, several useful parameters can be extracted to characterize the tuning properties of neurons (see subsection 3.2.10). These include the best frequency (BF), characteristic frequency (CF), Q10 and Q30 values. The Q10 value is the CF - normalized bandwidth for sounds presented 10dB louder than CF. Therefore neurons which have a broader tuning curve (i.e. respond to more frequencies) will tend to have smaller Q10 values (Figure 3.13A). The Q30 value is similar to Q10, but here CF is divided by the bandwidth sampled 30dB louder than CF. Again smaller Q30 values correspond to broader tuning.

A priori, there is no particular reason to expect that the BF or CF will differ between EXC and INH. However, we wanted to check that there was no frequency bias in the sampling of our data. Pooling the data across animals and penetrations, we did not observe a significant difference in the BF between putative EXC and INH from single unit microelectrode recordings data (Mann–Whitney U test<sup>4</sup>:  $z = 1.88$ ,  $p = 0.059$ , Figure 3.13B). Similarly, calcium imaging data pooled across animals and imaging areas did not reveal a significant difference in BF between EXC and INH (MW:  $z = 1.59$ ,  $p = 0.11$ , Figure 3.13C). We also tested for differences in CF, where again neither the electrophysiology data (MW:  $z = 1.12$ ,  $p = 0.26$ , Figure 3.13D) nor the imaging data showed significant differences (MW:  $z = -0.86$ ,  $p = 0.39$ , Figure 3.13E).

---

<sup>4</sup>Abbreviated as MW.



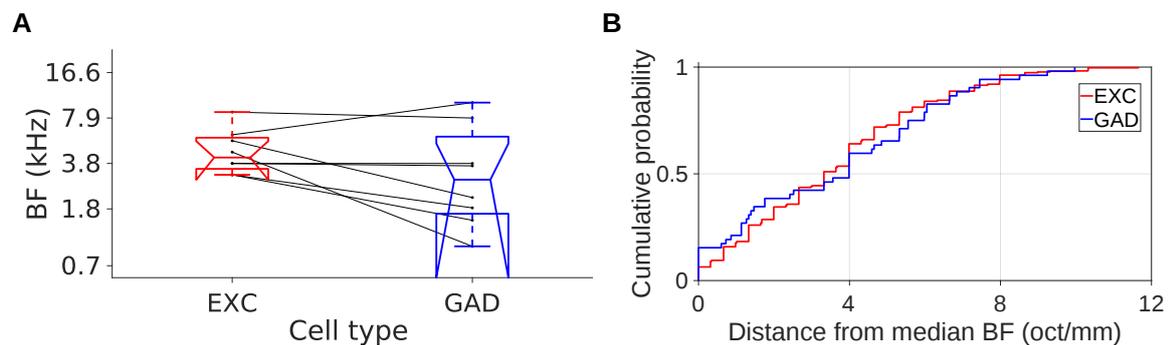
**Fig. 3.13** The tuning properties of excitatory and inhibitory neurons are similar.

**A**, Example Frequency Response Area (FRA) plot from microelectrode recordings demonstrating the features examined in B-G: Best frequency (BF); Characteristic frequency (CF); Q10, the bandwidth 10dB louder than CF normalized by CF ; Q30, the bandwidth 30dB louder than CF normalized by CF. The white line in each case delineates the ‘tuning curve’ (see subsection 3.2.10), from which the CF, Q10 and Q30 values were calculated. **B-C**, Violin plots comparing BF across all neurons identified as excitatory (EXC, red) or inhibitory (INH, blue) using electrophysiology (B) or calcium imaging (excitatory or GAD-expressing, C). **D-E**, Same as B-C, but here CF is compared. **F-G**, Same as to B-E, but here the comparison is for Q10 (F) or Q30 (G) only for the electrophysiology data. In all violin plots (B-G), the white dot represents the median, the horizontal thick line the mean, the thick gray lines the interquartile range, the thin gray lines 1.5x interquartile range, each individual dot is one neuron, the colored shaded area represents the distribution. Mann–Whitney U tests did not reveal significant differences in any of the comparisons (B-G), at  $\alpha=0.05$ .

Given that one of the main proposed differences between EXC and INH in the literature is their tuning width (Blackwell and Geffen, 2017), we also compared the Q10 and Q30 values between putative EXC and INH. The tuning widths of these 2 neuron classes in our microelectrode recordings data did not show significant differences, as evidenced by Q10 values (MW:  $z = 0.005$ ,  $p = 0.99$ , Figure 3.13F) or the Q30 values (MW:  $z = 0.53$ ,  $p = 0.59$ , Figure 3.13G). Due to the lower sensitivity of the GCaMP6f indicator compared to 6m and 6s versions (Chen et al., 2013), the FRA thresholds in our imaging data were not low enough to reliably calculate Q10 and Q30 values for many neurons in our imaging data. Combined with the small number of INH, this did not allow us to be able to compare the Q10 and Q30 values between EXC and INH for the imaging data as we did with electrophysiology (Figure 3.13F,G).

Together, the analyses presented in Figure 3.13 suggest that the basic tuning properties (BF, CF, Q10, and Q30) are largely similar between excitatory and inhibitory neurons in ferret primary AC.

In addition to examining the pooled data we wanted to explore in more detail the BF variability within imaging fields. This is similar to our analysis in Chapter 2 comparing single, double and complex neurons (see Figure 2.5C), but this time separating them into EXC and INH. This analysis is presented in Figure 3.14. First, we investigate whether the median BF in each imaging area was different for the EXC and INH. As expected, we did not observe a significant difference in median BF across imaging areas as we can see in Figure 3.14A (Wilcoxon signed-rank test:  $z = 1.18$ ,  $p = 0.30$ ). As EXC in primary AC are tonotopically organized, this suggests that INH are arranged globally on the same tonotopic map. We also explore the local variability in BF for the two neuronal types (excitatory and inhibitory) within an imaging field. We computed the distance between each neuron's BF and the median BF of all neurons of the same class (EXC or GAD) in the same imaging field (Figure 3.14B). As is visually evident from Figure 3.14B, we did not observe significant differences in local BF variability between the two neuronal classes (MW:  $z = 0.034$ ,  $p = 0.97$ ; Kolmogorov–Smirnov (KS) test:  $k = 0.12$ ,  $p = 0.51$ ). Thus, not only are the basic frequency tuning properties of EXC and INH are similar (Figure 3.13), but also their local variability within an area of primary AC.

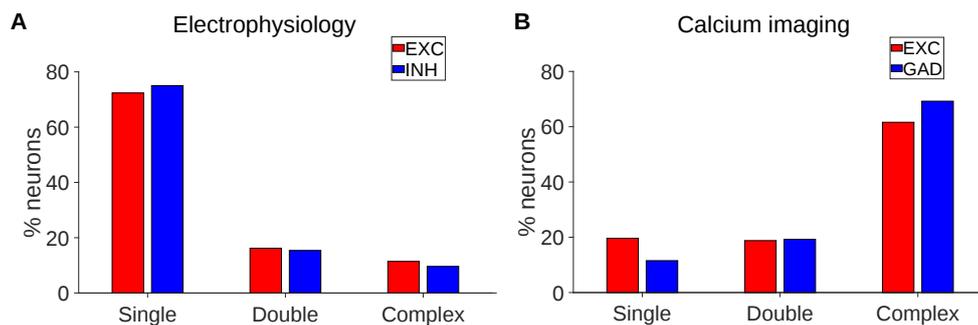


**Fig. 3.14 The best frequency preferences within an imaging field are similar for excitatory and inhibitory neurons.**

**A**, Box plots comparing the mean BF between excitatory (EXC, red) and inhibitory (GAD, blue) neurons within calcium imaging fields. Each black dot represents the median BF in one imaging field. The boxes represent the interquartile range, the line in the middle is the median, and the notch indicates the distribution about the median. The ‘whiskers’ extending above and below the box show the limits of the remaining percentiles. A Wilcoxon signed-rank test did not reveal a significant difference between EXC and GAD at  $\alpha=0.05$ . **B**, Cumulative probability plots of the difference (in octaves) between each neuron’s BF and the median BF of all neurons of the same class (EXC or GAD) in the same imaging field. The distance in octaves was also normalized for the size of the imaging field (oct/mm). Kolmogorov–Smirnov (KS) and Mann–Whitney U tests did not reveal a significant difference between the two neuronal classes at  $\alpha=0.05$ .

We next ask if there are differences in the complexity of frequency receptive fields between EXC and INH. The FRA of each neuron was classified in one of the three categories, as described in Chapter 2 (see subsection 2.2.8, “Classifying the frequency response”): 1. V-, I- or O-shaped FRAs, which all have a clear single BF (“single-peaked neurons”); 2. FRAs with response peaks at two distinct frequencies (“double-peaked neurons”), where one peak is usually substantially stronger than the other; and 3. “complex neurons”, whose FRAs often contained three response peaks and a poorly defined BF. We compared the relative proportions of the three FRA types amongst EXC and INH in electrophysiology (Figure 3.15A) and calcium imaging data (Figure 3.15B). In the electrophysiology data, Likelihood Ratio Tests did not reveal significant differences in the proportions of single-peaked ( $\chi^2 = 0.12$ ,  $p = 0.73$ ), double-peaked ( $\chi^2 = 0.017$ ,  $p = 0.90$ ) and complex ( $\chi^2 = 0.12$ ,  $p = 0.73$ ) FRAs between the EXC and INH (Figure 3.15A). Analogously, in the calcium imaging data there was no significant difference in the proportion of single-peaked ( $\chi^2 = 2.27$ ,  $p = 0.13$ ), double-peaked ( $\chi^2 = 0.006$ ,  $p = 0.94$ ) and complex ( $\chi^2 = 1.22$ ,  $p = 0.27$ )

FRAs (Figure 3.15B) between EXC and GAD. Apart from the lack of difference in the proportions of FRA types amongst EXC and INH, another result that can be observed in Figure 3.15 was the vastly different proportions of single-peaked and complex FRA types between the electrophysiology and imaging data. We observed this in Chapter 2 (Figure 2.5F) and I will explore this important difference further in the next sections and the discussion.



**Fig. 3.15 Electrophysiology and calcium imaging reveal similar complexity of frequency receptive fields between excitatory and inhibitory neurons.**

**A**, Percentage of neurons classified into each FRA class (Single, Double or Complex) from single unit microelectrode recordings. The putative excitatory neurons are in red (EXC) and putative inhibitory in blue (INH). **B**, Same as A, but the data are from two-photon calcium imaging where neurons were separated according to the GAD marker (excitatory in red, EXC; inhibitory in blue, GAD). In both A&B, Likelihood Ratio Tests did not reveal significant differences at  $\alpha=0.05$ .

Apart from the tuning properties, it is also very instructive to look at the interactions between neurons. This can allow us to measure which groups of neurons might form interacting networks, which could in turn also influence their tuning properties. A simple, yet powerful analysis is to explore the signal and noise correlations between different groups of neurons (Rothschild et al., 2010; Maor et al., 2016; Panniello et al., 2018; Cohen and Kohn, 2011). The Signal Correlation (SC) between two neurons measures roughly the correlation between their mean responses to different stimuli<sup>5</sup>. This measure is often used to quantify the extent to which a pair of neurons has similar response properties (co-tuning). Noise Correlation (NC) between two neurons measure the correlation between fluctuations around the mean response to the same stimulus. NC are often taken as a proxy for functional connectivity and under certain conditions

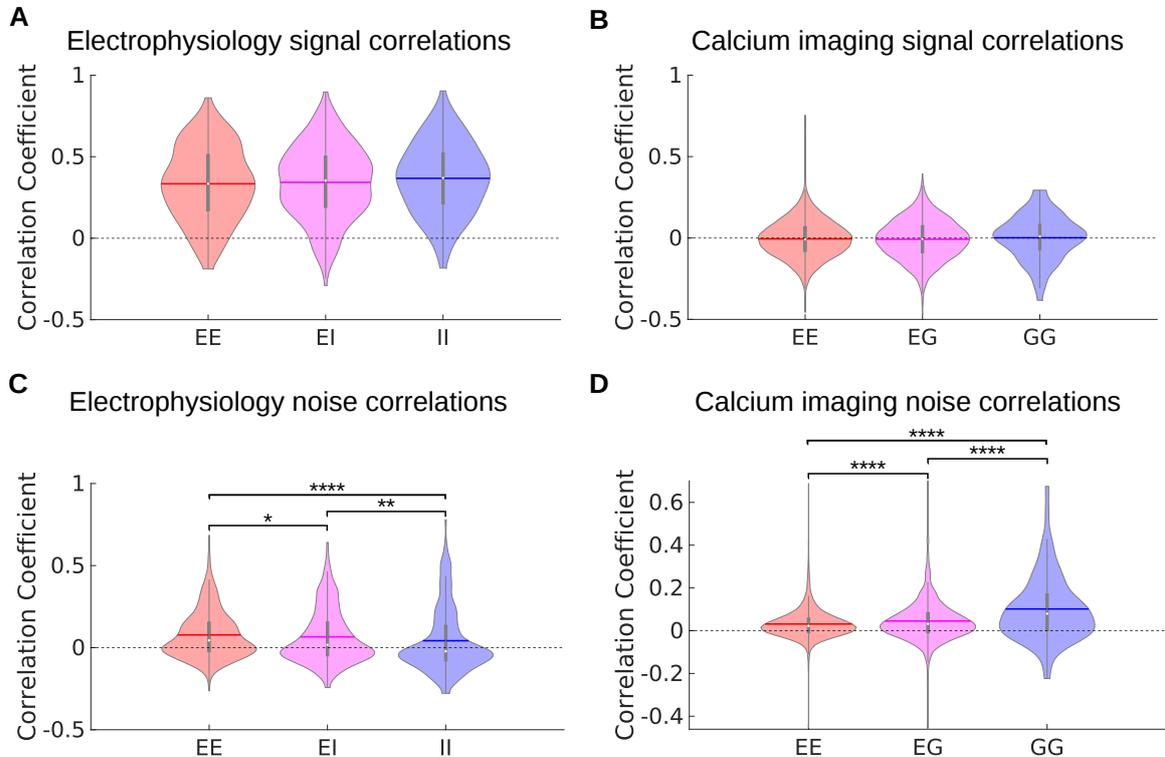
<sup>5</sup>Taking the mean response removes some of the noise, however to more accurately estimate the SC some corrections are necessary as described by Rothschild et al. (2010).

can be indicative of synaptic connectivity, similar to the crosscorrelogram approach described earlier (section 3.3.1 and Figure 3.6).

In order to quantify the functional interactions between EXC and INH we computed SC and NC within and across neuronal classes: excitatory-to-excitatory, EE; excitatory-to-inhibitory, EI (electrophysiology) or EG (calcium imaging); inhibitory-to-inhibitory, II (electrophysiology) or GG (calcium imaging), (Figure 3.16). In the electrophysiology data (Figure 3.16A), a Kruskal–Wallis (KW) test did not reveal a significant difference in signal correlations across the three pairs of neuronal classes ( $\chi^2(2) = 3.24$ ,  $p = 0.20$ ). Furthermore, post hoc pairwise Tukey’s Honestly Significant Difference (THSD) tests found no significant pairwise differences between groups: EE-EI ( $p = 0.70$ ), EE-II ( $p = 0.17$ ) and EI-II ( $p = 0.45$ ). Similarly, in the calcium imaging data (Figure 3.16B) we also did not observe significant differences in signal correlations overall (KW:  $\chi^2(2) = 1.72$ ,  $p = 0.42$ ) or in pairwise post hoc comparisons (THSD: EE-EG ( $p = 0.88$ ), EE-GG ( $p = 0.45$ ) and EG-GG ( $p = 0.40$ )). The noise correlations in the microelectrode data (Figure 3.16C) showed a significant effect of the neuron groups (KW:  $\chi^2(2) = 27.44$ ,  $p = 1.1 \times 10^{-6}$ ), including differences between each of the 3 pairwise comparisons (THSD: EE-EI ( $p = 0.014$ ), EE-II ( $p = 9.2 \times 10^{-7}$ ) and EI-II ( $p = 0.0051$ )). The noise correlations in the calcium imaging data (Figure 3.16D) also showed significant differences between the neuron groups (KW:  $\chi^2(2) = 106.49$ ,  $p = 7.5 \times 10^{-24}$ ; THSD: EE-EG ( $p = 9.5 \times 10^{-10}$ ), EE-GG ( $p = 1.84 \times 10^{-9}$ ) and EG-GG ( $p = 2.87 \times 10^{-5}$ )).

Thus, the co-tuning to sound features seems to be very similar between excitatory and inhibitory neurons (Figure 3.16A,B). Despite this similarity being evident in both electrophysiology and calcium imaging data, it is worth pointing out that the microelectrode data had much higher mean signal correlation values (electrophysiology: EE (0.33); EI (0.34); II (0.37); calcium imaging: EE (-0.0055); EG (-0.0077); GG (0.0016)). This difference is not entirely surprising. Both the electrophysiology and the calcium imaging data come from L2/3 primary AC. However, due to the sampling bias described in the previous chapter, there are more “single-peaked” than “complex” neurons in the electrophysiology data compared to the calcium imaging data (see Figure 3.15 and Figure 2.5F). As we saw in Figure 2.6A,C from Chapter 2, neurons with single- and double-peaked FRAs tend to have higher SC comparing to the complex ones.

Unlike signal correlations, the noise correlation values were significantly different between excitatory and inhibitory neurons in both the electrophysiology and calcium

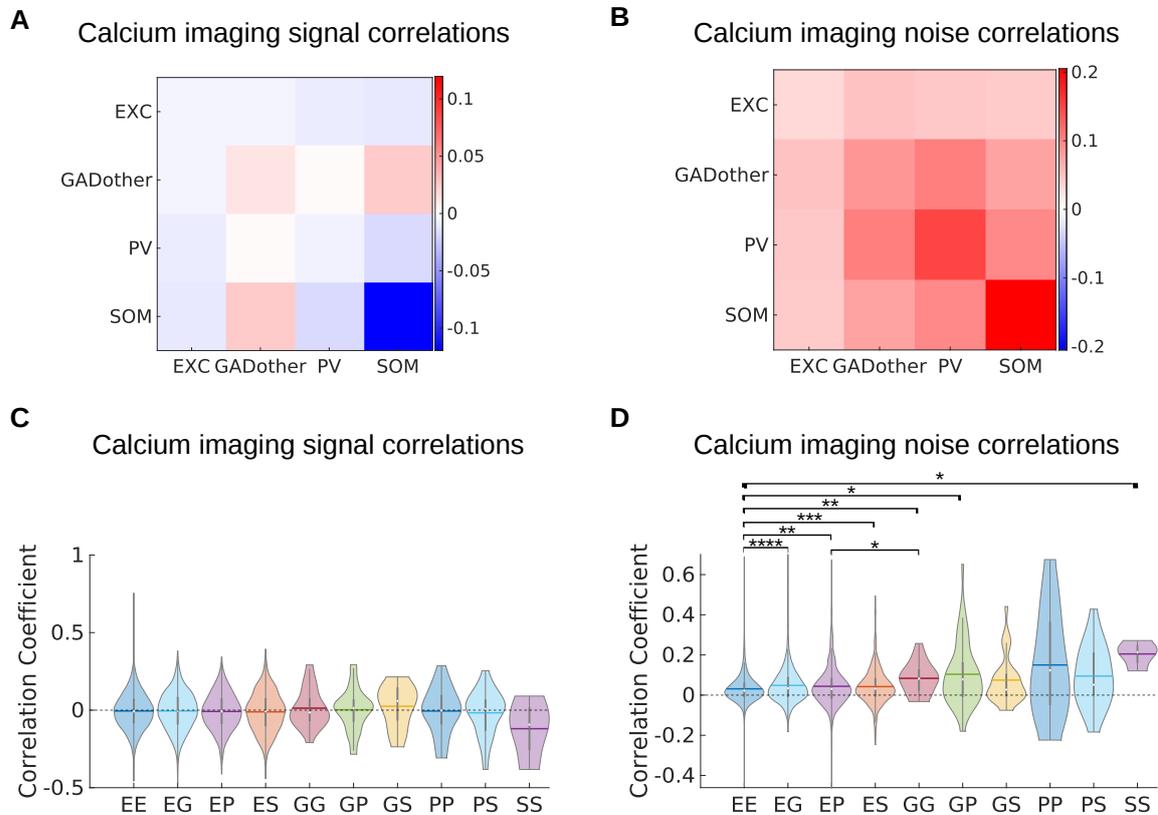


**Fig. 3.16 Correlations in neural activity from electrophysiology and calcium imaging data reveal higher connectivity in different classes of neurons.**

**A**, Violin plots showing the signal correlations between the two classes of neurons identified by k-means clustering in single unit microelectrode data. The signal correlations are shown for all excitatory-excitatory (EE, red), excitatory-inhibitory (EI, magenta) and inhibitory-inhibitory (II, blue) neuronal pairs. The data are pooled across animals and penetrations. **B**, Same as **A**, but showing the signal correlations for data collected by calcium imaging and neurons separated by GAD staining (G). **C**, Violin plots analogous to those in **A**, showing the noise correlations in single unit microelectrode recordings. **D**, Violin plots analogous to those in **B**, showing the noise correlations in calcium imaging data. In all violin plots (**A-D**), the white dot represents the median, the horizontal thick line the mean, the thick gray lines the interquartile range, the thin gray lines 1.5x interquartile range, the colored shaded area represents the distribution. The results of Kruskal–Wallis tests followed by multiple comparisons using Tukey’s Honestly Significant Difference (HSD) correction procedure are indicated above the bars in **C** and **D** (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\*\* $p < 0.0001$ ). The same procedure did not reveal significant differences in the signal correlations in **A** and **B** at  $\alpha=0.05$ .

imaging data (Figure 3.16C,D). The relative order of the groups was different, however, (electrophysiology: EE (0.077) > EI (0.066) > II (0.042); calcium imaging: EE (0.031) < EG (0.045) < GG (0.1)). Following the logic that noise correlations are indicative of neuronal connectivity, our findings seem to suggest that amongst the neurons sampled by electrophysiology, EE coupling is tighter than EI and II coupling, and EI coupling is stronger than II coupling. Interestingly, in our calcium imaging data we observe the opposite trend where the inhibitory-to-inhibitory (GG) coupling is stronger than EE or EG, while EG is stronger than EE. We will come back to this point in the Discussion section of this Chapter.

Finally, we wanted to dissect in more detail the interactions between EXC and different subclasses of INH (Figure 3.17). First, we explored the signal correlations, as summarised in Figure 3.17A,C. As we can see in Figure 3.17A, which shows the mean value of the SC for each interaction, there were no strong differences apart from a slightly lower SC for the SOM-to-SOM (SS) interaction. Similar to our more coarse-grained analysis in Figure 3.16B, we did not observe significant differences in the SC across the different groups of neurons as revealed by a Kruskal–Wallis (KW) test ( $\chi^2(9) = 7.66$ ,  $p = 0.57$ ), followed by post hoc pairwise Tukey’s Honestly Significant Difference (THSD) tests at  $\alpha=0.05$  (Figure 3.17C). In contrast, the noise correlations did show some interesting effects (Figure 3.17B). Just as we observed earlier (Figure 3.16D), the excitatory-to-excitatory (EE) coupling was the weakest, that between excitatory and the inhibitory neurons, PV (P), SOM (S) and GADother (G, GAD-positive, non-PV and non-SOM) was stronger, with the inhibitory-to-inhibitory coupling being the strongest (Figure 3.17D). A KW test revealed a significant overall difference in NC between the groups ( $\chi^2(9) = 120.16$ ,  $p = 1.2 \times 10^{-21}$ ). Furthermore, THSD post hoc pairwise comparisons showed significant differences in the following pairs of interactions: EE-EG ( $p = 1.3 \times 10^{-7}$ ), EE-EP ( $p = 0.0032$ ), EE-ES ( $p = 5.9 \times 10^{-4}$ ), EE-GG ( $p = 0.0031$ ), EE-GP ( $p = 0.028$ ), EE-SS ( $p = 0.039$ ), EP-GG ( $p = 0.026$ ). Thus, the excitatory-to-excitatory coupling is weaker than that between the excitatory and the different subclasses of inhibitory neurons, PV, SOM and GADother (mostly VIP and non-VIP 5HT cells). The excitatory-to-excitatory coupling was also weaker compared to that within pairs of GADother neurons, GADother-PV pairs, and the SOM-SOM coupling. We also observed that in the comparison of the excitatory-to-inhibitory coupling to that of inhibitory-to-inhibitory coupling, only the excitatory to PV (EP) was different from that of GADother-GADother coupling. All the inhibitory-to-inhibitory couplings tended to have higher average NCS, especially that of PV-to-PV and SOM-to-SOM (Figure 3.17B,D)



**Fig. 3.17 Signal and noise correlations in subclasses of inhibitory neurons reveal distinct interactions.**

**A**, Matrix of signal correlations from calcium imaging data from 2 animals between the excitatory and different subclasses of inhibitory neurons: PV, SOM, and non-PV and non-SOM GAD positive (GADother). Each entry represents the mean value of the signal correlations between two types of neurons (the mean of each violin plot in **C**). **B**, Matrix of noise correlations between the same types of neurons as in **A**. Each entry is the mean value of a violin plot in **D**. **C**, Violin plots of the signal correlations between subclasses of neurons: EXC-EXC (EE); EXC-GADother (EG); excitatory-PV (EP); EXC-SOM (ES); GADother-GADother (GG); GADother-PV (GP); GADother-SOM (GS), PV-PV (PP), PV-SOM (PS), SOM-SOM (SS). **D**, Violin plots of the noise correlations between the same types of neurons as in **C**. In all violin plots (**C**-**D**), the white dot represents the median, the horizontal thick line the mean, the thick gray lines the interquartile range, the thin gray lines 1.5x interquartile range, the colored shaded area represents the distribution. The results of Kruskal–Wallis tests followed by multiple comparisons using Tukey’s Honestly Significant Difference (HSD) correction procedure are indicated above the bars in **D** (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ ). The same procedure did not reveal significant differences in the signal correlations in **C**.

but we did not observe significant differences in these comparisons, likely due to the small numbers of neurons in each subclass.

### 3.4 Discussion

In this chapter, we developed electrophysiological and imaging approaches in order to separate excitatory from inhibitory neurons, and further divide the inhibitory class into PV and SOM subclasses in the imaging data. In the case of the electrophysiological dataset, we tested five different spike waveform features commonly used in the literature in order to determine which combination of features gave us the best separation of putative inhibitory and excitatory neurons. We found that the trough-to-peak and peak width features allowed us to discriminate between putative excitatory and inhibitory neurons using k-means clustering, as indicated by metrics of clustering quality. Further validation of our approach was provided by CCG measures of putative synaptic connectivity. Using a combination of targeted viral delivery of the calcium indicator GCaMP6f to populations of excitatory and inhibitory neurons, 2-photon calcium imaging, immunohistochemistry and registration techniques, we were able to successfully label, record from and identify post hoc excitatory and two of the main classes of inhibitory neurons, PV and SOM. This involved several refinement steps and resulted in a robust and reliable protocol which opens up many possibilities for the future investigation of inhibitory neurons in ferret AC.

Using these two approaches to identify neuron classes, we compared the properties of excitatory and inhibitory neurons in AC. Surprisingly, we did not find significant differences in the proportions of tuned neurons, their frequency tuning characteristics, such as tuning width, local tonotopic variability or complexity of frequency receptive fields. We also examined their co-tuning for pure tones using signal correlations and found that the responses within and between classes are very similar. Using noise correlations as an indicator of functional connectivity (coupling), we found that in our imaging data inhibitory subclasses tended to be more coupled to one another and form subnetworks, compared to excitatory. Interestingly, we found the opposite in our electrophysiology data. These differences could be potentially explained by the sampling of neurons with different complexity of their frequency receptive fields.

### 3.4.1 High density microelectrode recordings allow us to separate putative excitatory and inhibitory neurons based on spike waveform features

The novel high density microelectrodes, “Neuropixels” (Jun et al., 2017; Juavinett et al., 2019; Steinmetz et al., 2020), make it possible to record tens to hundreds of single neurons simultaneously, and can be applied in either an acute or chronic setting. Owing to the high number of channels (“pixels”), high sampling rate and excellent signal-to-noise quality, it is also possible to capture individual spike waveforms in exquisite detail. In addition, the advent of better, faster and more efficient spike-sorting algorithms such as “Kilosort” (Pachitariu et al., 2016) allows us to largely resolve some longstanding issues, such as temporally overlapping spikes recorded on the same channels and processing large electrophysiological datasets. We wanted to utilize these developments in order to characterize and classify putative excitatory and inhibitory neurons from extracellular recordings. Owing to their biophysical properties, excitatory and inhibitory neurons tend to exhibit distinct spike waveforms that should in theory allow us to separate them from one another, even if explicit markers of cell identity are not present (Peyrache and Destexhe, 2019).

We compared and evaluated five spike features that are commonly used in the literature: 1) Trough-to-peak duration (feature 1); 2) Peak width (feature 2); 3) Peak end slope (feature 3); 4) Trough-to-peak ratio (feature 4); 5) Trough width (feature 5); see Table 3.2 and Figure 3.2 (Sirota et al., 2008; Peyrache et al., 2012; Stark et al., 2013; Medrihan et al., 2017; Yu et al., 2019). Similar to previous studies, we observed a clear bimodal distribution in some of the features (see Figure 3.3). Features 1 and 2 showed such bimodal distribution, while features 3-5 were unimodal. We then tested the 3 sets of most promising feature pairs (features 1 and 2, features 1 and 5, features 2 and 5) in order to assign cluster identity as putative excitatory or inhibitory neurons. Using an unbiased clustering algorithm (k-means), we found that features 1 and 2 seemed to separate best the putative neuronal classes. Our visual intuition was confirmed by the quantitative silhouette score metric, which verified that features 1 and 2 provided the best degree of separation, followed by features 1 and 5, and features 2 and 5 (see Figure 3.4). However, in accordance with previous studies, we observed a high degree of overlap (91%, see Figure 3.5) between the classification produced by the

different spike features, as they tend to be correlated with one another (Sirota et al., 2008; Stark et al., 2013).

Similar to other studies, we wanted to corroborate our classification using some form of functional connectivity measure (Sirota et al., 2008; Peyrache et al., 2012; Mendoza et al., 2016). Using CCGs we identified putative monosynaptic excitatory and inhibitory interactions and compared those to the classes returned from our clustering approach, finding very good correspondence between the two (see Figure 3.6). It was especially reassuring that while we were towards the end of the current study, a paper by Yu et al. (2019) using optotagging (Deubner et al., 2019) of PV, SOM and VIP in mouse somatosensory cortex found that the two features we selected (Trough-to-peak duration and Peak width) provide good separation of most inhibitory neurons from excitatory ones. It is worth pointing out that the CCG analysis is just one approach and there are several possible ways to measure putative connectivity and interactions between neurons, with varying degrees of sophistication. A recent study by Kobayashi et al. (2019) used Generalised Linear Model (GLM) to fit “coupling filters” between the neurons to infer their connectivity from spiking data. Another study was able to infer dynamic connectivity assemblies by integrating techniques from adaptive filtering, compressed sensing, point process theory, and high-dimensional statistics (Sheikhattar et al., 2018). In the future, it would be interesting to supplement the CCG analysis with such approaches in order to arrive at a more robust validation of the neuronal types inferred from clustering and to explore further the neuronal connectivity between inhibitory and excitatory neurons.

Despite this, caution must be exercised in the interpretation of our results. Both the Yu et al. (2019) study and another optotagging study by Royer et al. (2012) found that PV fast-spiking (FS) neurons were very clearly separated from excitatory pyramidal regular spiking neurons (RS). However, SOM and VIP had more intermediate values of the two spike features and thus there is more ambiguity in their separation. Another complication stems from the fact that different inhibitory neurons have different distributions across layers (Tremblay et al., 2016). A potential improvement to the current procedure could involve the consideration of additional neuronal features such as firing rate, distribution of interspike intervals (ISIs), refractory period, burstiness, or coupling to brain oscillations (Peyrache and Destexhe, 2019), which all differ between the inhibitory subclasses.

### 3.4.2 Targeted viral delivery, calcium imaging, immunohistochemistry and registration techniques discriminate between excitatory and subclasses of inhibitory neurons

In this chapter, we also set out to record from and label different types of inhibitory neurons. In theory one could use promoter specific constructs and different fluorophores to achieve this. There are for example red-shifted calcium indicators such as jRGECO (Dana et al., 2016), which we could combine with the more traditional GCaMP indicator to image simultaneously the excitatory and inhibitory neurons. However, we would need to image 4 different fluorophores at the same time and given the existing 2-photon imaging setup this is not technically feasible. In addition, unlike the rodents there are no transgenic ferret lines so targeting the different types of neurons would be quite difficult based on viral strategies alone. That is why we set out to establish a protocol for targeted viral delivery, calcium imaging, immunohistochemistry and image registration in order to achieve this goal (Figure 3.8). Similar approaches have been previously applied to mouse (Khan et al., 2018) and ferret (Wilson et al., 2017) visual cortices. Despite this, due to anatomical and other differences such the location of AC vs visual cortex, the age of the animals involved (Wilson et al. (2017) used juvenile ferrets), differences in expression of the virus between areas, optimisation was needed to implement the approach successfully in our experiments. We found that the main factors that needed to be considered in the histology were primary antibody concentration, staining times, antigen retrieval, detergent concentration, temperature (see Figure 3.9 and subsection 3.2.5). The successful application of image registration through a 3D affine transformation was also critical for the process and presented its own challenges (Figure 3.10). In the end, we were able to establish a reliable protocol for labelling and recording from PV, SOM and GAD in adult ferret AC (Figure 3.11).

In our imaging experiments, we opted for the faster but less sensitive indicator GCaMP6f, comparing to the other two versions in the GCaMP6 indicator “family”, medium (6m) and slow (6s) (Chen et al., 2013). The main reason for this is that we were worried about the calcium buffering properties of the more sensitive versions (6m and 6s) on the inhibitory neuronal responses (Bootman et al., 2018; McMahon and Jackson, 2018), e.g. parvalbumin (PV) itself is a calcium binding protein. In addition, the most numerous INH type (PV basket cells) have faster channel kinetics and dynamics (Tremblay et al., 2016) so some of this dynamic range might be obscured

by the slower indicators. However, this comes at the cost of greatly reduced sensitivity and given the small number of inhibitory neurons, for future experiments it might be beneficial to utilise some of the more sensitive versions, such as GCaMP6m or GCaMP6s. A growing number of studies have used these indicators in inhibitory neurons (Wilson et al., 2018; Adler et al., 2019) without necessarily encountering difficulties related to calcium buffering. The slower kinetics might also not be very problematic as better models of spike inference from calcium data become available; for example, a recent study by Greenberg et al. (2018) was able to achieve state-of-the-art performance in spike inference with a detailed biophysical model of GCaMP6s. In addition, a newer generation of calcium indicators have become available, which have greatly improved sensitivity and kinetics (Dana et al., 2019) which could be combined with the mDlx construct (Dimidschstein et al., 2016) in order to target inhibitory neurons. Finally, our choice of sound stimuli can be improved by using AM tones rather than our current pure tone stimuli as these are known to be a lot more driving (Johnson et al., 2020). This will allow us to sample neuronal responses more effectively and provide us with more powerful statistics to investigate the similarities and difference between the different classes of neurons.

### 3.4.3 The tuning properties of excitatory and inhibitory neurons in ferret auditory cortex are remarkably similar

Given the conflicting evidence regarding the tuning properties of excitatory and inhibitory neurons in AC (Moore and Wehr, 2013; Li et al., 2015; Mesik et al., 2015; Maor et al., 2016; Liang et al., 2019), namely the number of tuned neurons in each neuronal class and tuning sharpness, we wanted to explore them in more detail. In addition, prior studies in mouse and ferret visual cortices showed significant differences between the properties of inhibitory neurons in rodents and higher mammals, with results in rodents reporting broader selectivity and more passive “pooling” of orientation tuning in inhibitory neurons, while inhibitory neurons in ferrets showed sharper tuning and participated in the same functional networks as excitatory neurons (Hofer et al., 2011; Kerlin et al., 2010; Scholl et al., 2015; Wilson et al., 2017). In both our microelectrode (Figure 3.12A) and calcium imaging data (Figure 3.12B), we did not observe significant differences in the proportions of responsive or selective neurons between excitatory and inhibitory neurons. In our calcium imaging data, where we could explore the properties of different subclasses of inhibitory neurons we also did not observe differences in the

relative proportions of different subclasses relative to each other and excitatory neurons (Figure 3.12C). It is noteworthy that the SOM subclass had a smaller proportion of responsive and tuned neurons and came close to but did not pass statistical significance (Table 3.3).

We also explored the properties of the neuronal FRAs (Figure 3.13A). We found that the preferred frequency of the neurons, measured either as the frequency eliciting the strongest response (BF) or the frequency that elicits a response at the lowest sound level (CF) did not differ between excitatory and inhibitory neurons in both our electrophysiology (Figure 3.13B,D) and calcium imaging data (Figure 3.13C,E). There is no a priori reason to expect such a difference so this finding is not entirely surprising. A more interesting comparison was that of the neuronal tuning widths. There, again we did not observe a significant difference in the excitatory compared to the inhibitory neurons in either the Q10 and Q30 measures (Figure 3.13F,G). Despite the lack of such a difference in tuning width in our aggregate electrophysiology data, we cannot rule out differences in the individual subclasses of inhibitory neurons. As we discussed in subsection 3.4.1, the extent of separation of different subclasses of inhibitory neurons based on spike features varies and PV fast-spiking (FS) neurons are probably best isolated while SOM, VIP and non-VIP will have more intermediate features and hence some classification error is inevitable (Yu et al., 2019; Royer et al., 2012). In addition, even if we had near perfect separation, given the highly conflicting accounts in the literature regarding the tuning widths of PV, SOM and VIP (Moore and Wehr, 2013; Li et al., 2015; Mesik et al., 2015; Maor et al., 2016; Liang et al., 2019) it is not evident that combining the different subclasses together would show a clear difference. However, the lack of difference in the proportions of tuned neurons (Figure 3.12) and the tuning properties themselves (Figure 3.13) are consistent with findings from ferret primary visual cortex where such differences between the two classes were also not observed (Wilson et al., 2017). Thus, similarly tuned excitatory and inhibitory neuronal responses might be a more general feature of the sensory cortices of higher mammals as opposed to rodents (but see Wehr and Zador (2003)). It is unfortunate that we could not obtain reliable measures of the Q10 and Q30 values in our imaging data as then we could have resolved some of this ambiguity. As mentioned in section 3.4.2, using a more sensitive calcium indicator (GCaMP6s, GCaMP6m or GCaMP7) in addition to stimuli that evoke stronger responses in AC cells (e.g. amplitude modulated tones) should allow us to estimate these properties and resolve this ambiguity in future experiments.

We also explored the global and local variability in BF in our imaging data across imaging fields between excitatory and inhibitory neurons (Figure 3.14). We did not observe significant differences in the median BF between areas (Figure 3.14A) or in the local variability within areas (Figure 3.14B). This is interesting, taking into account our results from the previous chapter where we saw that neurons with double-peaked and complex FRAs have higher BF variability comparing to single-peaked FRAs (Figure 2.5C). In this chapter, we saw that inhibitory and excitatory neurons have very similar proportions of single, double and complex FRAs in both electrophysiology and imaging data (Figure 3.15). Thus, the observation that inhibitory and excitatory neurons have similar local BF variability (Figure 3.14B) is consistent with our earlier results. Interestingly, a previous study by Maor et al. (2016) in the mouse primary AC found that PV had much higher homogeneity of their tuning, whereas excitatory neurons were much more heterogeneous. As we did not have enough neurons from the individual subclasses to test this and only looked at all neurons that express GAD, we cannot rule out such homogeneity amongst ferret cortical PV. Future experiments will be needed to resolve this question.

### 3.4.4 Signal and noise correlations in excitatory and subclasses of inhibitory neurons reveal distinct interactions

In addition to the tuning properties of excitatory and inhibitory neurons, we also investigated their co-tuning and functional coupling as measured by the neuronal pairwise signal correlations and noise correlations, respectively. In both our microelectrode data (Figure 3.16A) and calcium imaging data (Figure 3.16B), we did not observe significant differences in the signal correlations between the excitatory-to-excitatory (EE), excitatory-to-inhibitory (EI or EG) and inhibitory-to-inhibitory (II or GG) interactions. In addition, separating the inhibitory neurons further into subclasses in our imaging data (Figure 3.17A,C), we again did not observe significant differences in the signal correlations. As mentioned earlier, our microelectrode data had higher signal correlations. This finding can potentially be explained by the fact that in our microelectrode data there was a strong bias for single-peaked neurons whereas in our imaging data the proportion of double-peaked and complex neurons was higher (Figure 3.15). As we saw in the previous chapter, single-peaked neurons had much stronger signal correlations than either double-peaked or complex neurons (Figure 2.6A,C). A study that utilised two-photon targeted patch recording found that in mouse primary AC, PV tended to

have higher signal correlations than excitatory neurons, and thus exhibited a much higher degree of co-tuning (Maor et al., 2016), whereas in our data, we did not observe such a difference (Figure 3.17A,C). This discrepancy could arise from the difference in recording techniques. In the aforementioned study, Maor et al. (2016) found that single spike data was critical in order to obtain high signal correlations, whereas the indicator that we used (GCaMP6f) is not sensitive enough to consistently pick up single spikes (Chen et al., 2013). However, as discussed earlier, these differences could also represent genuine species differences as ferret inhibitory neurons in the visual cortex are co-tuned with excitatory neurons, unlike those in the mouse (Hofer et al., 2011; Kerlin et al., 2010; Scholl et al., 2015; Wilson et al., 2017).

As well as the co-tuning of neurons, we looked at the coupling between different classes of neurons using their noise correlations (Figure 3.16C,D). Here, unlike the signal correlations, we did observe some differences between cell types. In our electrophysiology data, excitatory-to-excitatory (EE) interactions had the highest noise correlations, followed by excitatory-to-inhibitory (EI), and inhibitory-to-inhibitory (II) exhibited the lowest noise correlations. In contrast, in the calcium imaging data, we observed the opposite order, with the GAD-to-GAD (GG) interactions showing the highest noise correlations. In this case, the difference in the proportions of single-peaked, double-peaked and complex neurons amongst microelectrode and imaging data could not offer an explanation since the noise correlations between these classes were not substantially different, as we saw in Chapter 2 (see Figure 2.6B,D).

A possible explanation might be offered by the analysis windows that we used for the two types of data. For the electrophysiology data, our analysis window included the period of the stimulus presentation, whereas for the imaging data the analysis window was twice the length of the stimulus presentation, similar to other calcium imaging studies (Romero et al., 2019; Tischbirek et al., 2019). This was dictated by the very different temporal resolution of the two techniques. While the electrophysiology data has a sub-millisecond precision, the imaging data is limited by the sampling rate of the microscope instrument ( $\sim 30$ ms) and the binding kinetics of the GCaMP6f indicator, mostly its decay time ( $\sim 100$ ms) (Chen et al., 2013; Greenberg et al., 2018). The deconvolution algorithm helps to sharpen this, but the temporal limitation is still at play. Due to its longer duration, the imaging time window includes offset responses, which are usually tuned differently than onset responses (Bizley et al., 2005), whereas the electrophysiology data includes only the onset responses during

stimulus presentation. As the noise correlation calculation includes a normalization of the responses to different tones, this could affect the estimates of noise correlations. In the future, we could try analysis windows of different sizes for the imaging and electrophysiology data in order to examine the effects of windowing on the noise correlations.

When we compared noise correlations amongst inhibitory neuronal subclasses in the imaging data, we found that the excitatory-to-inhibitory and inhibitory-to-inhibitory coupling tended to be stronger than excitatory-to-excitatory (Figure 3.17B,D). A 2-photon calcium imaging study in the mouse visual cortex demonstrated stronger coupling between inhibitory neurons compared to excitatory neurons, which increased after learning to a greater extent than for excitatory-to-excitatory or excitatory-to-inhibitory coupling (Khan et al., 2018). Most inhibitory neurons have electrical synapses between each other (gap junctions) and this is a defining feature of different inhibitory neuronal subtypes, representing a major component of the connectivity between them. While excitatory neurons do not show electrical coupling in mature animals, the electrical connection probability among related inhibitory neurons remains high, but mostly in the same class (Kepecs and Fishell, 2014; Tremblay et al., 2016). Thus, the higher noise correlations that we observe could be due to such coupling.

It is worth noting that the tuning of neurons within AC can potentially differ substantially between layers (Guo et al., 2012; Winkowski and Kanold, 2013), with deeper cortical layers exhibiting smoother tonotopy (but see (Tischbirek et al., 2019) reporting similar tonotopic organization across all layers of mouse AC). The connectivity between neurons can differ across layers, and also along the tonotopic vs the isofrequency (orthogonal to the tonotopic) axes (Oviedo et al., 2010; Oviedo, 2017). As our imaging data was limited to layers 2/3, we do not know whether our findings extrapolate to other layers.

Finally, caution must be taken when interpreting neuronal signal correlations, noise correlations and correlations more generally, as recently highlighted by Harris (2020). Harris (2020) reviewed and simulated methods for detecting correlations between neuronal time series. One of the main findings was that statistical tests that assume independence between time points can result in “nonsense correlations” of erroneous statistical significance, due to autocorrelations within neural time series. The author found that the most reliable method for detecting genuine correlations, the “pseudosession” method, required that one of the time series be randomly generated by

the experimenter. Other methods could also reduce but not eliminate the risk of false positive errors; of these, the linear shift method appeared to perform best. In future experiments and further analyses, we could implement some of these methods so as to arrive at more robust and accurate measures of signal and noise correlations.



# Chapter 4

## Adaptation to reverberation in auditory cortex

“Life is an echo. What you send out, comes back. What you sow, you reap. What you give, you get. What you see in others, exists in you. Remember, life is an echo. It always gets back to you.”

---

*Siddhārtha Gautama (The Buddha)*

### 4.1 Introduction

#### 4.1.1 Adaptation and robustness to reverberation in the auditory system

As described in subsection [1.6.1](#), reverberation is the persistence of a sound after its initial production and it is caused by the reflection of the sound emitted by the source from the surfaces of nearby objects in space, such as the room walls and floor, furniture, buildings, trees in forests and people ([Huisman and Attenborough, 1991](#); [Sakai et al., 1998](#)). Although reverberation can be seen as corrupting the sounds that humans and animals encounter, it is also an integral part of natural environments, so the auditory

system must have developed mechanisms to deal with it. At the perceptual level, various studies have demonstrated the ability of human listeners to perform well in mild-to-moderate reverberation in spatial localization tasks (Hartmann, 1982; Rakerd and Hartmann, 2005). Humans are also remarkably robust to reverberation in speech and auditory object recognition tasks (Houtgast and Steeneken, 1985; Bradley, 1986; Darwin and Hukin, 2000; Culling et al., 2003; Nielsen and Dau, 2010). In contrast, such comparable robustness remains beyond the capability of automatic speech recognition devices, where performance deteriorates rapidly even under modest reverberation (Yoshioka et al., 2012; Kinoshita et al., 2016).

Many of the previous neurophysiological studies have focused on the effects of reverberation on sound localization, mostly in the subcortical parts of the auditory system. The neural correlates of the “precedence effect”, the phenomenon where reflections arriving later at the ears are suppressed relative to the onset in sound localization, have been studied extensively in the IC (Yin, 1994; Litovsky and Yin, 1998; Fitzpatrick et al., 1999; Spitzer et al., 2004; Tollin et al., 2004; Pecka et al., 2007). A study by Devore et al. (2009) showed that robust encoding of directional information in the rate responses of IC neurons was sufficient to account for the performance of human listeners, as predicted by the precedence effect. A recent computational study argued for the importance of the MSO in a potential mechanistic implementation of the precedence effect using a hemispheric model, where stronger weighting of ITD information during rising sound-energy at lower frequencies improved performance in lateralisation task (Brughera et al., 2019). Others have explored the robustness of IC neurons to AM degradation in reverberant conditions in relation to distance coding (Kuwada et al., 2012; Kim et al., 2015).

In contrast to the wealth of studies exploring neuronal adaptation to reverb in sound localisation, much less is known about the neural processing of speech and other complex natural sounds in the presence of reverberation. A study that recorded from neurons in the ventral CN of guinea pigs showed that pitch encoding of harmonic tones was severely degraded in reverberation mostly due to attenuation of temporal envelope AM and smearing of the fine temporal structure, which was paralleled by changes in human performance in a discrimination task (Sayles and Winter, 2008). Another study explored the adaptation of IC neurons in the rabbit to envelope AM degradation by reverb and found that neurons compensated their responses (Slama and Delgutte, 2015). A human study using Electroencephalogram (EEG) reported

that the speech envelope of anechoic sounds was more faithfully decoded from cortical responses compared to reverberant versions of the same sounds for attended sources compared to unattended ones (Fuglsang et al., 2017). The most direct investigation of auditory cortical adaptation to reverb to date was performed by Mesgarani et al. (2014) in the ferret. Using speech and conspecific vocalizations, the authors found that ferret AC largely removes the distortions brought about by reverberation and the clean anechoic sounds were more readily decodable from neuronal responses compared to the reverberant ones.

Converging evidence from animal electrophysiology (Rabinowitz et al., 2013; Moore et al., 2013; Mesgarani et al., 2014), and human EEG (Khalighinejad et al., 2019) and fMRI (Kell and McDermott, 2019) studies suggests that noise-invariant representations of sounds emerge at later stages in the auditory pathway by neuronal adaptation to stimulus statistics (but see also Lohse et al. (2020a)). For example, Kell and McDermott (2019) found that robustness to non-stationary noise was a hallmark of secondary vs primary human AC. In contrast to other types of stationary noise (e.g. white, pink), the corruption caused by reverberation is highly non-stationary since it depends on the the sound source itself (via the convolution with the IR). These studies, combined with the study from Fuglsang et al. (2017), suggest that AC is likely a key player in the adaptation to reverberation seen perceptually, at least for speech and object recognition, though possibly less so for sound localisation.

### 4.1.2 Summary of the study

Despite its importance to daily life, speech recognition technology and hearing aids, there have been relatively few studies directly examining adaptation to reverberation at the neuronal level and the potential mechanisms for the implementation of such adaptation. Most of the previous studies have focused primarily on sound localisation within subcortical regions, while auditory object recognition has been relatively understudied. Another limitation of previous work is the use of artificial stimuli which do not necessarily capture the richness of natural sounds and their complex interactions with reverberant acoustic spaces. Some studies have used speech stimuli (Mesgarani et al., 2014), but none presented a rich set of diverse natural sounds.

To address the shortcomings of previous studies and better understand the neural mechanisms behind adaptation to reverberation under realistic conditions, we used

a normative computational modelling approach to first generate predictions, which we then test in the ferret AC. Normative models typically start with an analytical formulation of which problem the nervous system has to solve, and propose an answer: how the nervous system “could” solve this problem. Such a principled framework is particularly important for understanding complex systems, where pure phenomenological models often do not provide satisfying answers. Normative models have had great success in explaining a wide range of aspects of sensory neural processing. For example, the way sensory neurons encode information has been phrased as a problem of most efficiently representing sensory information or making a prediction about future input, and this has explained aspects of visual and auditory neural receptive fields (Olshausen and Field, 1996; Smith and Lewicki, 2006; Singer et al., 2018). Another example is the normative theory of the “Bayesian brain”, stating that our nervous system behaves optimally under the framework of Bayesian inference, and there are studies that support this, both behavioural and neurophysiological (Körding and Wolpert, 2004; Friston, 2012).

In the current study, we started with a rich data set of anechoic natural sounds, including speech, textures and other environmental sounds, and made versions with varied amounts of reverberation. We then asked the following questions: What would be the optimal linear filters for retrieving the clean anechoic sounds from their reverberant counterparts? What properties do they have and do they resemble the tuning properties of auditory cortical neurons? To achieve this, we trained a linear model to retrieve the clean anechoic sounds from their reverberant versions. We compared our model kernels to the STRF of ferret auditory cortical neurons using “Neuropixels” probes (Jun et al., 2017). We found that the model retrieves some known properties of neurons, such as frequency tuning and temporally asymmetric STRF profiles with an excitatory field followed by an inhibitory field. The model also made two novel predictions: (1) the inhibitory part of the STRF scales with the amount of reverberation, extending further back in time so as to cancel the excess reverberant energy, while the excitatory part is largely unchanged; (2) the inhibitory scaling of the STRF is frequency dependent. We tested these predictions by recording neuronal spiking responses in ferret AC. We also explored putative mechanisms allowing for the implementation of this adaptation that rely on local inhibitory interneurons. Together, our data suggest that auditory cortical neurons adapt to reverberation by adjusting their spectro-temporal filtering properties.

## 4.2 Materials and methods

### 4.2.1 Animals and surgical procedures

All animal procedures were approved by the local ethical review committee of the University of Oxford and performed under license from the UK Home Office. Four female and two male ferrets (*Mustela putorius furo*; Marshall BioResources, UK) were used in the electrophysiology experiments.

Terminal electrophysiological recordings were performed under general anaesthesia. Anaesthesia was induced with an intramuscular injection of ketamine (Vetalar; 5 mg/kg) and medetomidine (Domitor; 0.02 mg/kg), and was maintained with a continuous intravenous infusion of these two drugs in Hartmann's solution with 3.5% glucose and dexamethasone (0.5 mg/ml/hr). The animal was intubated and artificially ventilated. Respiratory rate, end-tidal CO<sub>2</sub>, electrocardiogram and blood oxygenation were continuously monitored throughout the recording session. Eye ointment (Maxitrol; Alcon, UK) was applied throughout and body temperature was maintained at 36-38°C. Atropine (Atrocare; 0.06 mg/kg i.m.) was administered every 6 hours, or when bradycardia or arrhythmia was observed. Ferrets were placed in a custom-built stereotaxic frame and head stability was achieved using ear bars and a mouthpiece. After shaving the scalp and injecting bupivacaine (Marcain, 2mg/kg s.c.), the skin was incised and the temporal muscle removed. A steel holding bar was secured to the skull using dental cement (SuperBond; C&B, UK) and a stainless steel bone screw (Veterinary Instrumentation, UK). A circular craniotomy 10 mm in diameter was drilled over ferret AC, and the dura over the AC was removed. After durotomy, the brain surface was covered with a solution of 1.25% agarose in 0.9% NaCl, and silicon oil was applied to the craniotomy regularly throughout recording. Ear bars were removed, and the ferret was placed in an electrically isolated anechoic chamber for recording. Recordings were carried out in the left AC. An Ag/AgCl external reference wire was inserted between the dura and the skull from the edge of craniotomy. A Neuropixels Phase 3a probe (Jun et al., 2017) was inserted orthogonally to the brain surface through the entire depth of AC. Data were acquired at a 30 kHz sampling rate using SpikeGLX software (<https://github.com/billkarsh/SpikeGLX>) and custom Matlab scripts.

## 4.2.2 Sound presentation

Stimuli were presented binaurally via Panasonic RP-HV094E-K earphone drivers, coupled to otoscope speculae inserted into each ear canal, and driven by a System 3 RP2.1 multiprocessor and headphone amplifier (Tucker-Davis Technologies). The output response of the speakers was measured using a Brüel & Kjær calibration system with a GRAS 40DP microphone coupled to the end of the silicone tube. An inverse filter was applied to the speaker output to produce a flat spectral response ( $\pm 3\text{dB}$ ) over the stimulus frequency range (200Hz-22kHz). Sound intensity was calibrated with an Iso-Tech TES-1356-G sound level calibrator. The speculae were sealed in place with Otoform (Dreve Otoplastik).

## 4.2.3 Spike sorting and data analysis

The recorded signal was processed offline by first digitally highpass filtering at 150Hz. Common average referencing was performed to remove noise across electrode channels (Ludwig et al., 2009). Spiking activity was then detected and clustered using Kilosort2 software (Pachitariu et al., 2016) (<https://github.com/MouseLand/Kilosort2>). Responses from single neurons were manually curated using Phy (<https://github.com/cortex-lab/phy>) if they had stereotypical spike shapes with low variance and their autocorrelation spike histogram showed a clear refractory period. Spikes from a given cluster were often measurable on 4-6 neighboring electrode channels, facilitating the isolation of single units. Only well isolated units were included in subsequent analyses.

To exclude unreliable and noisy units – we employed a method proposed by Sahani and Linden (2003) to estimate variance related to the signal (sensory input) and the variance related to ‘noise’. Following the terminology used by Sahani and Linden (2003), the variance of a response related to the input is called the signal power, and the variance of a response unrelated to the sensory input is called the noise power. The signal power is the variance of the mean PSTH relative to the variance seen in individual trials (i.e. total power):

$$SP = \text{Signal Power} = \frac{1}{N-1} (N\text{var}(y[t]) - TP) \quad (4.1)$$

$$TP = Total\ Power = \frac{1}{N} \sum_{n=1}^N \text{var}(R_n[t]) \quad (4.2)$$

Where  $R$  is the number of spikes elicited in time bin  $[t]$  in stimulus repetition number  $n$ , and  $N$  is the number of stimulus repetitions. The variable  $y[t]$  stands for the firing rate of the neuron averaged across the  $N$  trials,  $\text{var}()$  stands for variance. In equation 4.2 the variance is computed across time for each trial  $n$ .

The amount of variance which is then left is the noise power:

$$NP = Noise\ Power = TP - SP \quad (4.3)$$

This allows us to define a noise ratio (NR):

$$NR = Noise\ Ratio = \frac{NP}{SP} = \frac{total\ variance - explainable\ variance}{explainable\ variance} \quad (4.4)$$

An NR of 0 indicates that responses were identical for repeated stimulus presentations. Higher NR indicates that responses are less reliable. Units with  $NR > 40$  in any one stimulus condition, i.e., whose explainable variance was  $< 2.3\%$  of the total variance, were excluded from further analysis.

#### 4.2.4 Virtual acoustic space

In order to simulate reverberation, we used the “Roomsim” software (Campbell et al., 2005), a Virtual Acoustic Space (VAS). This software creates a cuboidal room of arbitrary  $x$ ,  $y$  and  $z$  dimensions and simulates its acoustic properties. The simulations are based on the room-image method (Allen and Berkley, 1979; Heinz, 1993; Shinn-Cunningham et al., 2001). Roomsim simulates the properties of the room for a listener at a particular point and orientation in space. One difference between the standard image method and Roomsim is that the latter incorporates the absorption properties of different materials. Different materials absorb sound differently and their properties can be summarized by their frequency-dependent absorption coefficients. In principle, the amount of reverberation in a room will depend on its size, shape and the material from which the walls are made. We decided to vary the amount of reverberation by

changing the room size whilst keeping the other parameters fixed. Three different rooms were created:

1. **Anechoic room** ( $RT_{60} = 30\text{ms}$ )
2. **Small room** ( $3 \times 0.3 \times 0.3\text{m}^3$ ,  $RT_{60} = 1.2\text{s}$ )
3. **Large room** ( $15 \times 1.5 \times 1.5\text{m}^3$ ,  $RT_{60} = 3.1\text{s}$ )

Note that the anechoic room does not have a clearly defined “shape”, as the walls are completely absorbing the sound energy and the resulting IRs look identical in the small anechoic room ( $3 \times 0.3 \times 0.3\text{m}^3$ ) and the large anechoic room ( $15 \times 1.5 \times 1.5\text{m}^3$ ), since they are only determined by the relative orientation and distances of the sound source and receiver. Roomsim also simulates the acoustic properties of the receiver’s head, represented by the Head-Related Transfer Function (HRTF). This is important, because in realistic acoustic conditions the sound interacts with the subject and becomes distorted before it reaches the ear drum. In all simulations, we used the same ferret HRTF provided from measurements previously made in the lab on a real ferret. The joint filtering properties of the ferret’s head and the room were simulated simultaneously by Roomsim to produce Binaural Room Impulse Response (BRIR), one for each ear. The ferret head was simulated in the VAS, positioning it at  $1/2$  of the room’s width (0.15m for the small, 0.75m for the large),  $1/4$  of the room’s length (0.75m for the small, 3.75m for the large) and 0.15m from the floor. The sound source was positioned at a distance of 1.5m,  $0^\circ$  azimuth and  $0^\circ$  elevation relative to the ferret’s head in all three conditions.

The anechoic BRIRs was generated in exactly the same way (room size, source and head positions were the same) as the echoic input but the difference was that the absorption coefficients for all the walls were set to 1, i.e. all the sound was absorbed and there was no echo. It was important to do this to ensure that the only difference between the sounds in the different conditions was the presence/absence of echo, since the BRIR is dependent on the distance and positions of the source and the head.

#### 4.2.5 Sound stimuli

We used the same stimuli to train the normative model and in the ferret experiments. The dataset consisted of anechoic sounds, containing human speech and other natural

sounds, such as cracking branches, footsteps, rolling stones, running water and others. The total duration of the sounds was 600s. The stimuli presented to the ferrets were a representative subset of those used to train the normative model, for a total of 80s. These were divided into two different 40s blocks. Each block was made by concatenating 20 snippets of 2s duration each. In addition, a smoothing function (cosine ramp) was applied from 0–0.25s and 1.75–2s to each snippet before concatenation. The sounds were also band passed between 200Hz–20,000kHz. The sounds were then convolved with the BRIR made in Roomsim as described above, for the three different conditions: anechoic, small room and large room. The two 40s blocks were presented to the ferrets in the three different conditions, thus there were six different 40s blocks presented. The presentation of the six blocks was repeated ten times, with the order of the six blocks randomized in each repeat and 5s of silence between blocks.

#### 4.2.6 Cochlear model

We used a power-spectrogram based model of cochlear processing as described in [Rahman et al. \(2020\)](#). Briefly, a spectrogram was produced from the sound waveform by taking the power spectrum through a short-time Fourier transform (STFT) using 10-ms Hanning windows, overlapping by 5 ms. The power of adjacent log-spaced frequency channels was summed using overlapping triangular windows (using code adapted from `melbank.m`, <http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html>) containing frequency channels ranging from 400Hz to 19kHz centre frequencies. The power in each time-frequency bin was converted to log values and any value below a threshold (-40) was set to that threshold.

#### 4.2.7 Model kernels

The dereverberation normative model consisted of a set of linear kernels, one for each frequency channel in the anechoic cochleagram, i.e. a total of 30 different kernels for each reverberant condition. That is, the two different reverberant conditions (small and large) were used to fit a separate set of kernels. These can be summarised by the following equation:

$$\hat{A}_{f't} = \sum_{f=1}^{f_{max}} \sum_{h=1}^{h_{max}} K_{f'fh} R_{f(t-h)} + b_{f'} \quad (4.5)$$

Here,  $\hat{A}_{f't}$  represents the predicted anechoic cochleagram for time bin  $t$  and frequency channel  $f'$ . The reverberant cochleagram is represented by  $R_{ft}$ , where  $t$  is the time bin and  $f$  is the frequency channel.  $K_{f'fh}$  is a 3D tensor which contains all the model weights for a given reverberant condition, where  $f'$  is the anechoic cochleagram frequency channel,  $f$  is the reverberant frequency channel for this condition and  $h$  is the history (time lag). Tensor  $K_{f'fh}$  is composed of a number of kernels  $k_{fh}$ , one for each frequency channel  $f'$  in the anechoic cochleagram. The  $b_{f'}$  stands for the bias term, one for each anechoic cochlear frequency channel that we are trying to predict. The model weights were fitted so as to minimize the Mean Squared Error (MSE) between the predicted anechoic cochleagram  $\hat{A}_{f't}$  and the actual anechoic cochleagram  $A_{f't}$ , subject to L2 regularisation (“ridge” regression). To select the regularisation hyperparameter  $\lambda$ , we performed 10-fold cross-validation, using 90% of the data for the training set and 10% for the “tuning” set<sup>1</sup>. The weights were optimized using the glmnet package (GLM, J. Qian, T. Hastie, J. Friedman, R. Tibshirani, and N. Simon, Stanford University, Stanford, CA; see [http://web.stanford.edu/~hastie/glmnet\\_matlab/index.html](http://web.stanford.edu/~hastie/glmnet_matlab/index.html)). Note that here onward we will typically refer to the individual kernels  $k_{fh}$ <sup>2</sup> rather than the full tensor  $K_{f'fh}$ .

## 4.2.8 Neuronal STRFs

We fitted spectro-temporal receptive fields (STRFs) to the neuronal data which we could then compare to the normative model kernels (Theunissen et al., 2001). Various extensions of the basic STRF model exist, most of them adding a non-linearity function to account for the known non-linearities of neurons, capturing some of the interactions between neurons or other contextual factors (Ahrens et al., 2008; Calabrese et al., 2011; Harper et al., 2016; Rabinowitz et al., 2011; Williamson et al., 2016; Willmore et al., 2016). In our study, we used the basic linear STRF, to allow comparison with our normative model. The STRF model can be summarised by the following equation:

---

<sup>1</sup>We refer to the data set used for optimization of the hyperparameters as “tuning” and the one for testing the performance as “reporting”. These are more traditionally known in machine learning as “development”/“validation” set and “test” set.

<sup>2</sup>Note that we also sometimes refer to the model kernels simply as  $k$ . The neuronal STRFs are represented by  $w$  instead, in order to distinguish them from the model kernels  $k$ .

$$\hat{y}_{nt} = \sum_{f=1}^{f_{max}} \sum_{h=1}^{h_{max}} W_{nfh} R_{f(t-h)} + b_n \quad (4.6)$$

Here,  $\hat{y}_{nt}$  stands for the predicted firing rate of neuron  $n$  at time  $t$ .  $R_{ft}$  again stands for the reverberant cochleagram at time  $t$  and frequency channel  $f$ .  $W_{nfh}$  describes the STRF weights for neuron  $n$ , frequency  $f$  and history (time lag)  $h$ . The  $b_n$  stands for the bias term of neuron  $n$ . Notice the similarity to the normative model kernels from section 4.2.7. In both cases, we use the reverberant cochleagram as an input (“small” or “large”) and try to find the best linear mapping to the output. The important difference is that in the case of neuronal STRFs we are trying to predict the neuronal firing rate, whereas in the model kernel case we are trying to predict the power in cochlear anechoic frequency bands. In this case, we perform the fit by finding the weights that minimise the MSE between the predicted firing rate  $\hat{y}$  and the actual firing rate  $y$ , subject to L2 regularisation, as for the model kernels. Again, 10-fold cross-validation was performed with a 90%/10% split into training and tuning sets. Similar to the model kernels, we will analyse the STRFs individually in the form  $w_{fh}$  rather than collected together as a tensor  $W_{nfh}$ .

### 4.2.9 Quantification of model kernels and STRFs

To quantify the temporal profile evolution of the model kernels and neuronal STRFs, we chose three different measures:

1. Center of Mass (COM)
2. Peak Time (PT)
3. Peak Height (PH)

To compute them, we first obtained the averaged excitatory and inhibitory temporal profiles of the model kernels/STRFs as follows:

$$w_h^+ = \frac{1}{f_{max}} \sum_{f=1}^{f_{max}} [w_{fh}]_+ \quad (4.7)$$

$$w_h^- = \frac{1}{f_{max}} \sum_{f=1}^{f_{max}} [w_{fh}]_- \quad (4.8)$$

where  $w_{fh}$  is the neuronal STRF (the equation is the same for the model kernels but with  $k$  instead of  $w$ ) with  $f$  and  $h$  subscripts denoting frequency and history,  $f_{max}$  is the number of frequencies in the kernel/STRF  $w_{fh}$  (30). The notation  $[w_{fh}]_+$  and  $[w_{fh}]_-$  stand for the element-wise operations  $\max(w_{fh}, 0)$  and  $\min(w_{fh}, 0)$ :

$$[w_{fh}]_+ = \begin{cases} w_{fh} & \text{if } w_{fh} \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (4.9)$$

$$[w_{fh}]_- = \begin{cases} w_{fh} & \text{if } w_{fh} \leq 0 \\ 0 & \text{otherwise} \end{cases} \quad (4.10)$$

The variables  $w_h^+$  and  $w_h^-$  represent the frequency averaged positive-only and negative-only parts of the kernels,  $[w_{fh}]_+$  and  $[w_{fh}]_-$ .

The *COM* was defined as follows:

$$COM^+ = \frac{1}{\sum_{h=1}^{h_{max}} w_h^+} \sum_{h=0}^{h_{max}} L_h w_h^+ \quad (4.11)$$

$$COM^- = \frac{1}{\sum_{h=1}^{h_{max}} w_h^-} \sum_{h=0}^{h_{max}} L_h w_h^- \quad (4.12)$$

The  $COM_{large}^+$  and  $COM_{small}^+$  variables denote the center of mass for the positive (excitatory) component of the kernels for the large and small rooms while  $COM_{large}^-$  and  $COM_{small}^-$  the negative (inhibitory) center of mass.  $L_h$  denotes the time lag in the history of the STRF/model kernel, ranging from  $L_1 = 0\text{ms}$  to  $L_{h_{max}} = 200\text{ms}$ .

The peak time (*PT*) was the time at which the excitation and inhibition in the frequency averaged model kernels/neuronal STRFs  $w_h$  peaked:

$$PT^+ = \underset{h}{\operatorname{argmax}} w_h^+ \quad (4.13)$$

$$PT^- = \operatorname{argmax}_h |w_h^-| \quad (4.14)$$

The peak height ( $PH$ ) was the maximum value of the excitation-only and the maximum absolute value of the inhibition-only frequency averaged model kernels/neuronal STRFs:

$$PH^+ = \max(w_h^+) \quad (4.15)$$

$$PH^- = \max(|w_h^-|) \quad (4.16)$$

#### 4.2.10 Simulated model neuron

In order to confirm that the changes that we observe are truly adaptive, we used a simulated model neuron to generate responses that we would then fit using our standard procedure and apply the same downstream analyses. This was done in the following way. First, we fitted a single STRF for every neuron as described in section 4.2.8. However, in this case we use the data from the “small” and “large” conditions together rather than fitting separate STRFs to the two conditions as we did previously. Next, we fitted an output sigmoid non-linearity by first generating a firing rate prediction  $\hat{y}$  according to equation 4.6 from section 4.2.8 using this single STRF and then finding the sigmoid that best fits the actual firing rate  $y$  according to the following equation:

$$y = a + \frac{b}{1 + e^{-(\hat{y}-c)d}} \quad (4.17)$$

Where  $y$  is the output of the point non-linearity,  $\hat{y}$  the predicted firing rate from the linear stage (see equation 4.6), and the four parameters  $a$ ,  $b$ ,  $c$  and  $d$  have the following meaning:  $a$  = y-axis offset;  $b$  = y-axis range;  $c$  = x-axis offset;  $b/d/4$  = gain. Next, in order to simulate realistic neuronal noise, we use the  $y(t)$  output at each time point  $t$  as the parameter ( $\lambda$  = mean and variance) of a Poisson distribution to generate 10 “virtual” trials. We do this for the small and large conditions. Finally, we fit STRFs for the two conditions separately using these simulated model responses in place of the actual firing rate as we did previously and we then analyse the resulting STRFs as outlined in section 4.2.9 above.

## 4.3 Results

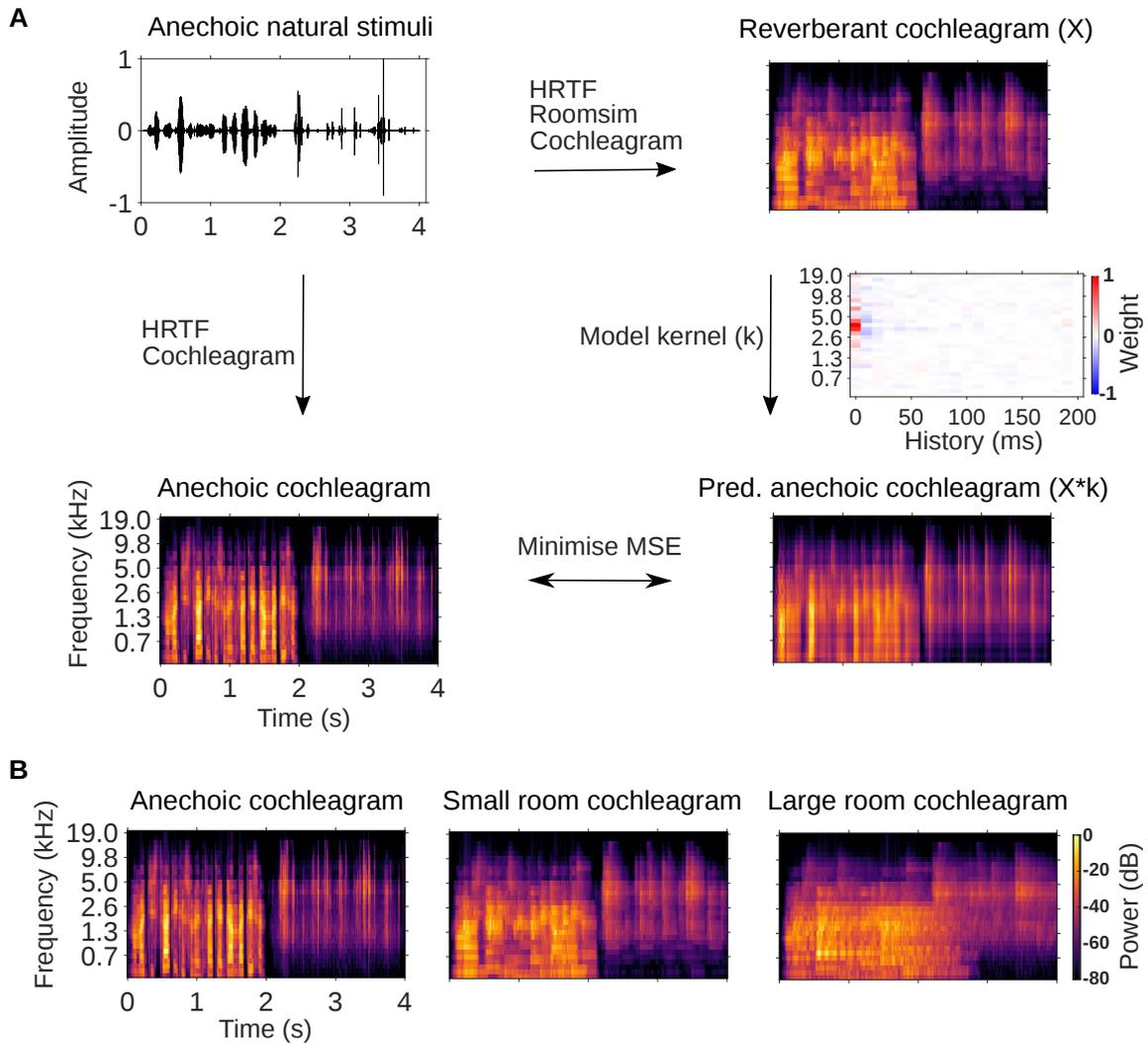
### 4.3.1 Normative dereverberation model

When building normative models of neuronal function, the first step is to optimise the model to perform a certain task. Therefore, we first set out to obtain the kernels that best predict the anechoic sounds from the reverberant sounds. In order to do so, we train an echo-removal model, which we will refer to as the “normative dereverberation model” or “normative model” for short.

To train this model, we used stimuli comprised of a rich data set of natural sounds, including speech, texture sounds and other environmental sounds. There were a total of 600s of natural sounds. The original sounds were anechoic, so in order to simulate the effect of adding reverberation to them we used the VAS simulator Roomsim (Campbell et al., 2005). This software creates a rectangular room of inputted dimensions and based on the room-image method it simulates the room’s acoustics (Allen and Berkley, 1979; Heinz, 1993; Shinn-Cunningham et al., 2001). This also includes a representation of the filtering properties of the head of a ferret captured by the HRTF (Cheng and Wakefield, 2001). The joint filtering properties of the ferret’s head and the room are simulated simultaneously by Roomsim to produce Binaural Room Impulse Responses (BRIRs). The final sounds are generated by convolving the BRIRs with the raw anechoic sounds.

There were three reverberation conditions (Figure 4.1B): “anechoic”, “small” - a small tunnel-shaped room with modest reverberation, and “large” - a large tunnel-shaped room with severe reverberation (see Materials and Methods subsection 4.2.4). The sounds were identical across conditions apart from the amount of reverb present. It is well known that sounds are louder in more reverberant environments as less of their energy is absorbed per unit time (Kuttruff, 2019). Therefore, to control for sound pressure level differences across conditions, the resulting sounds were scaled to have the same overall level in the three reverberant conditions.

After the different sounds were generated, we converted them to “cochleagrams”. A cochleagram is similar to the more familiar spectrogram and is intended to capture the representation of a sound by the auditory nerve (Brown and Cooke, 1994; Rahman et al., 2020). The cochleagrams were generated by processing sound waveforms with a filter bank that mimicked the general signal processing properties of the cochlea



**Fig. 4.1 Dereverberation normative model.**

**A**, Schematic of the workflow for generating the dereverberation normative model. The raw sound waveform (top left panel) shows a 4s sample of our anechoic audio recordings, which consisted of 600s of natural sounds, such as speech, textures and other environmental sounds. Simulated reverberation in either a small ( $0.27\text{m}^3$ ;  $RT_{60}=1.2\text{s}$ ) or large ( $33.75\text{m}^3$ ;  $RT_{60}=3.1\text{s}$ ) room was added to these anechoic sounds using Roomsim (Campbell et al., 2005), which also included a ferret HRTF (see subsection 4.2.4). The resulting 3 sets of sounds (anechoic, small room, and large room) were then filtered using a model cochlea to produce 3 cochleagrams (top right and bottom left panels). A linear model was fitted to predict the different anechoic frequency bands in the anechoic cochleagram from their reverberant counterparts ( $X$ ), resulting in model kernels similar to spectro-temporal receptive fields (STRF-like,  $k$ , see also subsections 4.2.7, 4.2.8). All reverberant frequency bands were used to predict each anechoic frequency separately. The predicted anechoic cochleagram (bottom right panel) was generated by the convolution of each kernel ( $X * k$ ) with the reverberant cochleagram (small or large room). **B**, Sample cochleagrams from the anechoic (left panel), small room (middle panel), and large room (right panel) reverberant conditions.

and extracted the amplitude envelope from each filter (see subsection 4.2.6 for more details).

We next trained the normative model to recover the anechoic cochleagram using as an input either the small or large room cochleagrams. The normative model was comprised of a set of kernels, one for each frequency in the anechoic cochleagram (see subsection 4.2.7 and equation 4.5). In order to predict the power in a specific anechoic cochlear frequency band, all frequency bands in the reverberant cochleagrams were used along with their recent history, up to 200ms. This results in a set of positive and negative weights in the prediction kernels. The prediction for each anechoic frequency band is obtained by the convolution ( $X * k$ ) of the dereverberation kernel  $k$  with each reverberant cochleagram  $X$  (Figure 4.1A). The model was trained separately using the small and large room cochleagrams as inputs.

### 4.3.2 Comparison of model kernels to neuronal receptive fields

After training the normative dereverberation model, we wanted to explore its properties and compare them to those of ferret auditory cortical neurons. This can tell us whether the goal of the model (reverberation invariance) could be also one of the presumed roles of AC. In order to do this comparison, we presented natural sounds to anaesthetized ferrets in the different reverberant conditions (“anechoic”, “small” and “large”, see also subsection 4.2.5) and simultaneously recorded the spiking activity of neurons in primary AC using the high-density Neuropixels probes (Jun et al., 2017). This allowed us to fit spectro-temporal receptive fields<sup>3</sup> (STRFs) to the neuronal data, which we could then compare to the normative model kernels (see subsections 4.2.7, 4.2.8 and equations 4.5, 4.6).

STRFs are linear kernels relating the cochleagram of the sound stimulus to the instantaneous firing rate of the neuron (Theunissen et al., 2001). They measure the time delay and the frequency tuning of the neuron to auditory inputs and can be thought of as a matched filter. The positive regions represent the preferred time lags and frequency channels where increasing the input increases the neuron’s spike rate. Conceptually, we can think of this as the “excitatory” part of the receptive field. Similarly, negative regions indicate the time lags and frequency channels where

---

<sup>3</sup>We will use the term STRFs to refer to the neuronal receptive fields and the term model kernels to refer to the ones generated by the normative dereverberation model.

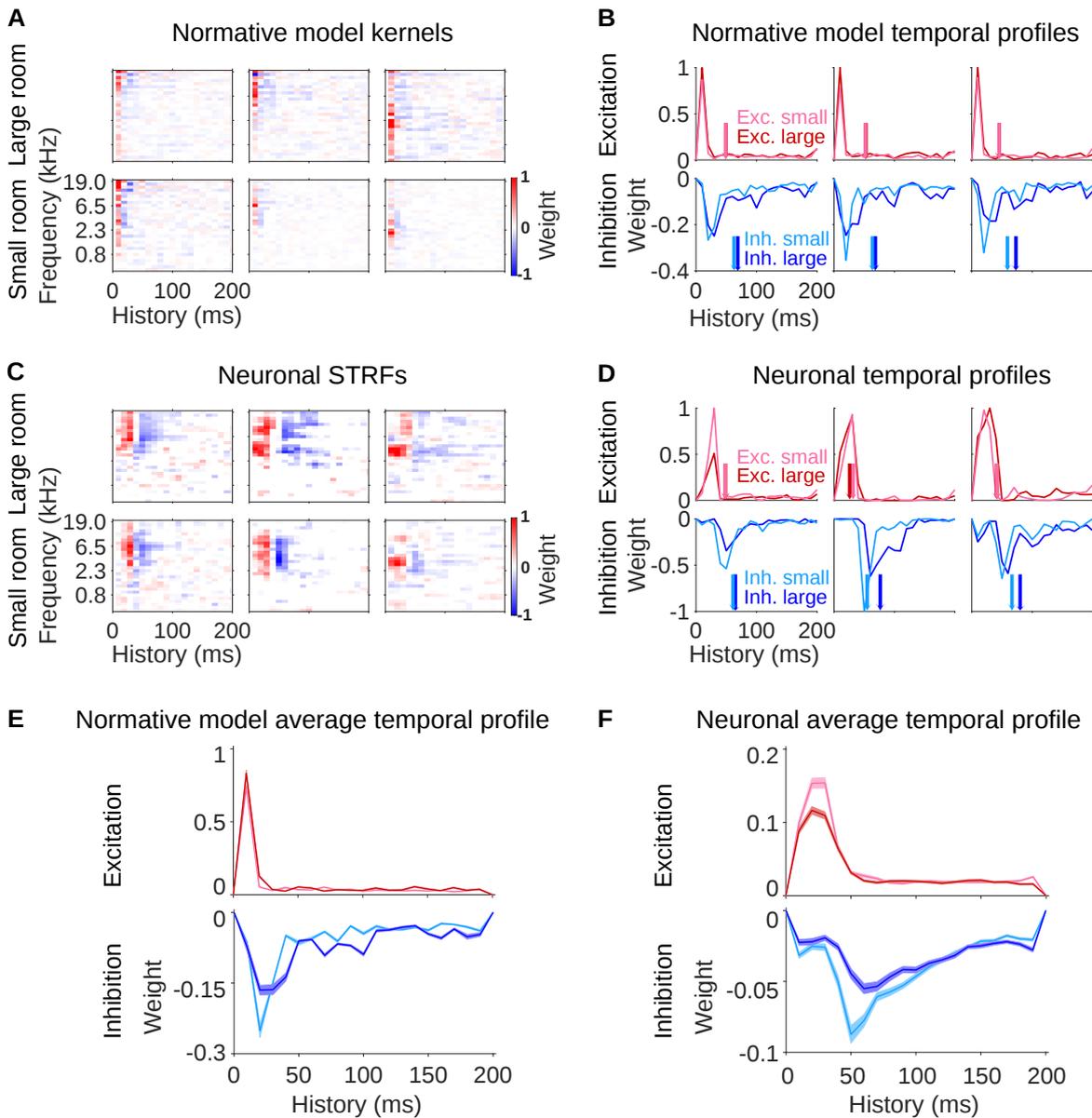
increasing the input decreases the neuron’s spike rate. We can think of this as the “inhibitory” part of the receptive field. In the case of the normative model, these “excitatory” and “inhibitory” regions of the kernel would represent spectro-temporal features in the reverberant cochleagrams that are associated with increased or decreased power in the anechoic cochleagram respectively (Figure 4.1A).

Typical example model kernels and neuronal STRFs are shown in Figure 4.2A and C, respectively. The majority of the model kernel and neuronal STRF weights were localized within a given frequency region i.e. they both exhibited “frequency tuning”, which is a well known phenomenon for auditory cortical neurons (Bizley et al., 2005). In the case of the model, this is also to be expected since we are trying to predict the power in a given anechoic cochleagram frequency band from the different frequency bands in the reverberant cochleagrams (Figure 4.1A). Naturally, most of the predictive power will lie around the given frequency band. This was reassuring and gave us confidence that our model was doing something sensible.

### 4.3.3 Model kernels capture the temporal adaptive properties of auditory cortical neurons

We next sought to investigate in more detail the temporal differences between the two reverberant conditions in terms of the excitatory and inhibitory profiles of the model kernels and neuronal STRFs. There were a few very salient features present in both model kernels and STRFs. Generally, the model kernels and STRFs exhibited a temporally asymmetric structure, where excitatory weights tended to occur initially (Figure 4.2A,C) and then reverse sign to a longer spanning inhibitory field. This pattern is commonly observed in the STRFs of auditory cortical neurons (Linden et al., 2003; Harper et al., 2016; Rahman et al., 2019). These temporal excitatory and inhibitory effects in the neuronal STRFs and model kernels became very apparent when we averaged all the positive and negative weights separately across frequencies (Figure 4.2B,D). This pattern was observed across frequencies for both the model and the neuronal data (Figure 4.3, Figure 4.4). This further increased the confidence in our model as it was reproducing another known property of auditory cortical neurons.

An important prediction we observed in the model was that the excitatory part of both model kernels seemed to remain constant between the “small” and “large” echoic conditions, whereas the inhibitory field tended to be more delayed and temporally



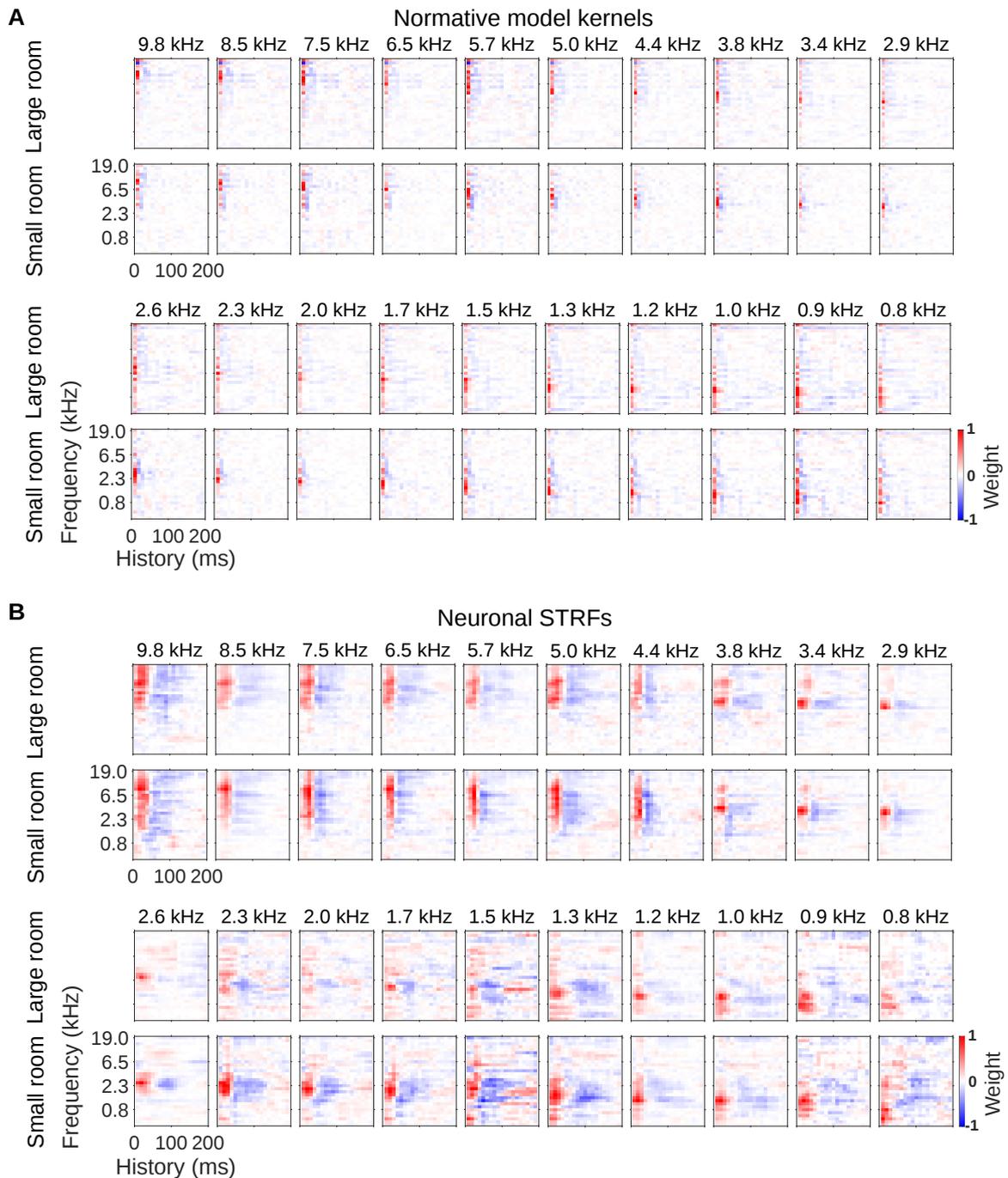
**Fig. 4.2 Comparison of normative model kernels and neuronal STRFs.**

**A**, Sample model kernels resulting from the dereverberation model. The model kernels for 3 example frequency channels (normative model kernels) are shown for the large (top row) and small (second row) rooms. Weight coefficients for each frequency (y-axis) and history (x-axis) combinations are plotted. Red indicates positive coefficients (i.e. “excitation”), and blue indicates negative coefficients (i.e. “inhibition”; colour bar right).

**B**, The first row of graphs shows the temporal profile of the averaged positive kernel coefficients for the large (dark red) and small (light red) rooms. The center of mass of the excitation,  $COM^+$ , is indicated by the arrows, which follow the same colour scheme. The second row plots the temporal profiles of the averaged negative kernel coefficients for the large (dark blue) and small (light blue) rooms. The  $COM^-$  is indicated by the coloured arrows.

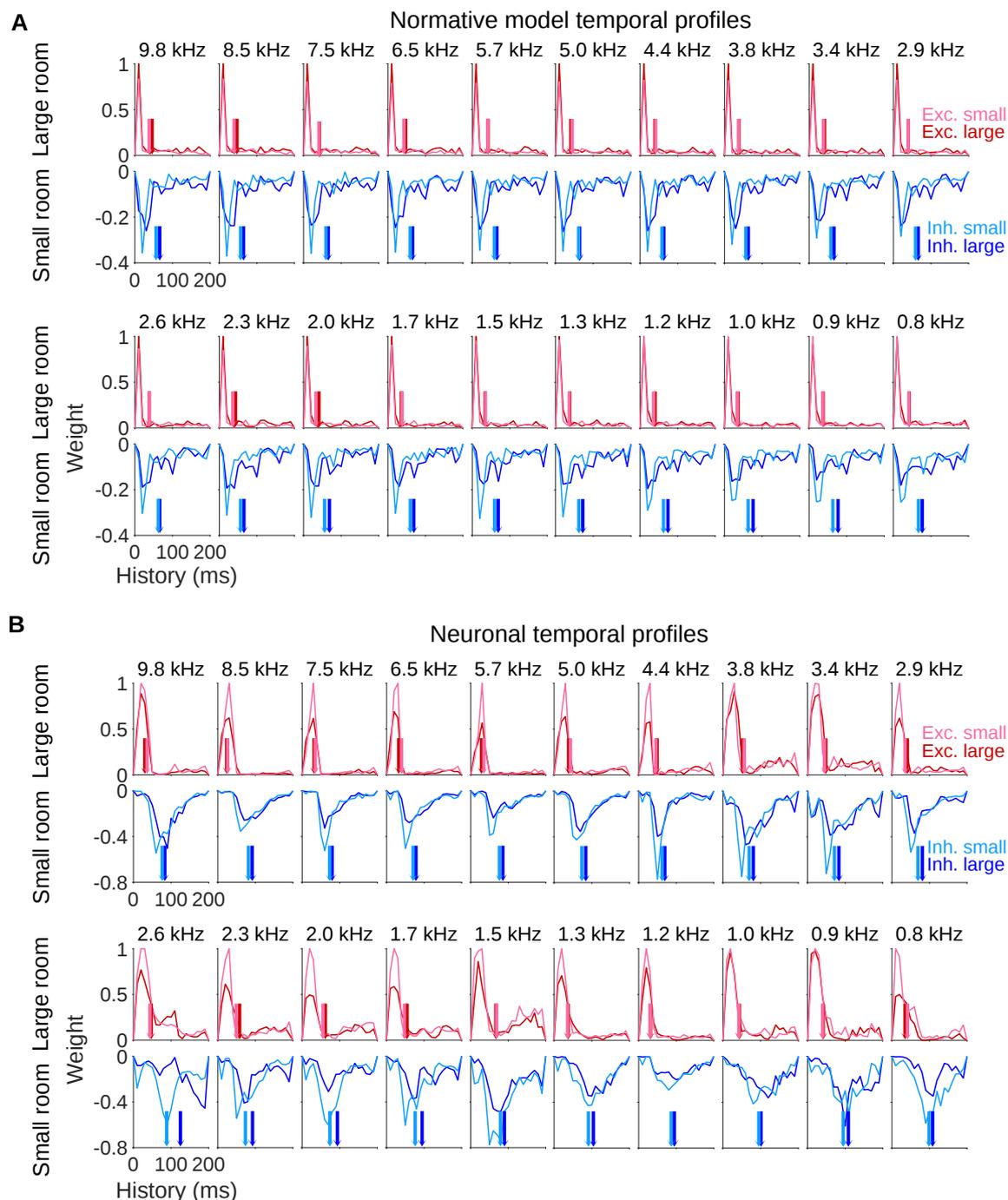
**C**, **D** Analogous to **A** and **B** but for example neuronal STRFs.

**E**, **F** Mean excitatory and inhibitory temporal profiles for the normative model kernels and the neuronal STRFs. The solid lines represent the average temporal profile across all normative kernels (**E**) or neuronal STRFs (**F**) and the shaded area shows  $\pm SEM$ , colour scheme same as in **B** and **D**.



**Fig. 4.3** Frequency breakdown of the normative model kernels and neuronal STRFs.

**A**, Normative model kernels arranged by the anechoic frequency they were trained to predict (BF). The rows immediately below the frequency labels correspond to the kernels obtained from the large room and those below them from the small room. Frequency is on the vertical axis and history on the horizontal. **B**, Neuronal STRFs arranged by their BF. The STRFs of all the neurons with the same BF were averaged together to produce these plots. The arrangement is the same as in A.



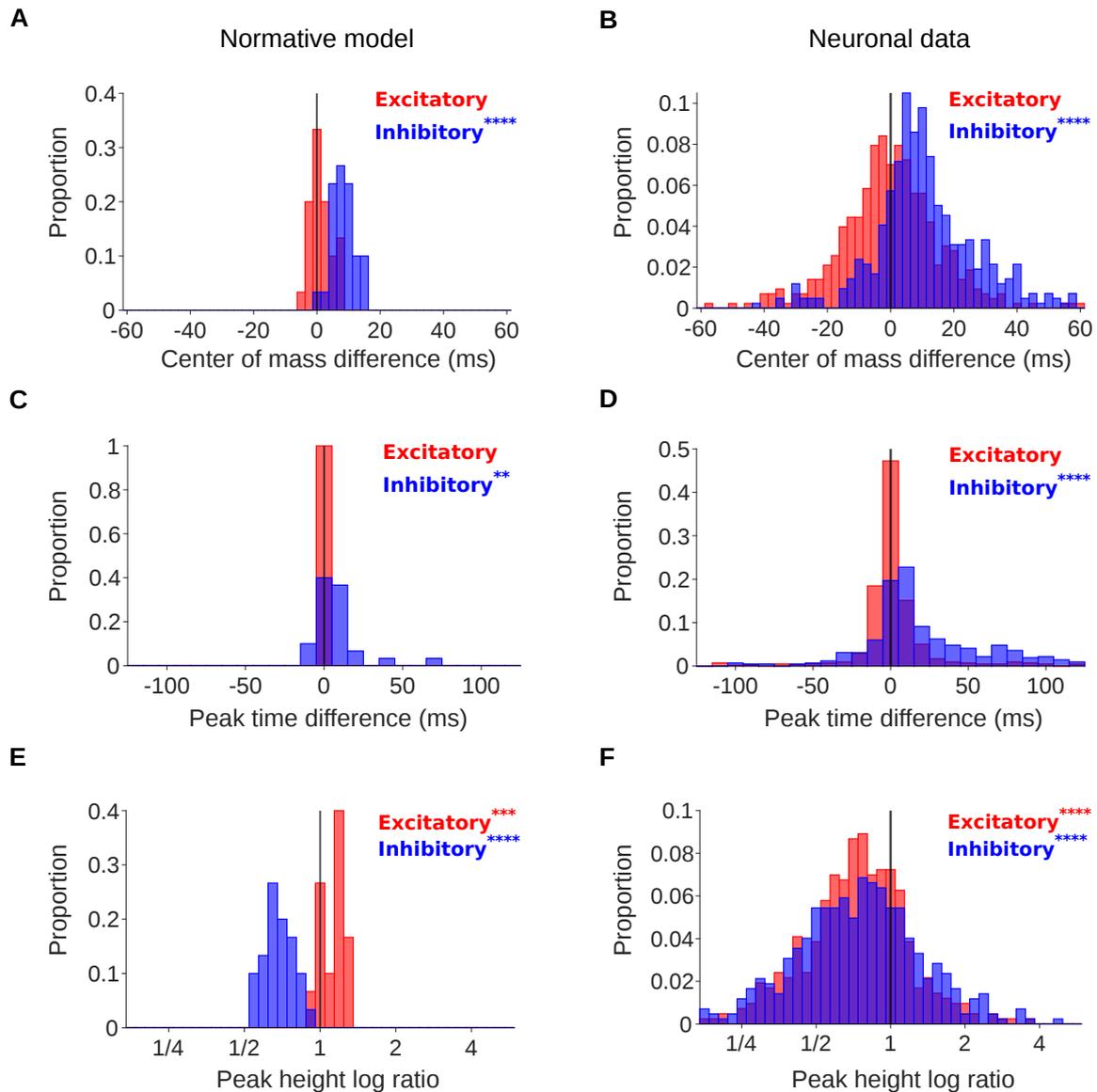
**Fig. 4.4** Frequency breakdown of the normative model and neuronal temporal profiles.

**A**, Temporal profiles of the normative model kernels. For each kernel from Figure 4.3A the excitatory and inhibitory components were averaged separately across frequency. Colour code: small room excitation, light red; large room excitation, dark red; small room inhibition, light blue; large room inhibition, dark blue similar to Figure 4.2B. The center of mass (*COM*) for each condition for the excitation and the inhibition is indicated by the arrows following the same colour scheme. **B**, Temporal profiles of the neuronal STRFs. The excitatory and inhibitory temporal profiles averaged across frequencies are plotted for each kernel from Figure 4.3B, along with their *COM* indicated by the arrows. Colour code is the same as in **A**.

spread out in the larger, more echoic room versus the smaller less echoic one. This is clearly visible in Figure 4.2B, which shows the frequency-averaged positive and negative temporal parts of the model kernels. Strikingly, we see the same pattern in the neuronal STRFs in Figure 4.2D, consistent with the model’s prediction. These temporal effects become very apparent in the average temporal profile across all model kernels (n=30, one for each frequency in the cochleagram, Figure 4.2E) and across all recorded neurons (n=430, Figure 4.2F). This observation also holds across different frequency channels in the model and the data. (Figure 4.3, Figure 4.4).

Intrigued by these qualitatively different responses in the two reverberant conditions in both model and neurons, we sought to quantify them using several different metrics. The three metrics we chose (see subsection 4.2.9) were: Center of mass (*COM*); Peak time (*PT*); Peak height (*PH*). The *COM* measures the relative temporal delay of the bulk of the excitatory ( $COM^+$ ) and inhibitory ( $COM^-$ ) parts of the kernels/STRFs (see equations 4.11, 4.12). The *PT* is the time at which the maximal excitation ( $PT^+$ ) or inhibition ( $PT^-$ ) occur, respectively (see equations 4.13, 4.14). The *PH* is the maximal value of the excitatory ( $PH^+$ ) and inhibitory ( $PH^-$ ) temporal components of the kernels/STRFs (see equations 4.15, 4.16).

For the *COM*, in both model kernels and neuronal STRFs there was little difference in the average excitatory temporal profile between “small” and “large” reverberant conditions. In contrast, there was a clear temporal extension/spread in the inhibitory part between the “small” and “large” reverb conditions in both model kernels (Figure 4.2B) and neuronal STRFs (Figure 4.2D). To quantify the difference between the two reverberant conditions across all model kernels and neurons, we performed a Wilcoxon signed-rank test and plotted this difference as histograms (Figure 4.5A, normative model, one model kernel for each predicted frequency for a total of 30; Figure 4.5B, neurons, 430 neurons from 6 different animals). For the model kernels, the inhibitory centers of mass were higher in the larger room (median  $COM^-$ -difference = 7.9ms,  $p = 2 \times 10^{-6}$ ), while the center of mass of the excitatory components did not differ significantly between the two rooms (median  $COM^+$ -difference = 1ms,  $p = 0.07$ ). The neuronal STRFs followed a similar pattern, the inhibitory centers of mass were higher in the larger room (median  $COM^-$ -difference = 9.5ms,  $p = 1.79 \times 10^{-37}$ ), while the center of mass of the excitatory components did not differ between rooms (median  $COM^+$ -difference = -0.5ms,  $p = 0.8$ ). In general, there was more spread in the *COM*, *PT* and the *PH* in the neuronal data comparing to the dereverberation model.



**Fig. 4.5 Quantification of the temporal profiles of the normative model kernels and neuronal STRFs.**

**A**, Histograms of the difference in center of mass (inhibitory,  $COM^-$ , blue; excitatory,  $COM^+$ , red) between the two different room conditions, large - small. The  $COM^-$  increased in the larger room with a median difference = 7.9ms;  $COM^+$  did not significantly differ, median difference = 1ms. **B**, Same as **A** but for the neuronal data. The  $COM^-$  increased in the larger room, median difference = 9.5ms;  $COM^+$  did not significantly differ, median difference = -0.5ms. **C**, Histograms of the difference in peak time (inhibitory,  $PT^-$ , blue; excitatory,  $PT^+$ , red). The  $PT^-$  were longer in the larger room, median difference = 10ms, whereas  $PT^+$  did not significantly differ, median difference = 0ms. **D**, Neuronal data, the  $PT^-$  were longer in the larger room, median difference = 10ms, while  $PT^+$  did not significantly differ, median difference = 0ms. **E**, Histograms of peak height  $\log_2$  ratio (inhibitory,  $PH^-$ , blue; excitatory,  $PH^+$ , red). The  $PH^-$  were smaller in the larger room, median  $\log_2$  ratio = -0.56, while  $PH^+$  was bigger, median  $\log_2$  ratio = 0.20. **F**, Neuronal data, The  $PH^-$  were smaller in the larger room, median  $\log_2$  ratio = -0.42, as was  $PH^+$ , median  $\log_2$  ratio = -0.45. Significance of Wilcoxon signed-rank tests: \*\*\*\* $p < 0.0001$ , \*\*\* $p < 0.001$ , \*\* $p < 0.01$ .

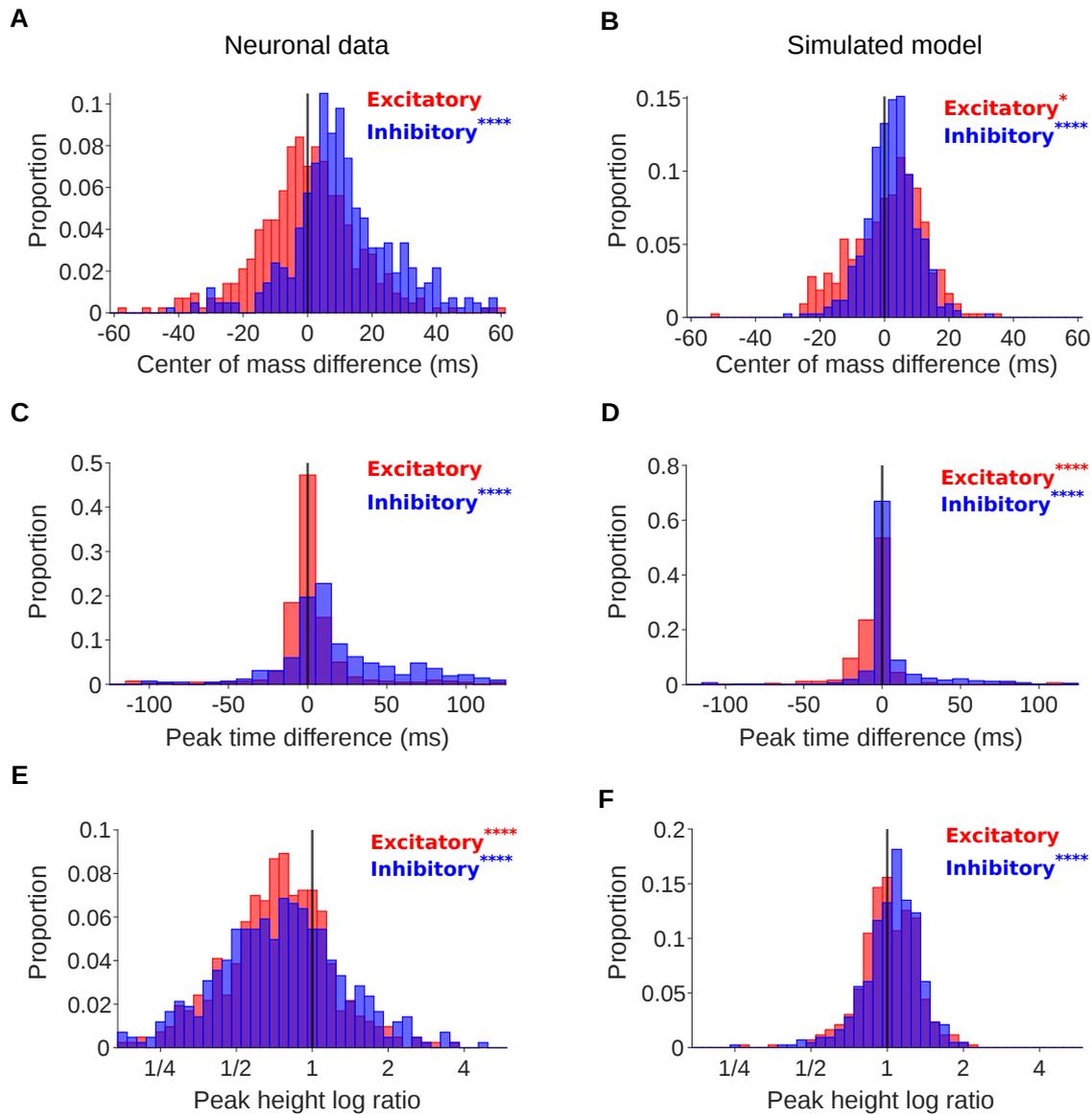
Another measure of the temporal spread of the excitatory and inhibitory STRF/kernel components is the  $PT$ . The excitatory peak time ( $PT^+$ ) was largely unchanged between the small and large conditions for both model and neurons. In contrast, the inhibitory component had a delayed  $PT$  in the large room vs the small room (Figure 4.2B,D,E,F and Figure 4.4A,B). The histograms in Figure 4.5C,D show this clearly for model and neurons respectively as do the Wilcoxon signed-rank tests. For the model kernels, the  $PT^-$  were longer in the larger room (median  $PT^-$  difference = 10ms,  $p=3.7\times 10^{-3}$ ), whereas  $PT^+$  did not differ (median  $PT^+$  difference = 0ms,  $p=1$ ). The neuronal data had a similar trend,  $PT^-$  were longer in the larger room (median  $PT^-$  difference = 10ms,  $p=4.6\times 10^{-24}$ ), whereas  $PT^+$  did not differ (median  $PT^+$  difference = 0ms,  $p=0.1$ ).

The peak height ( $PH$ ) also changed between the two different conditions (Figure 4.2E,F and Figure 4.4A,B). Figure 4.5E,F and the Wilcoxon signed-rank tests summarize this. The inhibitory peak height showed the same trend between model and neurons, decreasing overall in the large room vs the small room (Model: median  $\log_2$  ratio = -0.56,  $p=1.7\times 10^{-6}$ ; Neuronal data: median  $\log_2$  ratio = -0.42,  $p=6.7\times 10^{-25}$ ). The excitatory peak height for the model changed in the opposite direction (median  $\log_2$  ratio = 0.2,  $p=1.2\times 10^{-4}$ ), while it decreased for the neurons (median  $\log_2$  ratio = -0.45,  $p=3.9\times 10^{-38}$ ).

#### 4.3.4 The observed changes are consistent with a truly adaptive process

Next, we wanted to test whether the temporal changes in the inhibitory profile predicted by the normative model and observed in the neuronal data are truly adaptive. In principle, there could be several reasons why the temporal profile of the inhibitory field is broader in the more reverberant room. For example, the cochleagrams of more reverberant sounds are more temporally blurred. This could perhaps cause the “slower” neuronal STRFs, purely due to systematic model fitting artefacts, even if the underlying neuronal responses themselves are not adaptive. Another possibility is that the changes could be due to a static non-linearity that the STRF estimates are not able to capture.

To address these possible confounds, we set out to build a simulated model neuron with a static STRF and a static non-linearity (see subsection 4.2.10). We fitted a single STRF using the small and large room reverberation stimuli together rather than



**Fig. 4.6 Simulated model neurons confirm the adaptive nature of the real neuronal responses to reverberation.**

**A**, Histograms of the difference in center of mass for the real neuronal data (inhibitory,  $COM^-$ , blue; excitatory,  $COM^+$ , red) between the two different room conditions, large - small. The  $COM^-$  were higher in the larger room, median difference = 9.5ms,  $COM^+$  did not significantly differ, median difference = -0.5ms. **B**, The same histogram as in **A** but for the simulation model. The  $COM^-$  were slightly higher in the larger room, median difference = 1.90ms, and the  $COM^+$  were slightly elevated too, median difference = 2.13ms. **C**, Histograms of the difference in peak time for the real neuronal data (inhibitory,  $PT^-$ , blue; excitatory,  $PT^+$ , red). The  $PT^-$  were bigger in the larger room, median difference = 10ms,  $PT^+$  did not significantly differ, median difference = 0ms. **D**, Simulated model, the  $PT^-$  and  $PT^+$  were marginally higher in the larger room but centered at 0,  $PT^-$  median difference = 0ms,  $PT^+$  median difference = 0ms. **E**, Histograms of peak height  $\log_2$  ratio (inhibitory,  $PH^-$ , blue; excitatory,  $PH^+$ , red) for the neuronal data. The  $PH^-$  were smaller in the larger room, median  $\log_2$  ratio = -0.42, as was  $PH^+$ , median  $\log_2$  ratio = -0.45. **F**, Simulated model, the  $PH^-$  were slightly bigger in the larger room, median  $\log_2$  ratio = 0.096, as was  $PH^+$ , median  $\log_2$  ratio = 0.022. Significance of Wilcoxon signed-rank tests: \*\*\*\* $p < 0.0001$ , \*\*\* $p < 0.001$ , \* $p < 0.05$ .

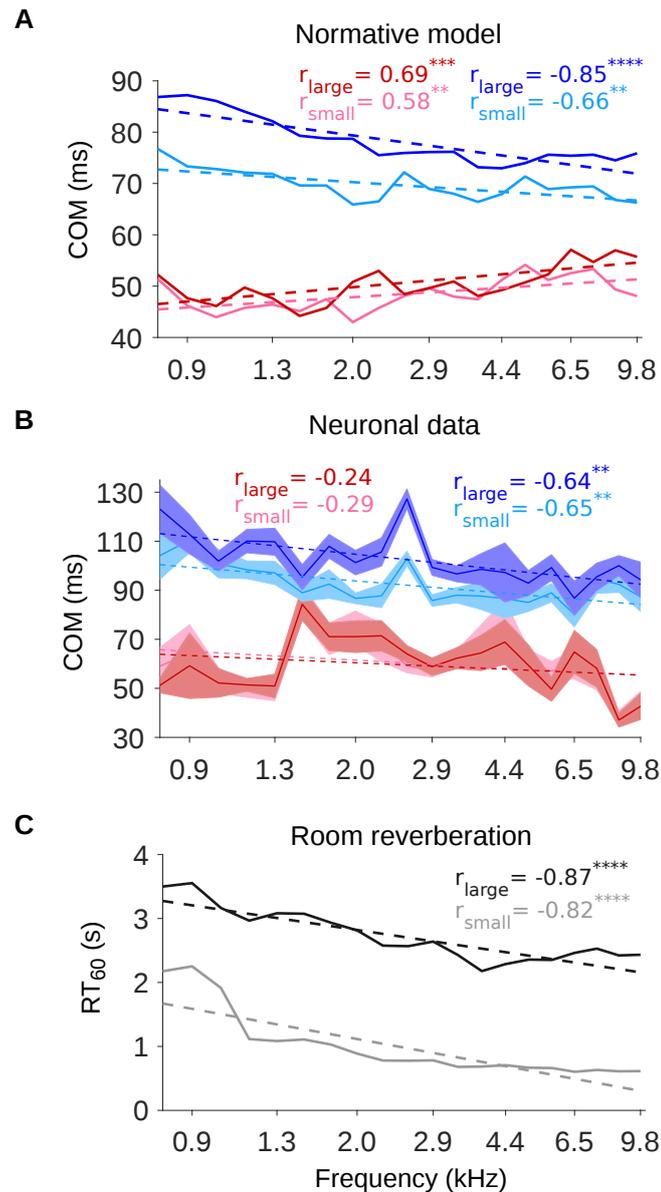
fitting them separately as we did before. We then used this single STRF to generate a predicted linear output response for each neuron to the small and large reverberation cochleagrams. Finally, we fitted a single sigmoid non-linearity for each neuron, again using all the data together. To generate realistic simulated responses, we added Poisson noise to the output of the linear-nonlinear model and created ten “virtual trials”, the same number as for the ferret experiments. These were then averaged to generate the final simulated neural response. We then used the exact same fitting procedure as we did to fit the ferret cortical STRFs. After fitting STRFs to the simulated static responses, we extracted the same three features as we previously did for the real neuronal STRFs: Center of mass (*COM*); Peak time (*PT*); Peak height (*PH*). The logic is that if we simulate a purely static neuron, we should not be able to recover the differences in STRF properties to the two conditions as previously seen.

We observed the following results, summarised in Figure 4.6 and performed Wilcoxon signed-rank tests as before. The  $COM^-$  values were slightly higher in the large room compared to the small room, but the effect size was 5 times smaller compared to the real neuronal data (simulated model median  $COM^-$  difference = 1.90ms,  $p=1.04 \times 10^{-9}$  vs real neuronal data median  $COM^-$  difference = 9.5ms,  $p=1.79 \times 10^{-37}$ ). The  $COM^+$  were slightly elevated too in the large room compared to the lack of statistical significance in the neuronal data (simulated model median  $COM^+$  difference = 2.13ms,  $p=0.023$  vs real neuronal data median  $COM^+$  difference = -0.5ms,  $p=0.8$ ). The changes in  $PT^-$  were statistically significant in the larger room but the median was centered at zero (simulated model median  $PT^-$  difference = 0ms,  $p=5.06 \times 10^{-9}$  vs real neuronal data median = 10ms,  $p=4.57 \times 10^{-24}$ ). The changes in  $PT^+$  were statistically significant but again with a zero median (simulated model median  $PT^+$  difference = 0ms,  $p=4.78 \times 10^{-14}$  vs real neuronal data median  $PT^+$  difference = 0ms,  $p=0.1$ ). The  $PH^-$  values were slightly bigger in the larger room whereas they were decreased in the neuronal data (simulated model median  $\log_2$  ratio = 0.096,  $p=2.79 \times 10^{-6}$  vs real neuronal data median  $\log_2$  ratio = -0.42,  $p=6.66 \times 10^{-25}$ ). The  $PH^+$  were not significantly different between large and small room, whereas they decreased in the large room for the real neurons (simulated model median  $\log_2$  ratio = 0.022,  $p=0.37$  vs real neuronal data median  $\log_2$  ratio = -0.45,  $p=3.88 \times 10^{-38}$ ). Overall, these results show that even though there is a slight bias, our main findings are likely due to a truly adaptive phenomenon.

### 4.3.5 Frequency dependence of the temporal profile of adaptation

In all our analyses so far, we have treated the model kernels and neuronal STRFs together without separating them by BF. However, there were some clear changes that we observed in the model kernels that happened as a function of frequency. If we look at Figure 4.3A, we can see that while the temporal spread of the excitatory part of the kernels stays relatively constant, the inhibitory part tends to extend further back in time as the frequency decreases from high to low. This temporal broadening of the inhibitory fields is seen for the model kernels derived from both the large and the small room. We can observe this even more clearly in the frequency averaged temporal profiles in Figure 4.4A. Following these interesting observations in our model, we also investigated the temporal spread with frequency in our neuronal data. When we examined the neuronal STRFs arranged by frequency, we observed a similar temporal spread of the inhibitory part of the STRFs with decreasing frequency, while the excitatory part stayed relatively constant as predicted by our model (Figure 4.3B). This is also evident in the frequency averaged temporal profiles of the STRFs (Figure 4.4B).

In order to quantify this relationship, we decided to plot the excitatory and inhibitory center of mass values ( $COM^+$ ,  $COM^-$ ) as a function of the BF of the model kernels and neuronal STRFs. This relationship is depicted in Figure 4.7A for the normative model and Figure 4.7B for the neuronal data. For both the normative model and the neuronal data there was a clear decrease in the inhibitory ( $COM^-$ ) values for both the large and the small room with increasing BF values. The excitatory ( $COM^+$ ) values did not show the same trends in the normative model and the neuronal data. They increased with increasing BF for the normative model and decreased or stayed flat with increasing BF for the neuronal data. Given that these relationships looked relatively linear, we decided to fit linear regression lines to quantify them. There was a significant decrease of  $COM^-$  with increasing frequency for the normative model in both conditions ( $COM_{large}^-$ , Pearson's correlation:  $r = -0.85$ ,  $p = 1.9 \times 10^{-6}$ ;  $COM_{small}^-$ ,  $r = -0.66$ ,  $p = 0.0016$ ; Figure 4.7A) and the same trend was observed for the neuronal data ( $COM_{large}^-$ ,  $r = -0.64$ ,  $p = 0.0023$ ;  $COM_{small}^-$ ,  $r = -0.65$ ,  $p = 0.0018$ ; Figure 4.7B). In contrast, there was a significant increase of  $COM^+$  with BF for the normative model in both conditions ( $COM_{large}^+$ ,  $r = 0.69$ ,  $p = 7.9 \times 10^{-4}$ ;  $COM_{small}^+$ ,  $r = 0.58$ ,  $p = 0.007$ ; Figure 4.7A) while there was no significant change for the neuronal data ( $COM_{large}^-$ ,  $r = -0.24$ ,  $p = 0.31$ ;  $COM_{small}^-$ ,  $r = -0.29$ ,  $p = 0.22$ ; Figure 4.7B). It is also



**Fig. 4.7 The excitatory and inhibitory COM profiles are frequency dependent and scale with the amount of reverberation.**

**A**, Center of mass values (COM, vertical axis) are plotted against the best frequency (BF, horizontal axis) for the excitatory and inhibitory components of each normative model kernel for the two different rooms. These are colour coded as follows: excitatory COM ( $COM_{large}^+$ , dark red;  $COM_{small}^+$ , light red) and their inhibitory counterparts ( $COM_{large}^-$ , dark blue;  $COM_{small}^-$ , light blue). The dashed lines show a linear regression fit for each room, and the Pearson's  $r$  value for each fit is given in the top right corner of the plot. **B**, COM values (vertical axis) are plotted against the BF (horizontal axis) for the neuronal data as in **A**. Each neuron was assigned a BF and the COM values measured. The solid lines represent the mean COM value for each BF, the shaded areas  $\pm$ SEM, colour scheme as in **A**. **C**,  $RT_{60}$  values are plotted as a function of cochlear frequency bands, for the large (black line) and small (grey line) rooms. Linear regression fit (dotted line) was used as in **A** and **B** to calculate  $r$ . Significance of Pearson's correlation:  $****p < 0.0001$ ,  $***p < 0.001$ ,  $**p < 0.01$ .

noteworthy that the large room inhibitory values ( $COM_{large}^-$ ) were higher than their small room counterparts ( $COM_{small}^-$ ) across BFs for both normative model and neuronal data, whereas the excitatory values ( $COM_{large}^+$ ,  $COM_{small}^+$ ) were mostly overlapping (Figure 4.7A,B). This is in agreement with our previous observations when analyzing all the data together regardless of BF (Figure 4.5A,B)

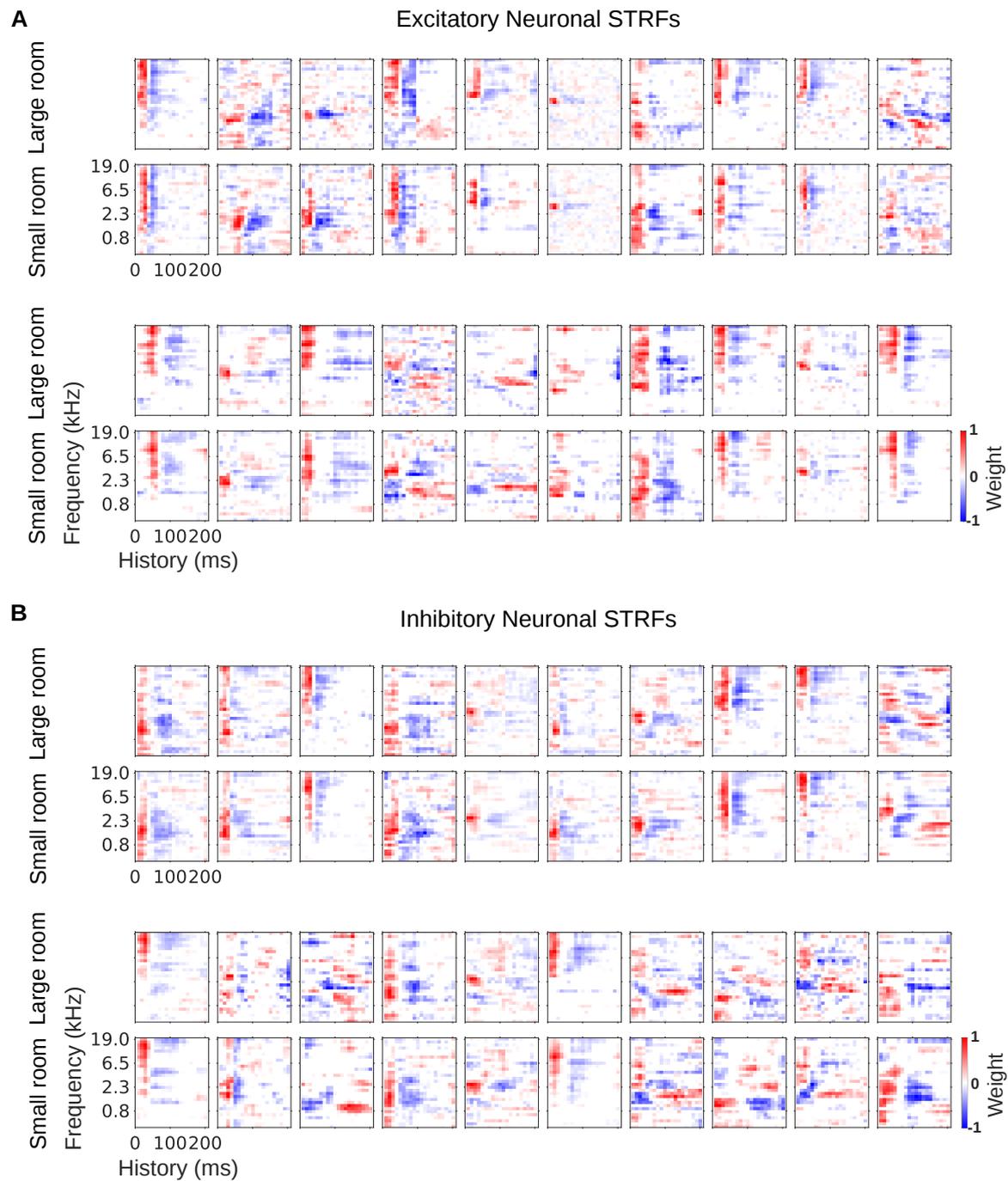
Given the frequency dependence of the  $COM$  measures, we sought to investigate whether this could be explained by the acoustical properties of the two different rooms. A previous study by [Traer and McDermott \(2016\)](#) found that the IR of natural and man-made environments exhibit strong regularities that can be exploited by human listeners to separate sound source and environment. Notably, there was a clear relationship between the decay rate of the power in the IR as measured by the  $RT_{60}$  and the particular frequency band. At the low end of the frequency spectrum, typically below the range we examine, the  $RT_{60}$  values tended to increase with frequency. Above a certain frequency (typically at the bottom end of the frequencies we examined), the  $RT_{60}$  values dropped as the frequency became higher. Therefore, we decided to measure the  $RT_{60}$  values in our two rooms, large and small. We wanted to see if they would follow a similar trend, and whether this relates to the  $COM$  values of the model kernels/neuronal STRFs. The impulse response of each room was converted to a cochleagram following the same procedure that was used for the generation of the model kernels and the STRFs (see subsection 4.2.6). We then measured the decay profile in each cochlear frequency band and calculated  $RT_{60}$  values for the IR of each room. These are plotted in Figure 4.7C. Both the large and small room exhibited a decrease in the  $RT_{60}$  values as a function of frequency. This was captured by the highly significant negative correlation coefficients (Large room, Pearson's correlation:  $r = -0.87$ ,  $p = 6.0 \times 10^{-7}$ ; small room,  $r = -0.82$ ,  $p = 8.5 \times 10^{-6}$ ; Figure 4.7C). Overall, the inhibitory  $COM^-$  seemed to track the  $RT_{60}$  frequency profile much better than the excitatory  $COM^+$  for both normative model and neurons.

### 4.3.6 Exploring the role of putative excitatory and inhibitory neurons in the adaptation to reverberation

Following our normative approach to the reverberation problem, we wanted to explore potential mechanisms for implementing the neuronal adaptation that we see. In principle, several biological mechanisms could be in play. These include synaptic

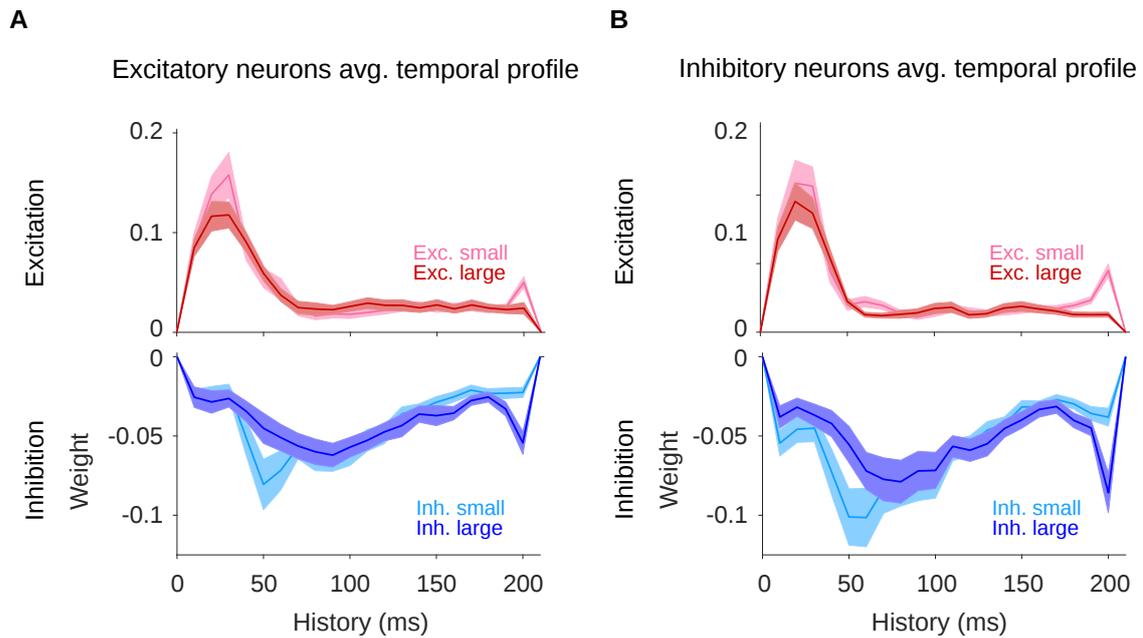
depression (David et al., 2009), intrinsic dynamics of membrane  $K^+$  channels (Abolafia et al., 2011), hyperpolarizing inputs from inhibitory neurons (Li et al., 2015; Natan et al., 2015; Gwak and Kwag, 2020) or adaptation inherited from subcortical regions such as the IC or MGB (Dean et al., 2008; Devore et al., 2009; Lohse et al., 2020a) or even from higher cortical areas. The main adaptive effect that we observe in the model kernels and neuronal STRFs is that the inhibitory component in the larger, more reverberant room, is more delayed and temporally spread out as shown by the COM and PT measures (Figure 4.2E,F; Figure 4.5B,D). Hence we may be likely looking at a phenomenon involving inhibitory neurons. Therefore, we decided to explore separately the STRFs of putative excitatory and inhibitory neurons and the potential interaction between them.

In order to examine the STRFs of the putative excitatory and inhibitory neurons and to see if there are any differences in their adaptation profiles, we used the spike waveform clustering approach described in the previous chapter (see subsection 3.3.1 and Figures 3.4 and 3.6). Our waveform clustering analysis used only single units and as before we considered only units that had a noise power to signal power (NPSP)  $< 40$ . There were a total of 118 such single units with reliable responses as defined by the NPSP measure. Out of them, 62 were classified as putative excitatory and 56 as putative inhibitory. Twenty example STRFs are plotted for both classes in Figure 4.8 (panel A excitatory, panel B inhibitory). As before (Figure 4.2C; Figure 4.3B), we fitted STRFs for the small and large conditions separately and these are plotted respectively in the bottom and top rows of each subpanel in Figure 4.8. At first sight, there were no obvious differences between the two classes. STRFs from both putative excitatory and inhibitory neurons included units with narrow and broad frequency tuning, different level of complexity, some being more independent (separable) in frequency and history, whilst others exhibited diagonal and other spectro-temporal patterns. Amongst both putative excitatory and inhibitory neurons, it was quite common to see the pattern of initial fast excitatory field followed by a more diffuse inhibitory field in the STRF. Out of the neurons that exhibited this typical pattern, many showed temporal broadening of their inhibitory fields in the large, more reverberant condition. This was the case for both putative excitatory and inhibitory neurons. Given these observations, we suspected that if there are differences between these two classes, they would be largely temporal.



**Fig. 4.8 Excitatory and inhibitory neuronal STRFs.**

**A**, Example putative excitatory neuronal STRFs identified by waveform clustering as described in subsection 3.3.1 (see also Figures 3.4 and 3.6). The top rows correspond to the STRFs obtained from the large room and those below them from the small room. Frequency is on the vertical axis and history on the horizontal. **B**, Example putative inhibitory neuronal STRFs. The arrangement is the same as in A.



**Fig. 4.9 Excitatory and inhibitory neuronal temporal profiles.**

**A**, Mean excitatory (top panel) and inhibitory (bottom panel) temporal profiles of the STRFs of putative excitatory neurons. The solid lines represent the average temporal profile across all putative excitatory neuronal STRFs ( $n=62$ ) and the shaded area shows  $\pm$ SEM. The colour scheme is the same as in earlier figures: the excitatory component for the small room in light red, for the large room in dark red; the inhibitory component for the small room in light blue, for the large room in dark blue. **B**, Mean excitatory (top panel) and inhibitory (bottom panel) temporal profiles of the STRFs of putative inhibitory neurons ( $n=56$ ). The colour scheme is the the same as in A. The putative excitatory and inhibitory neurons were identified using the clustering analysis described in Figures 3.4 and 3.6.

In order to explore the temporal differences between putative excitatory and inhibitory neurons, we looked at the averaged STRF excitatory and inhibitory field temporal profiles for both neuronal classes as shown in Figure 4.9. For both putative excitatory neuronal STRFs (Figure 4.9A, top row) and putative inhibitory ones (Figure 4.9B, top row) the excitatory field had a quick onset and short duration. In contrast, the inhibitory temporal profile had a longer latency and was more spread out in time for both putative excitatory (Figure 4.9A, bottom row) and inhibitory neurons (Figure 4.9B, bottom row). For both neuronal classes, the inhibitory temporal profile in the larger room was more extended in time compared to the small room condition. We measured the same three features of the temporal profile as we did before: Center of mass (*COM*); Peak time (*PT*); Peak height (*PH*).

To quantify the difference between the two reverberant conditions in the two putative neuronal classes, we performed Wilcoxon signed-rank tests (62 putative excitatory neurons and 56 putative inhibitory from 6 different animals). For both putative excitatory and inhibitory neurons, the inhibitory COM was higher in the larger room (median  $COM^-$  difference = 9.6ms,  $p = 3.1 \times 10^{-8}$ , excitatory neurons; median  $COM^-$  difference = 9.6ms,  $p = 2.7 \times 10^{-5}$ , inhibitory neurons), while the COM of the excitation was smaller in the larger room but only significant for the excitatory neurons (median  $COM^+$  difference = -3.4ms,  $p = 0.0076$ , excitatory neurons; median  $COM^+$  difference = -1.1ms,  $p = 0.38$ , inhibitory neurons). The excitatory peak time ( $PT^+$ ) was largely unchanged between the small and large conditions for both classes. In contrast, the inhibitory component had a delayed  $PT$  in the large room vs the small room. For both putative excitatory and inhibitory neurons, the  $PT^-$  were longer in the larger room (median  $PT^-$  difference = 10ms,  $p=6.0 \times 10^{-6}$ , excitatory neurons; median  $PT^-$  difference = 10ms,  $p=0.0034$ , inhibitory neurons), whereas  $PT^+$  did not differ ( $PT^+$  difference = 0ms,  $p=0.84$ , excitatory neurons; median  $PT^+$  difference = 0ms,  $p=0.086$ , inhibitory neurons). The peak height ( $PH$ ) also changed between the two different conditions and showed the same trend between putative excitatory and inhibitory neurons. The  $PH^-$  decreased overall in the large room vs the small room (median  $\log_2$  ratio = -0.32,  $p=0.002$ , excitatory neurons; median  $\log_2$  ratio = -0.17,  $p=0.021$ , inhibitory neurons). The  $PH^+$  for both classes changed in the same direction (median  $\log_2$  ratio = -0.32,  $p=4.4 \times 10^{-4}$ , excitatory neurons; median  $\log_2$  ratio = -0.34,  $p=0.0023$ , inhibitory neurons).

To more directly compare the two putative classes we performed Mann–Whitney U (Wilcoxon rank-sum) tests for the three measures in the small and large reverberant conditions. None of the tests revealed significant differences between the two classes, excitatory vs inhibitory:  $COM^-_{small}$ ,  $p=0.43$ ;  $COM^-_{large}$ ,  $p=0.55$ ;  $COM^+_{small}$ ,  $p=0.53$ ;  $COM^+_{large}$ ,  $p=0.46$ ;  $PT^-_{small}$ ,  $p=0.3$ ;  $PT^-_{large}$ ,  $p=0.18$ ;  $PT^+_{small}$ ,  $p=0.22$ ;  $PT^+_{large}$ ,  $p=0.75$ ;  $PH^-_{small}$ ,  $p=0.2$ ;  $PH^-_{large}$ ,  $p=0.5$ ;  $PH^+_{small}$ ,  $p=0.17$ ;  $PH^+_{large}$ ,  $p=0.17$ . Thus, these results suggest that the putative EXC and INH show similar adaptation to reverberation.

As we mentioned earlier in Chapter 3, subsection 3.3.1, an important caveat to our analysis is that the electrophysiological separation of the neuronal types is far from definitive. Ground truth recordings using optotagging of the three main classes of INH (PV, SOM, VIP) has shown that the fast-spiking PV (basket and chandelier cells) separate much more clearly compared with SOM and VIP (Yu et al., 2019). Thus, our

EXC cluster is likely to contain not only pyramidal neurons but also some proportion of SOM, VIP and non-VIP. Given the lack of difference in the temporal profiles that we observe in Figure 4.9, it is likely that PV do not contribute substantially to the adaptation to reverberation that we observe. This “negative” finding suggests that future efforts can focus on exploring the other INH types (SOM, VIP and non-VIP).

## 4.4 Discussion

In this chapter, we applied a normative modelling approach to ask the question: If the auditory system is optimised to remove reverberation from natural sounds, what linear filtering adaptive properties would be best suited to achieve this goal? To answer this question we used a rich dataset of anechoic speech and environmental natural sounds, adding different amounts of reverberation to them. We then optimised a linear model to remove this reverberation. We constructed our model in a way such that the model “neurons” (kernels) after optimisation can be compared to the filtering properties of real auditory cortical neurons in the ferret, specifically their STRFs (Figure 4.1). We confirmed the validity of our normative model by showing that it recapitulated known properties of auditory cortical neurons, such as frequency tuning and temporally asymmetric STRFs with excitation followed by inhibition (Figure 4.2). Interestingly, our model also makes two novel predictions: (1) the inhibitory part of the STRF shifts in time with the amount of reverberation so as to cancel the excess echoes in more reverberant conditions (Figure 4.5); (2) this inhibitory temporal scaling is dependent on frequency (Figures 4.3, 4.4, 4.7). We verified the predictions of the model using electrophysiological recordings from ferret auditory cortical neurons, fitting STRFs and comparing them qualitatively and quantitatively to the model kernels. In addition, using fitted simulated neurons, we showed that the observed changes in the neuronal STRFs are consistent with a truly adaptive dynamic process rather than a static non-linearity or stimulus dependent model-fitting artefact (Figure 4.6). We also explored putative mechanisms allowing for the implementation of the adaptation by comparing the properties of putative excitatory and inhibitory neurons (Figures 4.8, 4.9). Thus, our data show that most primary auditory cortical neurons seem to adapt to reverberation by adjusting their filtering properties in order to minimize its negative impact. In the following, we explore these findings in the broader context of previous studies and proposed mechanisms for adaptation to reverberation.

#### 4.4.1 Auditory cortical neurons adapt their responses to reverberation

Previous studies have shown that humans are remarkably robust to reverberation when discriminating speech and naturalistic sounds (Houtgast and Steeneken, 1985; Bradley, 1986; Darwin and Hukin, 2000; Culling et al., 2003; Nielsen and Dau, 2010). Our neurophysiological results in the ferret AC are consistent with such robust representation. We find that the majority of neurons recorded in primary AC tend to adapt their responses in a way which is consistent with the computational goal of removing reverberation from natural sounds (Figures 4.2, 4.5), even in anaesthetised animals. Our results are also in good agreement with a previous study in awake passive listening ferrets, showing that anechoic speech and vocalizations were more readily decodable from the responses of auditory cortical neurons than their echoic counterparts (Mesgarani et al., 2014). A similar study in humans using EEG corroborated these findings, showing better decoding accuracy of anechoic speech envelope compared to distorted reverberant speech, but only when listeners attended to the sound sources (Fuglsang et al., 2017).

Interestingly, a human MEG study found that both anechoic and reverberant sounds are represented in AC, as both of them were readily decodable (Puvvada et al., 2017). In addition, Traer and McDermott (2016) found that humans were able to discriminate different reverberant conditions quite well with both familiar and unfamiliar sounds. In line with this, we did find a small number of neurons which did not seem to adapt their responses or followed the opposite direction than expected, i.e. their  $COM^-$  and  $PT^-$  decreased in the large room comparing to the small room (Figure 4.5B,D,F). Thus, although most cortical neurons seem to be reverberation invariant, it is possible that some of them might carry information about the reverberant environment or even represent it more explicitly.

#### 4.4.2 Largely inhibitory phenomena underlie the adaptation to reverberation

The main change that we observe in both model kernels and neuronal STRFs is that the severe reverberation condition (large room) has a more temporally delayed inhibitory profile than the moderate reverberation (small room). In contrast, the excitatory part of the kernels and STRFs does not change in time (Figures 4.2, 4.3, 4.4, 4.5A-D).

In addition, there also seems to be a small multiplicative gain effect, which causes the maximum amount of inhibition to decrease in the large room compared to the small room in both model and neurons (Figure 4.5E,F). In contrast, the excitatory components change in opposite directions, increasing in maximum value from small to large reverberation for the model while decreasing for the neurons.

Our findings are consistent with those of [Mesgarani et al. \(2014\)](#), who put forward a combination of subtractive synaptic depression and multiplicative gain change as a potential mechanism for the observed adaptation in their study. However, they did acknowledge that other functionally equivalent mechanism might be feasible too. Although exploring reverberation in a similar setting to this current study, there were several important differences. A limitation of the [Mesgarani et al. \(2014\)](#) study was the use of an exponentially decaying Gaussian to simulate the effects of reverb. Although modelling well the tail reverberation, this fails to take into account the early reflections and the frequency-dependent decay profile of the IR, which has important perceptual effects and contributes substantially to the perceived realism as judged by human listeners ([Traer and McDermott, 2016](#)). Furthermore, their study did not test different echoic conditions with varying amounts of reverb. Therefore, the time constants of the synaptic depression and gain components in their model, two critical parameters, were fixed. The authors highlighted that these time constants might have an important impact in conditions with different amounts of reverberation. This is indeed one of our main novel findings: more reverberant environments require temporally delayed inhibitory responses within the STRFs, and this is what is seen in primary auditory cortical neurons. Finally, [Mesgarani et al. \(2014\)](#) only used speech and vocalizations to probe adaptation to reverberation, while humans and animals experience a much richer set of sounds, including textures and other environmental sounds that can have very different statistics ([Attias and Schreiner, 1996](#); [Turner, 2010](#)). In the current study, we use a more diverse stimulus set, which includes many environmental sounds, thus showing that the adaptation to reverb we describe is a stimulus-general mechanism.

Our finding that a primarily inhibitory temporal shift is an important adaptive mechanism in AC is also consistent with a recent study by [Rahman et al. \(2019\)](#). The authors found that a neural network model consisting of a small number of hidden units with STRFs of only 25 ms duration followed by an exponential decay fitted the neuronal firing rate over time as well as models with STRFs of 200ms duration. In addition, there was a clear bimodal distribution of the exponential time constants,

with the majority of long time constants belonging to units whose STRF was largely inhibitory, while excitatory STRFs had very short time constants.

### 4.4.3 Adaptation to reverberation is frequency dependent

One interesting observation that we have made is that the changes which we observed both in the normative model kernels and neuronal STRFs are frequency dependent (Figures 4.3, 4.4). Thus, the temporal extension of inhibition but not the excitation temporally, as captured by our center of mass (COM) measure, seemed to track the reverberant profile of the acoustic spaces as measured by the reverberation time ( $RT_{60}$ ) of the small and large rooms (Figure 4.7). As discussed earlier, natural and man-made environments exhibit certain regularities, and frequency-dependent reverberation, as measured by  $RT_{60}$ , is one of them (Traer and McDermott, 2016). Given the similarity of the overall shape of the  $RT_{60}$  vs frequency profile (that is, an inverted parabola) across acoustic spaces, it is likely that the observations we make in this study would generalise to the IRS of other environments. This is something that can be tested in future studies, where IRS with slightly different  $RT_{60}$  frequency profiles are used to see whether neurons will adapt their response according to frequency. Interestingly, time constants of adaptation to the mean sound level in IC seem to follow a similar trend, with neurons tuned to lower frequencies adapting more slowly (Dean et al., 2008). A study by Willmore et al. (2016) found that including an input pre-processing mimicking this frequency-dependent adaptive process improved model predictions of the responses of ferret auditory cortical neurons to various natural and artificial sound stimuli with undefined amount of reverberation. The authors observed that these IC time constants were anti-correlated with the autocorrelation width in the different frequency bands of natural sounds, with higher frequencies in natural sounds exhibiting larger autocorrelation widths but higher frequency tuned neurons showing shorter time constants (Willmore et al., 2016). It is likely that stronger reverberation in a given frequency band will increase the autocorrelation width of that frequency. Hence, these findings may be linked to the effects we see here.

### 4.4.4 Conclusion

In summary, in this chapter we took a normative approach to ask the following question: what would the optimal linear transformation from reverberant to anechoic

sounds look like, and are the predictions of such a normative approach telling us something about how AC achieves adaptation to reverberation? We find that our model recapitulates known properties of auditory cortical neurons, namely frequency tuning and temporally asymmetric STRF profile. Our model also made two novel predictions which we confirmed using electrophysiological recordings: (1) the inhibitory component of neuronal STRFs should scale in time in order to cancel the effects of reverberation, with the scaling being proportional to the amount of reverberation; (2) the inhibitory scaling is frequency dependent and tracks the reverberation profile of the acoustic space. The observed changes are likely due to a dynamic adaptive process and not a static non-linearity. In addition, we find that inhibitory and excitatory neurons seem to adapt in the same way, following a similar temporal profile. Thus, our work extends the findings of previous studies on the neural mechanisms of adaptation to reverberation and suggests avenues for future research. Future studies can focus on exploring the temporal shift in the inhibitory component of the STRF in a number of different acoustic environments in order to confirm the generality of the phenomenon that we observe here. In addition, it would be interesting to examine the biological mechanisms behind the inhibitory effects in the STRFs.



# Chapter 5

## General Discussion

### 5.1 Summary of findings

In this thesis, I explored the organisation and properties of ferret AC, namely its tonotopic local and global maps and the tuning of different classes of inhibitory and excitatory neurons, in addition to the role of AC in the robust representation of complex natural sounds in realistic acoustic environments containing reverberation. I utilised 2-photon calcium imaging and the novel Neuropixels microelectrodes to record the responses of auditory cortical neurons while simultaneously presenting different sound stimuli. The main findings of the thesis are summarised below.

In Chapter 2, I explored the tonotopic organisation of AC at the local and global scale in the mouse and the ferret, and I found that both mice and ferrets have equivalent local heterogeneity in frequency preferences (“salt and pepper”) while retaining a globally tonotopic map. In addition, much of the observed heterogeneity in the best frequency (BF) maps could be explained by the existence of neurons with several stereotypical frequency receptive fields (“single-peaked”, “double-peaked” and “complex”), where the more complex tuning was associated with higher local heterogeneity. Finally, the different proportions of each type of neuron sampled in microelectrode and calcium imaging data pointed to a potential explanation for the long-standing discrepancy in frequency tuning maps observed in the literature using these two methods.

In Chapter 3, I developed electrophysiological and imaging approaches in order to separate and study the properties of putative excitatory and inhibitory neurons. Using microelectrode data of spike waveforms, clustering and correlogram analysis

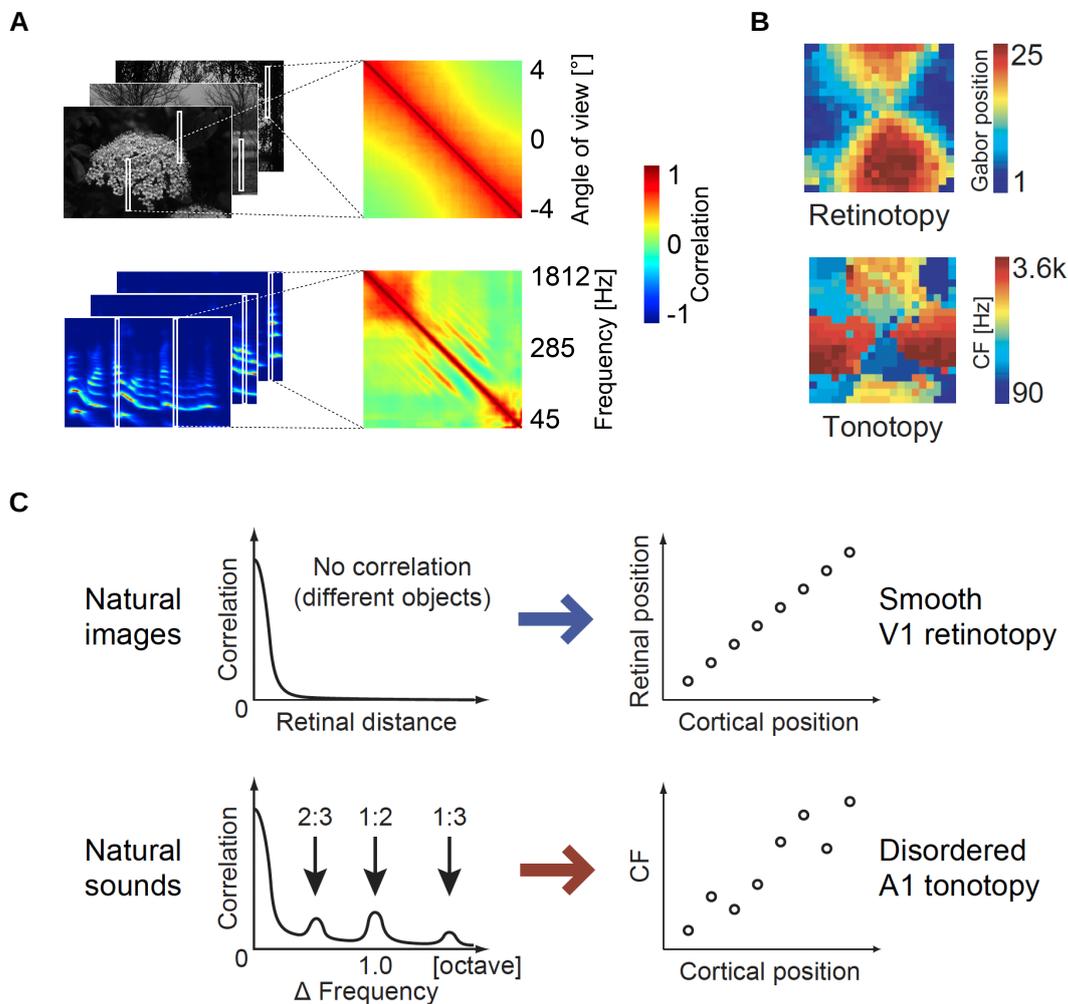
I tested five different waveform features commonly used in the literature and found that the trough-to-peak and peak width features allowed us to discriminate putative excitatory and inhibitory neurons. Using a combination of targeted viral delivery of the calcium indicator GCaMP6f to populations of excitatory and inhibitory neurons, 2-photon calcium imaging, immunohistochemical and registration techniques, I was able to successfully label, record from and identify post hoc excitatory neurons and two of the main classes of inhibitory neurons - PV and SOM. Surprisingly, I did not find significant differences in the proportions of tuned neurons, or their frequency tuning characteristics, such as tuning width, local tonotopic variability or complexity of frequency receptive fields. I also examined their co-tuning using signal correlations and found that the responses within and between classes are very similar. Using noise correlations as an indicator of functional connectivity, I found that inhibitory neuronal subclasses tended to be more coupled to one another and form sub-networks, compared to excitatory neurons. Interestingly, I found the opposite in our electrophysiology data and that these differences could be potentially explained by the sampling of neurons with different complexity of their receptive fields.

Finally, in Chapter 4 I took a normative approach to ask the following question: What would the optimal transformation from reverberant to anechoic sounds look like and are the predictions of this model telling us something about how AC achieves robust representation of natural sounds despite the presence of reverberation? I found that our model recapitulates known properties of auditory cortical neurons, namely frequency tuning and temporally asymmetric receptive field profile. Our model also made two novel predictions which I confirmed using electrophysiological recordings: (1) The inhibitory component of neuronal receptive fields should shift in time in order to cancel the effects of reverberation, with the shifting being proportional to the amount of reverberation; and (2) The inhibitory shifting is frequency dependent. The observed changes are likely due to a dynamic adaptive process and not a static non-linearity. In addition, I found that inhibitory and excitatory neurons seem to adapt to reverberation in the same way, following a similar temporal profile.

## 5.2 Cortical tonotopic maps: disorder without rhyme or reason?

In Chapter 2, I found that both the ferret and mouse AC display the so called “salt and pepper” tonotopic organisation, characterised by local heterogeneity but global order. This is in good agreement with previous studies in the mouse (Bandyopadhyay et al., 2010; Rothschild et al., 2010; Panniello et al., 2018; Romero et al., 2019). Although initial studies suggested that such local disorder is more prevalent in AC layer 2/3 (Guo et al., 2012; Winkowski and Kanold, 2013) than the granular layer 4, a recent tour de force 2-photon calcium imaging study instead found that this organisation is generalised across layers (Tischbirek et al., 2019). In contrast to our findings in ferrets and mice, Zeng et al. (2019) found much smoother tonotopic organisation in the marmoset. However, this result might have been strongly influenced by the way the data was analysed. In their study, Zeng et al. (2019) did not apply any correction for eliminating neuropil contamination of somatic calcium signals. This contamination results from the diffuse activation of peri-somatic structures, primarily dendrites. Because it is diffuse, it is also synchronous across the entire imaging field and thus, when added to signals from cell bodies, it makes the sensory responses (i.e. frequency preferences in our case) appear more similar across neurons. Interestingly, one of the few mouse studies that did not find disordered local tonotopy, but a very tight one instead (Issa et al., 2014), also did not perform such neuropil correction, potentially offering an explanation for these discrepancies.

Even if we assume that locally disordered tonotopy is a general feature of AC across species, is this really a feature or is it a bug? There have been some interesting theoretical arguments put forward by Terashima and Okada (2012) that argue for the former. The authors explored the correlation structure of both natural images and natural sounds. They argue that unlike natural images, which have mostly local pixel correlations that fall sharply with distance, natural sounds exhibit strong off-diagonal peaks mostly corresponding to the frequency harmonics (Figure 5.1A). They then trained an unsupervised learning model based on topographic independent component analysis (Hyvärinen et al., 2001; Hyvärinen and Hoyer, 2001), which consisted of a two-hidden layer neural network. This network was trained to minimize a sparsity cost function following the principles of “sparse coding” I described in section 1.1 (Olshausen and Field, 1996). They then found using the same network that visual inputs lead



**Fig. 5.1** Is the locally heterogeneous structure of auditory cortex a consequence of its input?

**A**, The correlation matrix of image strips (right) demonstrate only local correlation in the field of view, top panel. The correlation matrix of the human voice spectra (right) demonstrate not only local correlation but also off-diagonal distant correlations produced by harmonics, bottom panel. **B**, Results from the modelling. The retinotopy of the visual map is smooth, whereas the tonotopy of the auditory map is more disordered, although global tonotopy still exists. **C**, The suggested relationships between natural stimulus statistics and topographies. Adapted from [Terashima and Okada \(2012\)](#).

to the development of smooth retinotopic maps, while natural sound inputs lead to a locally heterogeneous, but globally tonotopic map (Figure 5.1B). Thus, Terashima and Okada (2012) speculate that the much smoother V1 retinotopic maps (Smith and Häusser, 2010; Bonin et al., 2011) compared to the “salt and pepper” tonotopic maps are a consequence of the correlation structure of the input stimuli (Figure 5.1C).

The fact that the input to a given cortical area can strongly modulate its structure and properties is underscored by experimental studies that exchanged the peripheral inputs of the visual and auditory systems and suggested the sensory experiences had a dominant effect on cortical organization (Sur et al., 1988; Angelucci et al., 1998; Sharma et al., 2000). Terashima and Okada (2012) also found that the distribution of CFs is not random - instead “artificial neurons” that are closer to one another are more likely to be harmonically related. Unfortunately, the frequency spacing of the tones I presented in this thesis and those in other studies does not allow for the testing of this model prediction, but future studies could investigate this by sampling the frequency space more finely. Finally, the authors found that the specific tonotopic organisation of the network allows the emergence of “pitch neurons” in the second neural network layer, which sample locally and selectively from harmonically related units in the first, in a manner analogous to “simple” and “complex” cells in V1. Such “harmonic template” neurons have in fact been observed in the primary AC of the marmoset (Feng and Wang, 2017). An interesting potential experiment would be to perform dendritic imaging in ferret AC and observe whether the tuning of the dendrites of putative pitch neurons do indeed display tuning to harmonics related to the fundamental frequency of the pitch.

Thus, an important advantage of the more spectrally heterogeneous tonotopic maps, is that neurons in AC can integrate different frequencies in order to generate more complex receptive fields. An advantage of such broad spectral integration is that it could allow AC neurons to form representations of behaviorally-relevant sound sources (Las et al., 2005; Bar-Yosef et al., 2002). In addition, there have been a number of studies in various species showing that AC frequency representations are highly plastic over many timescales (Dahmen and King, 2007). What is more, individual auditory cortical neurons can change their stimulus selectivity rapidly depending on behavioural context (Fritz et al., 2003). The above dynamic modulations of sound frequency processing are only possible if the AC neurons can access a broad range of spectral inputs (Metherate et al., 2005; Winer et al., 2005; Chen et al., 2011; Miller et al., 2001b; Intskirveli et al., 2016).

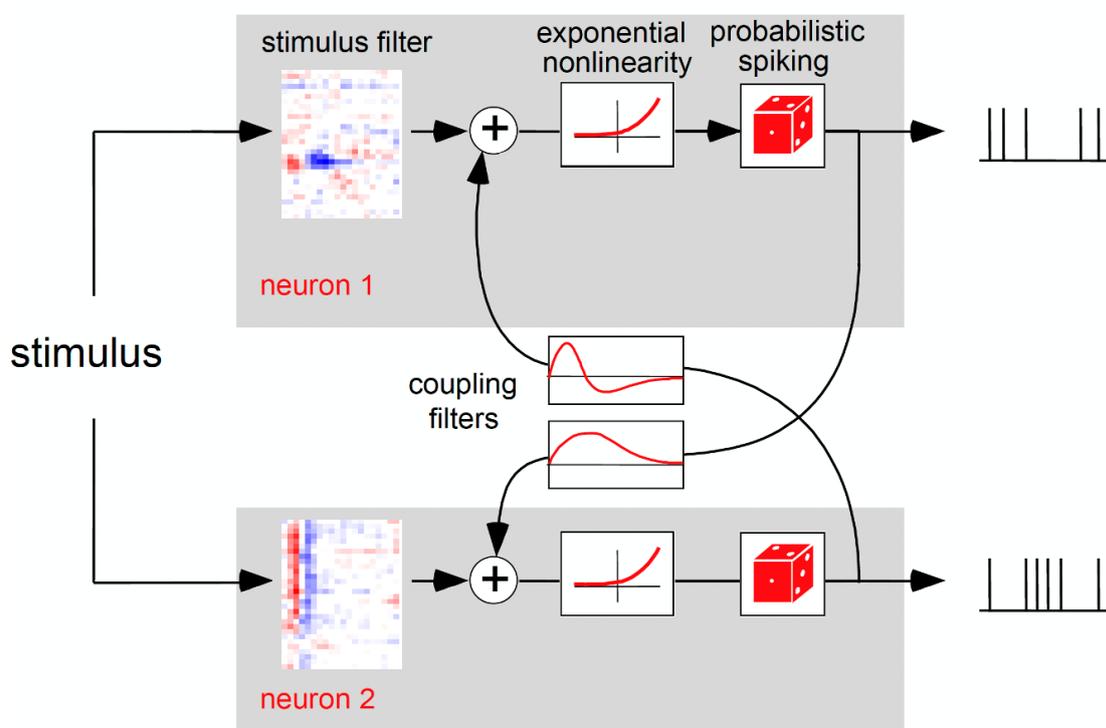
Another important aspect of the order/disorder in tonotopic maps is how these are established during development and what are the respective roles of nature vs nurture. How much of the tonotopic organisation is genetically predetermined and how much is shaped by the environment and experience? The timing of many of the critical properties related to tonotopy and how these develop have been described, however much less is known about the mechanisms that are at play in terms of acquiring specific features in hair cells, extracellular membranes or supporting cells at the sensory periphery (Mann and Kelley, 2011). In general, circuits in the auditory periphery and the brainstem are wired with high tonotopic (topographic) fidelity early in development and show little refinement later on in life. This view is based on a plethora of anatomical tract tracing studies in both mammalian and avian species where the growing axons innervate their tonotopically corresponding target areas from early development and rarely display aberrant, transient connectivity to the wrong nuclei (Friauf and Kandler, 1990; Kandler and Friauf, 1993; Gurung and Fritzsche, 2004; Rubel and Fritzsche, 2002). Spontaneous activity propagates through the auditory system early in development and a recent paper demonstrated the remarkable robustness of the tonotopic maps as inner hair cells with *Vglut3* KO (no glutamate release onto the spiral ganglion cells of the auditory nerve) did not prevent the activation of the auditory nerves and the formation of tonotopic maps (Babola et al., 2018). Even though the auditory system is fairly robust at the global scale, the acoustic environment can have important influences on the finer tonotopic organisation. Interestingly, raising gerbils in an environment with constant broadband noise exposure, reduced the frequency tuning of neurons in the inferior colliculus (Sanes and Constantine-Paton, 1983). Another interesting study comes from the MSO, where the tonotopic refinement of inhibitory inputs was impaired after early exposure to continuous omnidirectional noise (Werthat et al., 2008). Experience dependent refinement of tonotopic maps is much more prominent in the AC (Eggermont, 2008; Keuroghlian and Knudsen, 2007; Chang, 2003; Dahmen and King, 2007). For example, over-exposure to particular sound frequencies leads to an expansion in the AC representation of those frequencies (Zhang et al., 2001; Kim and Bao, 2013). Given the ideas laid out in Figure 5.1, an interesting thought experiment might be how can one induce more or less smooth tonotopic maps, based on the acoustic environment the animal is exposed to? If the animal is reared and exposed primarily to simpler sounds such as pure tones and FM sweeps, which lack first and higher order correlations between the different frequencies (Figure 5.1A, bottom

panel) one might expect to observe a smoother tonotopic organisation in AC. This experiment would be technically challenging, but in principle feasible.

## 5.3 Cortical excitatory-inhibitory interactions and their relationship to receptive field properties

In Chapter 3, I explored the frequency tuning properties of excitatory and inhibitory neurons, while in Chapter 4 I looked at their spectro-temporal feature selectivity in response to natural stimuli, in addition to describing their adaptive properties in the context of reverberation. In Chapter 3, I also carried out some preliminary exploration of their putative connectivity using crosscorrelogram analysis in the electrophysiology data and noise correlation analysis in the electrophysiology and imaging data. In order to have a complete understanding of sensory processing, however, it can be highly illuminating to combine both the neuron's stimulus feature selectivity and its underlying network interactions (Gilbert and Wiesel, 1985; Alonso, 2002; Hamilton et al., 2013).

A relatively straightforward extension exists of the Generalised Linear Model (GLM), similar to the one I used in Chapter 4 to fit neuronal STRFs, which can allow us to address this question (Figure 5.2). Originally described by Pillow et al. (2008), this extension builds on the classical GLM by adding so called “coupling filters” to model the interactions between neurons. Recently, this framework has been augmented in order to accommodate very big data sets and tens to hundreds of neurons, which would work well with both our Neuropixels and calcium imaging data sets (Zoltowski and Pillow, 2018). These coupling filters perform a weighted sum (dot product) of the firing rate of other neurons within a certain history time window and then this output is combined with the output of the linear STRF filtering stage before being fed to the point non-linearity (Figure 5.2). Negative weights in the coupling filters would correspond to an “inhibitory influence”, while positive weights indicate an “excitatory influence”. One could also perform fits with and without the coupling filters, in order to gauge the effects of the neuronal interactions on the STRFs. Relating this to our data, I could use the electrophysiology data where I have identified putative excitatory and inhibitory neurons in order to examine how the two classes of neurons influence each others' receptive fields, in addition to further validating the clustering and correlogram



**Fig. 5.2 Modelling the interactions between connectivity and receptive fields.**

Model schematic for two coupled neurons: each neuron has a stimulus filter, i.e. a STRF and coupling filters that capture dependencies on spiking in other neurons. Summed filter output passes through an exponential non-linearity, followed by a probabilistic process, e.g. Poisson or Bernoulli, in order to produce the instantaneous spike rate. Adapted from [Pillow et al. \(2008\)](#).

approaches themselves. Due to the flexibility of the GLM approach, it would be possible to use either the FRA or the STRF of different neurons to elucidate the interplay between connectivity and both simple and more complex stimulus response properties. With regards to the calcium imaging data, due to slower kinetics of the imaging indicator and the sampling rate of the microscope, it might not be possible to capture the neuronal interactions to the same extent, but this approach could be in theory applied there as well.

However, one must be careful in the interpretation of these coupling filters. The GLM gives us a descriptive statistical model that is oblivious to the underlying anatomy, biophysics or other biological details. The coupling filters reflect the statistical relationships between neurons and not necessarily their synaptic connectivity. To tackle

this issue, [Kobayashi et al. \(2019\)](#) used yet another flavour of the GLM that addressed the connectivity more directly. By applying their method to rat hippocampal data, the authors show that the types of estimated connections match the results inferred from other physiological cues. The [Kobayashi et al. \(2019\)](#) and the [Pillow et al. \(2008\)](#) models could be applied sequentially, such that first the synaptic connectivity is inferred and then the effects of the connections on the tuning properties are examined. These type of analyses could be further supplemented with approaches applying Granger causality to get even better estimates of connectivity and how the interactions between excitatory and inhibitory neurons influence their receptive fields ([Sheikhhattar et al., 2018](#)).

## 5.4 Possible biological implementations of the adaptation to reverberation

In Chapter 4, I have referred to the effects that I see in the neuronal STRFs as “excitatory” and “inhibitory”, but ultimately they are positive and negative values in a statistical model. What might be some plausible biological mechanisms for the cortical adaptation to reverberation that I saw in Chapter 4? In theory, the enhancement of the initial part of the sound (direct arrival and early reflections) and the suppression of late reverberation that I see could be compatible with a number of biological mechanisms. Some plausible mechanisms include synaptic depression ([David et al., 2009](#)), intrinsic dynamics of membrane  $K^+$  channels ([Abolafia et al., 2011](#)), hyperpolarizing inputs from inhibitory neurons ([Li et al., 2015](#); [Natan et al., 2015](#); [Gwak and Kwag, 2020](#)) or adaptation inherited from subcortical regions such as the IC or MGB ([Dean et al., 2008](#); [Devore et al., 2009](#); [Willmore et al., 2016](#); [Lohse et al., 2020a](#)). The present physiological data in Chapter 4 do not allow us to discriminate among these mechanisms.

In this thesis, I explored possible differences between excitatory and inhibitory neurons, which might provide us with some hints regarding mechanistic implementation. I did not find a difference in the adaptation properties that I have measured between the two types of neurons (Figures 4.8, 4.9). A previous study by [Li et al. \(2015\)](#) showed that different classes of inhibitory neurons have different temporal response properties. The PV in the rat auditory cortex showed faster adaptation dynamics compared to pyramidal neurons, whereas SOM showed slower dynamics than both of them. In our

clustering approach, I do not distinguish between different classes of inhibitory neurons, so it is likely that faster PV and slower SOM responses are averaged out, thus showing no difference from pyramidal neurons. Future studies using optogenetic approaches to phototag different classes of inhibitory neurons will prove valuable in elucidating further their possible involvement (Moore and Wehr, 2013; Natan et al., 2015; Kim et al., 2017; Bajo et al., 2019). In addition, I did not explore directly the interaction between putative excitatory and inhibitory neurons. As I have discussed in the section 5.3, extensions of the the basic GLM framework are available, which can allow us to gauge the contribution of inhibitory neurons to the STRF properties of pyramidal neurons (Pillow et al., 2008; Zoltowski and Pillow, 2018; Sheikhattar et al., 2018; Kobayashi et al., 2019).

Regardless of which biological mechanism might be responsible for the adaptation to reverberation, our data suggest that it has to be dynamic. I showed using a simulated linear-nonlinear Poisson (LNP) neuronal model that a static non-linearity is likely not sufficient to explain the adaptation that I observe (Figure 4.6). It is also worth noting, more generally, that a given computational algorithm can have many possible biological implementations (Marr, 1982). In addition, even if the underlying computation is conceptually and algorithmically simple, it may require a fairly sophisticated biological machinery. This point was elegantly shown by Rahman et al. (2020) using cochlear models of various sophistication to predict cortical neuronal responses, finding that simpler models predicted just as well as more biologically detailed ones. In a similar fashion, a recent study used deep neural networks to predict human auditory cortical responses from sound cochleagrams (Keshishian et al., 2020). The authors demonstrated that after fitting a deep neural network to predict the responses of each recording electrode, the resulting network can be transformed to a dynamic STRF, which changes with stimulus input and is mathematically equivalent to the original network, but has the interpretability of a normal STRF. Thus, despite the network having several layers and non-linearities, it can be compactly summarised and interpreted. The authors found that three main type of changes were observed in the dynamic STRFs: gain change, temporal hold, and shape change. Thus, even though our STRF model is relatively simple, it was able to accurately describe the underlying dynamic adaptive changes to reverberation at the conceptual and algorithmic level.

## 5.5 Subcortical and cortical contributions to the adaptation to reverberation

When we think about the biological implementation of the adaptation to reverberation, another important question is whether the changes in the temporal profile of inhibition in the neuronal STRFs are a cortical property and therefore whether adaptation to reverberation is found only at this level? In view of the fact that, in mice, contrast gain control adaptation is equally strong in both the midbrain and thalamus (Lohse et al., 2020a), it would be important to investigate adaptation to reverberation at subcortical levels too, namely the IC and the MGB. Previous research has shown adaptive effects to reverberation at the level of the IC (Slama and Delgutte, 2015). The authors used amplitude modulated broadband noise bursts presented in virtual acoustic space in both anechoic and reverberant conditions. Slama and Delgutte (2015) found that the degradation in temporal coding was smaller than the amplitude modulation attenuation in the stimulus. Thus, these results are consistent with the findings in Chapter 4 where I showed that the adaptation to reverberation is mostly evident in the temporal properties of the neuronal STRFs. Furthermore, it could be the case that the adaptation to reverberation becomes more complete as we progress along the auditory pathway. Rabinowitz et al. (2013) found that subcortical structures are generally affected more by the addition of noise than AC. This and other studies have also shown that the increase in adaptation to sound statistics from auditory nerve to midbrain to cortex is helping in the construction of noise-invariant sound representations in the higher auditory brain (Dean et al., 2005, 2008; Robinson et al., 2016).

## 5.6 Possible confounds from the use of anaesthesia

Given that all experiments in this thesis were performed in anaesthetised animals, one might ask how confident can we be that our results will translate to awake animals? Therefore, I will briefly touch on this issue here and comment on the general effects that anaesthesia might have on studying the properties of AC.

There have been a number of studies that have shown that the processing of information in AC can be sculpted by the spontaneous activity of the network, and such patterns of activity have been dubbed “cortical states” (Poulet and Crochet, 2019;

Issa and Wang, 2008; Marguet and Harris, 2011; Hubel et al., 1959). The dominant cortical state during both slow wave sleep and general anaesthesia is characterised by slow synchronous fluctuations amongst neurons with alternating phases of silence and spiking. Thus, this state is often defined as “synchronized”. Diametrically opposed to this state is the so called “desynchronized” state, which appears during wakefulness, where neurons fire at higher frequencies and in a non-synchronous fashion. Despite these results, it is not correct to view cortical states in this binary fashion as they form more of a continuum between these two extremes and include many possible patterns of activity (Harris and Thiele, 2011).

In addition to exploring the activity of the auditory cortical network as a whole, several studies have investigated how cortical states impinge on the activity of specific classes of neurons in primary AC. For example, Sakata and Harris (2012) found that the activity of excitatory neurons in rat primary AC is enhanced during synchronised states (similar to the state during anaesthesia), and suppressed during the desynchronized states (more akin to wakefulness). The transitions in cortical states occurred spontaneously, but could also be induced. Another study corroborated these findings by showing that “inhibition dominates sensory responses in the awake cortex” and directly measured synaptic conductances in layer 2/3 neurons of V1 in both anaesthetised and awake rats (Haider et al., 2013).

With regards to the results presented in Chapters 2 and 3, the effects of anaesthetics on the tuning properties of AC neurons has been previously explored. During Equithesin (barbiturate) anesthesia, Gaese and Ostwald (2001) found that only a portion of AC neurons were active, and only ~29% of the them exhibited clear frequency selectivity in addition to their tuning sharpness being increased under anesthesia. Similar to the Gaese and Ostwald (2001) study, Noda and Takahashi (2015) found that isoflurane anesthesia decreased the number of active single units as well as increasing their response onset latency. In terms of tuning properties, isoflurane anesthesia shifted the response threshold without changing the shape of the frequency response area. With regards to tonotopy, studies in which tonotopy was tighter were generally conducted under barbiturate anaesthesia (Hackett et al., 2011; Polley et al., 2007).

In the experiments in this thesis, a mix of ketamine and medetomidine was used. Rothschild et al. (2010), who used the same anaesthetic protocol, also reported maps characterized by an intermediate degree of BF variability that I observed in the ferret and the mouse. Recently, Tischbirek et al. (2019) found that recordings of neuronal BF

patterns obtained in awake animals are very similar to those recorded in conditions of light medetomidine, midazolam, and fentanyl (MMF) anaesthesia. Thus, given the similarity in chemical composition and mechanism of action of ketamine and medetomidine anaesthesia to the MMF anaesthesia, it is likely that our results from Chapter 2 regarding the tonotopic organisation of AC should generalise to awake animals as well. With regards to the results in Chapter 3, most previous studies exploring the properties of inhibitory neurons used various different anaesthetic protocols (Moore and Wehr, 2013; Li et al., 2015; Mesik et al., 2015; Maor et al., 2016; Liang et al., 2019). Therefore, it would be interesting to repeat our experiments in awake ferrets to see whether greater differences emerge between excitatory and inhibitory neurons.

When it comes to the results of Chapter 4, previous work has shown no effect of anaesthesia on contrast adaptation in either the ferret AC (Rabinowitz et al., 2011) or the mouse IC (Lohse et al., 2020a). There is therefore no a priori reason to expect that cortical adaptation to reverberation would be any different in awake ferrets. Nevertheless, the effects of attention and behaviour on AC STRFs in the ferret are well documented (David, 2018). These can manifest e.g. as gain changes and tuning shifts. Considering the importance of reverberation to perception and behaviour, it would be very interesting to repeat the experiments in awake and potentially behaving animals. The adaptation effects I observed in the anaesthetised AC might be supplemented by top-down attention or other mechanisms such that a more rich and dynamic adaptation is achieved.



# Bibliography

- A. Jukić, T. van Waterschoot, T. Gerkmann, and S. Doclo (2014). Speech dereverberation with multi-channel linear prediction and sparse priors for the desired signal. In *2014 4th Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA)*, pages 23–26. Journal Abbreviation: 2014 4th Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA).
- Abolafia, J. M., Vergara, R., Arnold, M. M., Reig, R., and Sanchez-Vives, M. V. (2011). Cortical Auditory Adaptation in the Awake Rat and the Role of Potassium Currents. *Cerebral Cortex*, 21(5):977–990.
- Adesnik, H., Bruns, W., Taniguchi, H., Huang, Z. J., and Scanziani, M. (2012). A neural circuit for spatial summation in visual cortex. *Nature*, 490(7419):226–231.
- Adesnik, H. and Naka, A. (2018). Cracking the Function of Layers in the Sensory Cortex. *Neuron*, 100(5):1028–1043.
- Adler, A., Zhao, R., Shin, M. E., Yasuda, R., and Gan, W.-B. (2019). Somatostatin-Expressing Interneurons Enable and Maintain Learning-Dependent Sequential Activation of Pyramidal Neurons. *Neuron*, 102(1):202–216.e7.
- Aertsen, A. M. and Johannesma, P. I. (1981). The spectro-temporal receptive field. A functional characteristic of auditory neurons. *Biological Cybernetics*, 42(2):133–143.
- Ahrens, M. B., Linden, J. F., and Sahani, M. (2008). Nonlinearities and contextual influences in auditory cortical responses modeled with multilinear spectrotemporal methods. *Journal of Neuroscience*, 28(8):1929–42. Publisher: Society for Neuroscience.
- Aizenberg, M., Mwilambwe-Tshilobo, L., Briguglio, J. J., Natan, R. G., and Geffen, M. N. (2015). Bidirectional Regulation of Innate and Learned Behaviors That Rely on Frequency Discrimination by Cortical Inhibitory Neurons. *PLOS Biology*, 13(12):e1002308.
- Aizenberg, M., Rolón-Martínez, S., Pham, T., Rao, W., Haas, J. S., and Geffen, M. N. (2019). Projection from the Amygdala to the Thalamic Reticular Nucleus Amplifies Cortical Sound Responses. *Cell Reports*, 28(3):605–615.e4.
- Alain, C., Arsenault, J. S., Garami, L., Bidelman, G. M., and Snyder, J. S. (2017). Neural Correlates of Speech Segregation Based on Formant Frequencies of Adjacent Vowels. *Scientific Reports*, 7(1):40790.
- Albouy, P., Weiss, A., Baillet, S., and Zatorre, R. J. (2017). Selective Entrainment of Theta Oscillations in the Dorsal Stream Causally Enhances Auditory Working

- Memory Performance. *Neuron*, 94(1):193–206.e5.
- Allen, J. B. and Berkley, D. A. (1979). Image method for efficiently simulating small-room acoustics. *The Journal of the Acoustical Society of America*, 65(4):943–950. Publisher: Acoustical Society of America.
- Alonso, J.-M. (2002). Neural connections and receptive field properties in the primary visual cortex. *The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, 8(5):443–456.
- Andermann, M. L., Gilfoy, N. B., Goldey, G. J., Sachdev, R. N., Wölfel, M., McCormick, D. A., Reid, R. C., and Levene, M. J. (2013). Chronic cellular imaging of entire cortical columns in awake mice using microprisms. *Neuron*, 80(4):900–913. Publisher: Cell Press.
- Anderson, L. A., Christianson, G. B., and Linden, J. F. (2009). Mouse auditory cortex differs from visual and somatosensory cortices in the laminar distribution of cytochrome oxidase and acetylcholinesterase. *Brain Research*, 1252:130–142. Publisher: Elsevier B.V.
- Angelucci, A., Clascá, F., and Sur, M. (1998). Brainstem inputs to the ferret medial geniculate nucleus and the effect of early deafferentation on novel retinal projections to the auditory thalamus. *The Journal of Comparative Neurology*, 400(3):417–439.
- Atencio, C. A. and Schreiner, C. E. (2010). Columnar Connectivity and Laminar Processing in Cat Primary Auditory Cortex. *PLoS ONE*, 5(3):e9521. Publisher: Public Library of Science.
- Atencio, C. A. and Sharpee, T. O. (2017). Multidimensional receptive field processing by cat primary auditory cortical neurons. *Neuroscience*, 359:130–141.
- Atencio, C. A., Sharpee, T. O., and Schreiner, C. E. (2012). Receptive field dimensionality increases from the auditory midbrain to cortex. *Journal of Neurophysiology*, 107(10):2594–2603.
- Attias, H. and Schreiner, C. E. (1996). Temporal Low-Order Statistics of Natural Sounds. In *NIPS*.
- Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, 61(3):183–193.
- Babola, T. A., Li, S., Gribizis, A., Lee, B. J., Issa, J. B., Wang, H. C., Crair, M. C., and Bergles, D. E. (2018). Homeostatic Control of Spontaneous Activity in the Developing Auditory System. *Neuron*, 99(3):511–524.e5.
- Bajo, V. M., Nodal, F. R., Korn, C., Constantinescu, A. O., Mann, E. O., Boyden, E. S., and King, A. J. (2019). Silencing cortical activity during sound-localization training impairs auditory perceptual learning. *Nature Communications*, 10(1):3075.
- Bandyopadhyay, S., Shamma, S. a., and Kanold, P. O. (2010). Dichotomy of functional organization in the mouse auditory cortex. *Nature Neuroscience*, 13(3):361–368. Publisher: Nature Publishing Group.
- Bar-Yosef, O., Rotman, Y., and Nelken, I. (2002). Responses of neurons in cat primary auditory cortex to bird chirps: effects of temporal and spectral context. *Journal of Neuroscience*, 22(19):8619–32. Publisher: Society for Neuroscience.

- Baraban, S. C. and Tallent, M. K. (2004). Interneuron Diversity series: Interneuronal neuropeptides – endogenous regulators of neuronal excitability. *Trends in Neurosciences*, 27(3):135–142.
- Barlow, H. B. (2012). Possible Principles Underlying the Transformations of Sensory Messages. In Rosenblith, W. A., editor, *Sensory Communication*, pages 216–234. The MIT Press.
- Barnstedt, O., Keating, P., Weissenberger, Y., King, A. J., and Dahmen, J. C. (2015). Functional microarchitecture of the mouse dorsal inferior colliculus revealed through in vivo two-photon calcium imaging. *Journal of Neuroscience*, 35(31):10927–10939.
- Barthó, P., Hirase, H., Monconduit, L., Zugaro, M., Harris, K. D., and Buzsáki, G. (2004). Characterization of Neocortical Principal Cells and Interneurons by Network Interactions and Extracellular Features. *Journal of Neurophysiology*, 92(1):600–608.
- Bartlett, E. L. (2013). The organization and physiology of the auditory thalamus and its role in processing acoustic features important for speech perception. *Brain and Language*, 126(1):29–48.
- Bendor, D. and Wang, X. (2008). Neural response properties of primary, rostral, and rostrotemporal core fields in the auditory cortex of marmoset monkeys. *Journal of Neurophysiology*, 100(2):888–906.
- BeraneK, L. L. (2004). *Concert halls and opera houses: music, acoustics, and architecture*. Springer, New York, NY, 2nd ed edition.
- Bieszczad, K. M. and Weinberger, N. M. (2010). Representational gain in cortical area underlies increase of memory strength. *Proceedings of the National Academy of Sciences*, 107(8):3793–3798.
- Biswal, P. K. and Banerjee, S. (2010). A parallel approach for affine transform of 3D biomedical images. In *2010 International Conference on Electronics and Information Engineering*, pages V1–329–V1–332, Kyoto, Japan. IEEE.
- Bizley, J., Walker, K., Nodal, F., King, A., and Schnupp, J. (2013). Auditory Cortex Represents Both Pitch Judgments and the Corresponding Acoustic Cues. *Current Biology*, 23(7):620–625.
- Bizley, J., Walker, K., Silverman, B., King, A., and Schnupp, J. (2009). Interdependent encoding of pitch, timbre, and spatial location in auditory cortex. *Journal of Neuroscience*, 29(7):2064–2075. Edition: 2009/02/21 ISBN: 1529-2401 (Electronic) 0270-6474 (Linking).
- Bizley, J. K. and King, A. J. (2008). Visual–auditory spatial processing in auditory cortical neurons. *Brain Research*, 1242:24–36.
- Bizley, J. K. and King, A. J. (2009). Visual influences on ferret auditory cortex. *Hearing Research*, 258(1-2):55–63.
- Bizley, J. K., Nodal, F. R., Bajo, V. M., Nelken, I., and King, A. J. (2007a). Physiological and anatomical evidence for multisensory interactions in auditory cortex. *Cerebral Cortex (New York, N.Y.: 1991)*, 17(9):2172–2189.
- Bizley, J. K., Nodal, F. R., Nelken, I., and King, A. J. (2005). Functional organization of ferret auditory cortex. *Cerebral Cortex*, 15(10):1637–1653.

- Bizley, J. K., Nodal, F. R., Parsons, C. H., and King, A. J. (2007b). Role of auditory cortex in sound localization in the midsagittal plane. *Journal of Neurophysiology*, 98(3):1763–1774.
- Blackwell, J. M. and Geffen, M. N. (2017). Progress and challenges for understanding the function of cortical microcircuits in auditory processing. *Nature Communications*, 8(1):2165.
- Blackwell, J. M., Taillefumier, T. O., Natan, R. G., Carruthers, I. M., Magnasco, M. O., and Geffen, M. N. (2016). Stable encoding of sounds over a broad range of statistical parameters in the auditory cortex. *The European Journal of Neuroscience*, 43(6):751–764.
- Blesser, B. and Salter, L.-R. (2009). *Spaces speak, are you listening? experiencing aural architecture*. MIT Press, Cambridge, Mass. OCLC: 845577268.
- Bonin, V., Histed, M. H., Yurgenson, S., and Reid, R. C. (2011). Local diversity and fine-scale organization of receptive fields in mouse visual cortex. *Journal of Neuroscience*, 31(50):18506–21.
- Bootman, M. D., Allman, S., Rietdorf, K., and Bultynck, G. (2018). Deleterious effects of calcium indicators within cells; an inconvenient truth. *Cell Calcium*, 73:82–87.
- Bos, H., Oswald, A.-M., and Doiron, B. (2020). Untangling stability and gain modulation in cortical circuits with multiple interneuron classes. preprint, Neuroscience.
- Boyden, E. S., Zhang, F., Bamberg, E., Nagel, G., and Deisseroth, K. (2005). Millisecond-timescale, genetically targeted optical control of neural activity. *Nature Neuroscience*, 8(9):1263–1268.
- Bradley, J. S. (1986). Speech intelligibility studies in classrooms. *The Journal of the Acoustical Society of America*, 80(3):846–854.
- Bradley, J. S., Sato, H., and Picard, M. (2003). On the importance of early reflections for speech in rooms. *The Journal of the Acoustical Society of America*, 113(6):3233.
- Brand, A., Behrend, O., Marquardt, T., McAlpine, D., and Grothe, B. (2002). Precise inhibition is essential for microsecond interaural time difference coding. *Nature*, 417(6888):543–547.
- Bronkhorst, A. W. and Houtgast, T. (1999). Auditory distance perception in rooms. *Nature*, 397(6719):517–520.
- Brown, G. J. and Cooke, M. (1994). Computational auditory scene analysis. *Computer Speech & Language*, 8(4):297–336.
- Brughera, A., Mikiel-Hunter, J., Dietz, M., and McAlpine, D. (2019). Auditory brainstem models: adapting cochlear nuclei improve spatial encoding by the medial superior olive in reverberation. preprint, Neuroscience.
- Brumm, H. and Naguib, M. (2009). Chapter 1 Environmental Acoustics and the Evolution of Bird Song. In *Advances in the Study of Behavior*, volume 40, pages 1–33. Elsevier.
- Buzsáki, G. and Wang, X.-J. (2012). Mechanisms of gamma oscillations. *Annual Review of Neuroscience*, 35:203–225.
- Cabreraa, D., Jeongb, D., Kwakc, H. J., and Kimd, J.-Y. (2005). Auditory Room Size Perception for Modeled and Measured Rooms.

- Calabrese, A., Schumacher, J. W., Schneider, D. M., Paninski, L., and Woolley, S. M. N. (2011). A Generalized Linear Model for Estimating Spectrotemporal Receptive Fields from Responses to Natural Sounds. *PLOS ONE*, 6(1):e16104. Publisher: Public Library of Science.
- Calford, M. and Aitkin, L. (1983). Ascending projections to the medial geniculate body of the cat: evidence for multiple, parallel auditory pathways through thalamus. *The Journal of Neuroscience*, 3(11):2365–2380.
- Campbell, D. R., Palomaki, K. J., and Brown, G. (2005). A MATLAB simulation of "shoebox" room acoustics for use in research and teaching. *Computing and Information Systems J*, 9:48–51.
- Caputi, A., Rozov, A., Blatow, M., and Monyer, H. (2009). Two calretinin-positive GABAergic cell types in layer 2/3 of the mouse neocortex provide different forms of inhibition. *Cerebral Cortex (New York, N.Y.: 1991)*, 19(6):1345–1359.
- Carandini, M. (2005). Do We Know What the Early Visual System Does? *Journal of Neuroscience*, 25(46):10577–10597.
- Carandini, M. and Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, 13(1):51–62.
- Caras, M. L. and Sanes, D. H. (2017). Top-down modulation of sensory cortex gates perceptual learning. *Proceedings of the National Academy of Sciences*, 114(37):9972–9977.
- Carbajal, G. V. and Malmierca, M. S. (2018). The Neuronal Basis of Predictive Coding Along the Auditory Pathway: From the Subcortical Roots to Cortical Deviance Detection. *Trends in Hearing*, 22:233121651878482.
- Carcea, I. and Froemke, R. C. (2013). Cortical plasticity, excitatory-inhibitory balance, and sensory perception. *Progress in Brain Research*, 207:65–90.
- Carr, C. E. (1993). Processing of Temporal Information in the Brain. *Annual Review of Neuroscience*, 16(1):223–243. Publisher: Annual Reviews.
- Carruthers, I. M., Laplagne, D. A., Jaegle, A., Briguglio, J. J., Mwilambwe-Tshilobo, L., Natan, R. G., and Geffen, M. N. (2015). Emergence of invariant representation of vocalizations in the auditory cortex. *Journal of Neurophysiology*, 114(5):2726–2740.
- Casseday, J. H. and Winer, J. A. (2005). The Inferior Colliculus: Past, Present, and Future. In Winer, J. A. and Schreiner, C. E., editors, *The Inferior Colliculus*, pages 626–640. Springer-Verlag, New York.
- Chang, E. F. (2003). Environmental Noise Retards Auditory Cortical Development. *Science*, 300(5618):498–502.
- Chechik, G. and Nelken, I. (2012). Auditory abstraction from spectro-temporal features to coding auditory entities. *Proceedings of the National Academy of Sciences*, 109(46):18968–18973.
- Chen, T.-W., Wardill, T. J., Sun, Y., Pulver, S. R., Renninger, S. L., Baohan, A., Schreiter, E. R., Kerr, R. A., Orger, M. B., Jayaraman, V., Looger, L. L., Svoboda, K., and Kim, D. S. (2013). Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature*, 499(7458):295–300.

- Chen, X., Leischner, U., Rochefort, N. L., Nelken, I., and Konnerth, A. (2011). Functional mapping of single spines in cortical neurons in vivo. *Nature*, 475(7357):501–505.
- Cheng, C. I. and Wakefield, G. H. (2001). Introduction to Head-Related Transfer Functions (HRTFs): Representations of HRTFs in Time, Frequency, and Space. *Journal of the Audio Engineering Society*, 49(4):231–249. Publisher: Audio Engineering Society.
- Christianson, G. B., Sahani, M., and Linden, J. F. (2011). Depth-Dependent Temporal Response Properties in Core Auditory Cortex. *Journal of Neuroscience*, 31(36):12837–12848.
- Christison-Lagay, K. L., Gifford, A. M., and Cohen, Y. E. (2015). Neural correlates of auditory scene analysis and perception. *International Journal of Psychophysiology*, 95(2):238–245.
- Clugnet, M. C., LeDoux, J. E., and Morrison, S. F. (1990). Unit responses evoked in the amygdala and striatum by electrical stimulation of the medial geniculate body. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 10(4):1055–1061.
- Cobos, I., Calcagnotto, M. E., Vilaythong, A. J., Thwin, M. T., Noebels, J. L., Baraban, S. C., and Rubenstein, J. L. R. (2005). Mice lacking *Dlx1* show subtype-specific loss of interneurons, reduced inhibition and epilepsy. *Nature Neuroscience*, 8(8):1059–1068.
- Cohen, M. R. and Kohn, A. (2011). Measuring and interpreting neuronal correlations. *Nature Neuroscience*, 14(7):811–819.
- Cohen, Y. E., Bennur, S., Christison-Lagay, K., Gifford, A. M., and Tsunada, J. (2016). Functional Organization of the Ventral Auditory Pathway. *Advances in Experimental Medicine and Biology*, 894:381–388.
- Connors, B. W. and Gutnick, M. J. (1990). Intrinsic firing patterns of diverse neocortical neurons. *Trends in Neurosciences*, 13(3):99–104.
- Constantinople, C. M. and Bruno, R. M. (2013). Deep Cortical Layers Are Activated Directly by Thalamus. *Science*, 340(6140):1591–1594.
- Cooke, J. E., King, A. J., Willmore, B. D. B., and Schnupp, J. W. H. (2018). Contrast gain control in mouse auditory cortex. *Journal of Neurophysiology*, 120(4):1872–1884.
- Cossell, L., Iacaruso, M. F., Muir, D. R., Houlton, R., Sader, E. N., Ko, H., Hofer, S. B., and Mrsic-Flogel, T. D. (2015). Functional organization of excitatory synaptic strength in primary visual cortex. *Nature*, 518(7539):399–403. Publisher: Nature Publishing Group.
- Couey, J. J., Witoelar, A., Zhang, S.-J., Zheng, K., Ye, J., Dunn, B., Czajkowski, R., Moser, M.-B., Moser, E. I., Roudi, Y., and Witter, M. P. (2013). Recurrent inhibitory circuitry as a mechanism for grid formation. *Nature Neuroscience*, 16(3):318–324.
- Culling, J. F., Hodder, K. I., and Toh, C. Y. (2003). Effects of reverberation on perceptual segregation of competing voices. *The Journal of the Acoustical Society of America*, 114(5):2871.
- Culling, J. F., Summerfield, Q., and Marshall, D. H. (1994). Effects of simulated reverberation on the use of binaural cues and fundamental-frequency differences for

- separating concurrent vowels. *Speech Communication*, 14(1):71–95.
- Da Costa, S., Clarke, S., and Crottaz-Herbette, S. (2018). Keeping track of sound objects in space: The contribution of early-stage auditory areas. *Hearing Research*, 366:17–31.
- Dahmen, J. C., Hartley, D. E. H., and King, A. J. (2008). Stimulus-timing-dependent plasticity of cortical frequency representation. *Journal of Neuroscience*, 28(50):13629–39.
- Dahmen, J. C. and King, A. J. (2007). Learning to hear: plasticity of auditory cortical processing. *Current Opinion in Neurobiology*, 17(4):456–464.
- Dana, H., Mohar, B., Sun, Y., Narayan, S., Gordus, A., Hasseman, J. P., Tsegaye, G., Holt, G. T., Hu, A., Walpita, D., Patel, R., Macklin, J. J., Bargmann, C. I., Ahrens, M. B., Schreiter, E. R., Jayaraman, V., Looger, L. L., Svoboda, K., and Kim, D. S. (2016). Sensitive red protein calcium indicators for imaging neural activity. *eLife*, 5:e12727.
- Dana, H., Sun, Y., Mohar, B., Hulse, B. K., Kerlin, A. M., Hasseman, J. P., Tsegaye, G., Tsang, A., Wong, A., Patel, R., Macklin, J. J., Chen, Y., Konnerth, A., Jayaraman, V., Looger, L. L., Schreiter, E. R., Svoboda, K., and Kim, D. S. (2019). High-performance calcium sensors for imaging activity in neuronal populations and microcompartments. *Nature Methods*, 16(7):649–657.
- Dantzker, J. L. and Callaway, E. M. (2000). Laminar sources of synaptic input to cortical inhibitory interneurons and pyramidal neurons. *Nature Neuroscience*, 3(7):701–707.
- Darwin, C. J. and Hukin, R. W. (2000). Effects of reverberation on spatial, prosodic, and vocal-tract size cues to selective attention. *The Journal of the Acoustical Society of America*, 108(1):335–342.
- David, S. V. (2018). Incorporating behavioral and sensory context into spectro-temporal models of auditory encoding. *Hearing Research*, 360:107–123.
- David, S. V., Mesgarani, N., Fritz, J. B., and Shamma, S. A. (2009). Rapid Synaptic Depression Explains Nonlinear Modulation of Spectro-Temporal Tuning in Primary Auditory Cortex by Natural Stimuli. *Journal of Neuroscience*, 29(11):3374–3386.
- de Boer, R. and Kuyper, P. (1968). Triggered correlation. *IEEE transactions on bio-medical engineering*, 15(3):169–179.
- Dean, I., Harper, N. S., and McAlpine, D. (2005). Neural population coding of sound level adapts to stimulus statistics. *Nature neuroscience*, 8(12):1684–1689. ISBN: 1097-6256 (Print).
- Dean, I., Robinson, B. L., Harper, N. S., and McAlpine, D. (2008). Rapid neural adaptation to sound level statistics. *J Neurosci*, 28(25):6430–6438. ISBN: 1529-2401 (Electronic)\r0270-6474 (Linking).
- DeFelipe, J. and Jones, E. G. (1988). *Cajal on the cerebral cortex: an annotated translation of the complete writings*. Number no. 1 in History of neuroscience. Oxford University Press, New York.
- Delgutte, B. (2002). Auditory Neural Processing of Speech.

- Deneux, T., Kempf, A., Daret, A., Ponsot, E., and Bathellier, B. (2016). Temporal asymmetries in auditory coding and perception reflect multi-layered nonlinearities. *Nature Communications*, 7(1):12682.
- Descartes, R. (1664). *Traite de l'homme*, volume 11. Paris: Angot.
- Deubner, J., Coulon, P., and Diester, I. (2019). Optogenetic approaches to study the mammalian brain. *Current Opinion in Structural Biology*, 57:157–163.
- Devore, S., Ihlefeld, A., Hancock, K., Shinn-Cunningham, B., and Delgutte, B. (2009). Accurate Sound Localization in Reverberant Environments Is Mediated by Robust Encoding of Spatial Cues in the Auditory Midbrain. *Neuron*, 62(1):123–134.
- Dimidschstein, J., Chen, Q., Tremblay, R., Rogers, S. L., Saldi, G.-A., Guo, L., Xu, Q., Liu, R., Lu, C., Chu, J., Grimley, J. S., Krostag, A.-R., Kaykas, A., Avery, M. C., Rashid, M. S., Baek, M., Jacob, A. L., Smith, G. B., Wilson, D. E., Kosche, G., Kruglikov, I., Rusielewicz, T., Kotak, V. C., Mowery, T. M., Anderson, S. A., Callaway, E. M., Dasen, J. S., Fitzpatrick, D., Fossati, V., Long, M. A., Noggle, S., Reynolds, J. H., Sanes, D. H., Rudy, B., Feng, G., and Fishell, G. (2016). A viral strategy for targeting and manipulating interneurons across vertebrate species. *Nature Neuroscience*, 19(12):1743–1749.
- Dodds, E. M. and DeWeese, M. R. (2019). On the Sparse Structure of Natural Sounds and Natural Images: Similarities, Differences, and Implications for Neural Coding. *Frontiers in Computational Neuroscience*, 13. Publisher: Frontiers.
- Douglas, R. J. and Martin, K. A. C. (2004). Neuronal circuits of the neocortex. *Annual Review of Neuroscience*, 27:419–451.
- Druga, R. and Syka, J. (1984). Projections from auditory structures to the superior colliculus in the rat. *Neuroscience Letters*, 45(3):247–252.
- Eggermont, J. J. (2008). The Role of Sound in Adult and Developmental Auditory Cortical Plasticity. *Ear & Hearing*, 29(6):819–829.
- Egner, T. and Summerfield, C. (2013). Grounding predictive coding models in empirical neuroscience research. *The Behavioral and Brain Sciences*, 36(3):210–211.
- Elgueda, D., Duque, D., Radtke-Schuller, S., Yin, P., David, S. V., Shamma, S. A., and Fritz, J. B. (2019). State-dependent encoding of sound and behavioral meaning in a tertiary region of the ferret auditory cortex. *Nature Neuroscience*, 22(3):447–459.
- F. J. Gall, G. S. (1815). Anatomie et Physiologie du Système Nerveux en général, et du Cerveau en particulier; avec des Observations sur la Possibilité de reconnoître plusieurs Dispositions intellectuelles et morales de l'Homme et des Animaux par la Configuration de leur Têtes. Par. *The Edinburgh Review*, 25(49):227. Place: Edinburgh, etc. Publisher: A. Constable, etc.
- Farb, C. R. and LeDoux, J. E. (1997). NMDA and AMPA receptors in the lateral nucleus of the amygdala are postsynaptic to auditory thalamic afferents. *Synapse (New York, N.Y.)*, 27(2):106–121.
- Feng, L. and Wang, X. (2017). Harmonic template neurons in primate auditory cortex underlying complex sound processing. *Proceedings of the National Academy of Sciences USA*, page 201607519.

- Field, D. J. (1994). What Is the Goal of Sensory Coding? *Neural Computation*, 6(4):559–601.
- Fishman, Y. I., Steinschneider, M., and Micheyl, C. (2014). Neural representation of concurrent harmonic sounds in monkey primary auditory cortex: implications for models of auditory scene analysis. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 34(37):12425–12443.
- Fitzpatrick, D. C., Kuwada, S., Kim, D. O., Parham, K., and Batra, R. (1999). Responses of neurons to click-pairs as simulated echoes: Auditory nerve to auditory cortex. *The Journal of the Acoustical Society of America*, 106(6):3460–3472.
- Flourens, P. (1842). *Auteur Du Texte. Recherches Expérimentales Sur Les Propriétés Et Les Fonctions Du Système Nerveux Dans Les Animaux Vertébrés*. Par P. Flourens. J.-B. Baillière, 2e édition, corrigée edition.
- Francis, N. A., Winkowski, D. E., Sheikhattar, A., Armengol, K., Babadi, B., and Kanold, P. O. (2018). Small Networks Encode Decision-Making in Primary Auditory Cortex. *Neuron*, 97(4):885–897.e6.
- Friauf, E. and Kandler, K. (1990). Auditory projections to the inferior colliculus of the rat are present by birth. *Neuroscience Letters*, 120(1):58–61.
- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1456):815–836.
- Friston, K. (2012). The history of the future of the Bayesian brain. *NeuroImage*, 62(2):1230–1233.
- Fritz, J., Shamma, S., Elhilali, M., and Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nature Neuroscience*, 6(11):1216–1223.
- Fritz, J. B. (2005). Differential Dynamic Plasticity of A1 Receptive Fields during Multiple Spectral Tasks. *Journal of Neuroscience*, 25(33):7623–7635.
- Fritz, J. B., Elhilali, M., and Shamma, S. A. (2007). Adaptive changes in cortical receptive fields induced by attention to complex sounds. *Journal of Neurophysiology*, 98(4):2337–2346.
- Froemke, R. C., Carcea, I., Barker, A. J., Yuan, K., Seybold, B. A., Martins, A. R. O., Zaika, N., Bernstein, H., Wachs, M., Levis, P. A., Polley, D. B., Merzenich, M. M., and Schreiner, C. E. (2013). Long-term modification of cortical synapses improves sensory perception. *Nature Neuroscience*, 16(1):79–88.
- Fuglsang, S. A., Dau, T., and Hjortkjær, J. (2017). Noise-robust cortical tracking of attended speech in real-world acoustic scenes. *NeuroImage*, 156:435–444.
- Fujisawa, S., Amarasingham, A., Harrison, M. T., and Buzsáki, G. (2008). Behavior-dependent short-term assembly dynamics in the medial prefrontal cortex. *Nature Neuroscience*, 11(7):823–833.
- Gaese, B. H. and Ostwald, J. (2001). Anesthesia changes frequency tuning of neurons in the rat primary auditory cortex. *Journal of Neurophysiology*, 86(2):1062–1066.
- Gao, X. and Wehr, M. (2015). A Coding Transformation for Temporally Structured Sounds within Auditory Cortical Neurons. *Neuron*, 86(1):292–303.

- Gaubitch, N. D., Thomas, M. R. P., and Naylor, P. A. (2010). Dereverberation Using LPC-based Approaches. In Naylor, P. A. and Gaubitch, N. D., editors, *Speech Dereverberation*, pages 95–128. Springer London, London. Series Title: Signals and Communication Technology.
- Gaucher, Q., Panniello, M., Ivanov, A. Z., Dahmen, J. C., King, A. J., and Walker, K. M. (2020). Complexity of frequency receptive fields predicts tonotopic variability across species. *eLife*, 9:e53462.
- George, S. S., Shivdasani, M. N., and Fallon, J. B. (2016). Effect of current focusing on the sensitivity of inferior colliculus neurons to amplitude-modulated stimulation. *Journal of Neurophysiology*, 116(3):1104–1116.
- Ghazanfar, A. A., Chandrasekaran, C., and Logothetis, N. K. (2008). Interactions between the Superior Temporal Sulcus and Auditory Cortex Mediate Dynamic Face/Voice Integration in Rhesus Monkeys. *Journal of Neuroscience*, 28(17):4457–4469.
- Gidon, A. and Segev, I. (2012). Principles Governing the Operation of Synaptic Inhibition in Dendrites. *Neuron*, 75(2):330–341.
- Gilbert, C. D. and Wiesel, T. N. (1985). Intrinsic connectivity and receptive field properties in visual cortex. *Neural Basis of Visual Perception*, 25(3):365–374.
- Giraud, A.-L. and Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience*, 15(4):511–517.
- Glennon, E., Carcea, I., Martins, A. R. O., Multani, J., Shehu, I., Svirsky, M. A., and Froemke, R. C. (2019). Locus coeruleus activation accelerates perceptual learning. *Brain Research*, 1709:39–49.
- Goldstein, M. H., Abeles, M., Daly, R. L., and McIntosh, J. (1970). Functional architecture in cat primary auditory cortex: tonotopic organization. *Journal of Neurophysiology*, 33(1):188–197.
- Goltz, F. (1892). Der hund ohne grosshirn. *Archiv für die gesamte Physiologie des Menschen und der Tiere*, 51(11):570–614. ISBN: 1432-2013 Publisher: Springer.
- Gourévitch, B. and Brette, R. (2012). The impact of early reflections on binaural cues. *The Journal of the Acoustical Society of America*, 132(1):9–27.
- Greenberg, D. S., Wallace, D. J., Voit, K.-M., Wuertenberger, S., Czubayko, U., Monsees, A., Handa, T., Vogelstein, J. T., Seifert, R., Groemping, Y., and Kerr, J. N. (2018). Accurate action potential inference from a calcium sensor protein through biophysical modeling: preprint, Neuroscience.
- Griesinger, D. (2009). The importance of the direct to reverberant ratio in the perception of distance, localization, clarity, and envelopment, Part two. *The Journal of the Acoustical Society of America*, 125(4):2483–2483.
- Griffiths, T. D. and Warren, J. D. (2004). What is an auditory object? *Nature Reviews Neuroscience*, 5(11):887–892.
- Guediche, S., Blumstein, S. E., Fiez, J. A., and Holt, L. L. (2014). Speech perception under adverse conditions: insights from behavioral, computational, and neuroscience research. *Frontiers in Systems Neuroscience*, 7(4):328–32.

- Guo, W., Chambers, a. R., Darrow, K. N., Hancock, K. E., Shinn-Cunningham, B. G., and Polley, D. B. (2012). Robustness of cortical topography across fields, laminae, anesthetic states, and neurophysiological signal types. *Journal of Neuroscience*, 32(27):9159–9172.
- Gurung, B. and Fritzsche, B. (2004). Time course of embryonic midbrain and thalamic auditory connection development in mice as revealed by carbocyanine dye tracing. *The Journal of Comparative Neurology*, 479(3):309–327.
- Gwak, J. and Kwag, J. (2020). Distinct subtypes of inhibitory interneurons differentially promote the propagation of rate and temporal codes in the feedforward neural network. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 30(5):053102.
- Hackett, T. A., Rinaldi Barkat, T., O’Brien, B. M. J., Hensch, T. K., and Polley, D. B. (2011). Linking topography to tonotopy in the mouse auditory thalamocortical circuit. *Journal of Neuroscience*, 31(8):2983–2995.
- Haggerty, D. L., Grecco, G. G., Reeves, K. C., and Atwood, B. (2020). Adeno-Associated Viral Vectors in Neuroscience Research. *Molecular Therapy - Methods & Clinical Development*, 17:69–82.
- Haider, B., Häusser, M., and Carandini, M. (2013). Inhibition dominates sensory responses in the awake cortex. *Nature*, 493(7430):97–100.
- Hamilton, L., Sohl-Dickstein, J., Huth, A., Carels, V., Deisseroth, K., and Bao, S. (2013). Optogenetic Activation of an Inhibitory Network Enhances Feedforward Functional Connectivity in Auditory Cortex. *Neuron*, 80(4):1066–1076.
- Harper, N. S., Schoppe, O., Willmore, B. D. B., Cui, Z., Schnupp, J. W. H., and King, A. J. (2016). Network Receptive Field Modeling Reveals Extensive Integration and Multi-feature Selectivity in Auditory Cortical Neurons. *PLOS Computational Biology*, 12(11):e1005113. Publisher: Public Library of Science.
- Harris, K. D. (2020). Nonsense correlations in neuroscience. *bioRxiv*, page 2020.11.29.402719.
- Harris, K. D. and Thiele, A. (2011). Cortical state and attention. *Nature Reviews Neuroscience*, 12(9):509–523.
- Hartmann, W. M. (1982). Localization of sound in rooms. *The Journal of the Acoustical Society of America*, 72(S1):S93–S93.
- Hattori, R., Kuchibhotla, K. V., Froemke, R. C., and Komiyama, T. (2017). Functions and dysfunctions of neocortical inhibitory neuron subtypes. *Nature Neuroscience*, 20(9):1199–1208.
- Heffner, H. E. and Heffner, R. S. (2007). Hearing ranges of laboratory animals. *Journal of the American Association for Laboratory Animal Science : JAALAS*, 46(1):20–2.
- Heil, P., Neubauer, H., and Irvine, D. R. F. (2011). An Improved Model for the Rate-Level Functions of Auditory-Nerve Fibers. *Journal of Neuroscience*, 31(43):15424–15437.
- Heilbron, M. and Chait, M. (2018). Great Expectations: Is there Evidence for Predictive Coding in Auditory Cortex? *Neuroscience*, 389:54–73.
- Heinz, R. (1993). Binaural room simulation based on an image source model with addition of statistical methods to include the diffuse sound scattering of walls and

- to predict the reverberant tail. *Applied Acoustics*, 38(2):145–159.
- Helfer, K. S. and Wilber, L. A. (1990). Hearing loss, aging, and speech perception in reverberation and noise. *Journal of speech and hearing research*, 33(1):149–55.
- Herrick, C. J. (1926). *Brains of rats and men*. Publisher: University of Chicago Press.
- Hertäg, L. and Sprekeler, H. (2020). Learning prediction error neurons in a canonical interneuron circuit. *eLife*, 9:e57541.
- Hind, J. E. (1953). An electrophysiological determination of tonotopic organization in auditory cortex of cat. *Journal of Neurophysiology*, 16(5):475–89.
- Hine, J. E., Martin, R. L., and Moore, D. R. (1994). Free-field binaural unmasking in ferrets. *Behavioral Neuroscience*, 108(1):196–205.
- Hofer, S. B., Ko, H., Pichler, B., Vogelstein, J., Ros, H., Zeng, H., Lein, E., Lesica, N. A., and Mrsic-Flogel, T. D. (2011). Differential connectivity and response dynamics of excitatory and inhibitory neurons in visual cortex. *Nature Neuroscience*, 14(8):1045–1052. Publisher: Nature Publishing Group.
- Houtgast, T. and Steeneken, H. J. M. (1985). A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. *The Journal of the Acoustical Society of America*, 77(3):1069–1077.
- Hromádka, T., Dewese, M. R., and Zador, A. M. (2008). Sparse representation of sounds in the unanesthetized auditory cortex. *PLoS biology*, 6(1):e16.
- Hu, Y. and Kokkinakis, K. (2014). Effects of early and late reflections on intelligibility of reverberated speech by cochlear implant listeners. *The Journal of the Acoustical Society of America*, 135(1):EL22–EL28.
- Hubel, D. H., Henson, C. O., Rupert, A., and Galambos, R. (1959). Attention units in the auditory cortex. *Science (New York, N.Y.)*, 129(3358):1279–1280.
- Hubel, D. H. and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. *The Journal of Physiology*, 160(1):106–154.
- Huffman, R. F. and Henson, O. (1990). The descending auditory pathway and acousticomotor systems: connections with the inferior colliculus. *Brain Research Reviews*, 15(3):295–323.
- Huisman, W. H. T. and Attenborough, K. (1991). Reverberation and attenuation in a pine forest. *The Journal of the Acoustical Society of America*, 90(5):2664–2677.
- Humes, L. E., Dirks, D. D., Bell, T. S., Ahlstrom, C., and Kincaid, G. E. (1986). Application of the Articulation Index and the Speech Transmission Index to the recognition of speech by normal-hearing and hearing-impaired listeners. *Journal of speech and hearing research*, 29(4):447–62.
- Hyvärinen, A. and Hoyer, P. O. (2001). A two-layer sparse coding model learns simple and complex cell receptive fields and topography from natural images. *Vision Research*, 41(18):2413–2423.
- Hyvärinen, A., Hoyer, P. O., and Inki, M. (2001). Topographic Independent Component Analysis. *Neural Computation*, 13(7):1527–1558.
- Imaizumi, K., Priebe, N. J., Crum, P. A. C., Bedenbaugh, P. H., Cheung, S. W., and Schreiner, C. E. (2004). Modular functional organization of cat anterior auditory

- field. *Journal of Neurophysiology*, 92(1):444–457. Publisher: American Physiological Society.
- Imig, T. J. and Morel, A. (1985). Tonotopic organization in ventral nucleus of medial geniculate body in the cat. *Journal of Neurophysiology*, 53(1):309–340.
- Insel, N. and Barnes, C. A. (2015). Differential Activation of Fast-Spiking and Regular-Firing Neuron Populations During Movement and Reward in the Dorsal Medial Frontal Cortex. *Cerebral Cortex*, 25(9):2631–2647.
- Intskirveli, I., Joshi, A., Vizcarra-Chacón, B. J., and Metherate, R. (2016). Spectral breadth and laminar distribution of thalamocortical inputs to A1. *Journal of Neurophysiology*, 115(4):2083–2094.
- Irvine, D. R. (2018). Plasticity in the auditory system. *Hearing Research*, 362:61–73.
- Irvine, D. R. and Gago, G. (1990). Binaural interaction in high-frequency neurons in inferior colliculus of the cat: effects of variations in sound pressure level on sensitivity to interaural intensity differences. *Journal of Neurophysiology*, 63(3):570–591.
- Isaacson, J. and Scanziani, M. (2011). How Inhibition Shapes Cortical Activity. *Neuron*, 72(2):231–243.
- Ison, J. R., Allen, P. D., and O’Neill, W. E. (2007). Age-related hearing loss in C57BL/6J mice has both frequency-specific and non-frequency-specific components that produce a hyperacusis-like exaggeration of the acoustic startle reflex. *Journal of the Association for Research in Otolaryngology*, 8(4):539–550.
- Issa, E. B. and Wang, X. (2008). Sensory Responses during Sleep in Primate Primary and Secondary Auditory Cortex. *Journal of Neuroscience*, 28(53):14467–14480.
- Issa, J. B., Haeffele, B. D., Agarwal, A., Bergles, D. E., Young, E. D., and Yue, D. T. (2014). Multiscale Optical Ca<sup>2+</sup> Imaging of Tonal Organization in Mouse Auditory Cortex. *Neuron*, 83(4):944–959. ISBN: 08966273 Publisher: Elsevier Inc.
- Ito, T. and Oliver, D. L. (2014). Local and commissural IC neurons make axosomatic inputs on large GABAergic tectothalamic neurons. *The Journal of Comparative Neurology*, 522(15):3539–3554.
- Jadi, M., Polsky, A., Schiller, J., and Mel, B. W. (2012). Location-Dependent Effects of Inhibition on Local Spiking in Pyramidal Neuron Dendrites. *PLoS Computational Biology*, 8(6):e1002550.
- Johnson, J. S., Niwa, M., O’Connor, K. N., and Sutter, M. L. (2020). Amplitude modulation encoding in the auditory cortex: comparisons between the primary and middle lateral belt regions. *Journal of Neurophysiology*, 124(6):1706–1726.
- Juavinett, A. L., Bekheet, G., and Churchland, A. K. (2019). Chronically implanted Neuropixels probes enable high-yield recordings in freely moving mice. *eLife*, 8:e47188.
- Jun, J. J., Steinmetz, N. A., Siegle, J. H., Denman, D. J., Bauza, M., Barbarits, B., Lee, A. K., Anastassiou, C. A., Andrei, A., Aydın, , Barbic, M., Blanche, T. J., Bonin, V., Couto, J., Dutta, B., Gratiy, S. L., Gutnisky, D. A., Häusser, M., Karsh, B., Ledochowitsch, P., Lopez, C. M., Mitelut, C., Musa, S., Okun, M., Pachitariu, M., Putzeys, J., Rich, P. D., Rossant, C., Sun, W.-l., Svoboda, K., Carandini, M., Harris, K. D., Koch, C., O’Keefe, J., and Harris, T. D. (2017). Fully integrated silicon probes for high-density recording of neural activity. *Nature*, 551(7679):232–236.

- K Harting, J. and Van Lieshout, D. P. (2000). Projections from the rostral pole of the inferior colliculus to the cat superior colliculus. *Brain Research*, 881(2):244–247.
- Kacelnik, O., Nodal, F. R., Parsons, C. H., and King, A. J. (2006). Training-Induced Plasticity of Auditory Localization in Adult Mammals. *PLoS Biology*, 4(4):e71.
- Kadia, S. C. and Wang, X. (2003). Spectral Integration in A1 of Awake Primates: Neurons With Single- and Multi-peaked Tuning Characteristics. *Journal of Neurophysiology*, 89(3):1603–1622.
- Kandler, K. and Friauf, E. (1993). Pre- and postnatal development of efferent connections of the cochlear nucleus in the rat. *The Journal of Comparative Neurology*, 328(2):161–184.
- Kanold, P. O., Nelken, I., and Polley, D. B. (2014). Local versus global scales of organization in auditory cortex. *Trends in Neurosciences*, 37(9):502–510. Publisher: Elsevier Ltd.
- Kawaguchi, Y. (1993). Physiological, morphological, and histochemical characterization of three classes of interneurons in rat neostriatum. *The Journal of Neuroscience*, 13(11):4908–4923.
- Kawaguchi, Y. and Kubota, Y. (1993). Correlation of physiological subgroupings of nonpyramidal cells with parvalbumin- and calbindinD28k-immunoreactive neurons in layer V of rat frontal cortex. *Journal of Neurophysiology*, 70(1):387–396.
- Kawaguchi, Y. and Kubota, Y. (1997). GABAergic cell subtypes and their synaptic connections in rat frontal cortex. *Cerebral Cortex (New York, N.Y.: 1991)*, 7(6):476–486.
- Kawaguchi, Y. and Kubota, Y. (1998). Neurochemical features and synaptic connections of large physiologically-identified GABAergic cells in the rat frontal cortex. *Neuroscience*, 85(3):677–701.
- Kayser, C., Logothetis, N. K., and Panzeri, S. (2010). Visual Enhancement of the Information Representation in Auditory Cortex. *Current Biology*, 20(1):19–24.
- Kayser, C., Petkov, C. I., and Logothetis, N. K. (2009). Multisensory interactions in primate auditory cortex: fMRI and electrophysiology. *Hearing Research*, 258(1-2):80–88.
- Kell, A. J., Yamins, D. L., Shook, E. N., Norman-Haignere, S. V., and McDermott, J. H. (2018). A Task-Optimized Neural Network Replicates Human Auditory Behavior, Predicts Brain Responses, and Reveals a Cortical Processing Hierarchy. *Neuron*, 98(3):630–644.e16.
- Kell, A. J. E. and McDermott, J. H. (2019). Invariance to background noise as a signature of non-primary auditory cortex. *Nature Communications*, 10(1):3958.
- Kelly, J. B., Judge, P. W., and Phillips, D. P. (1986a). Representation of the cochlea in primary auditory cortex of the ferret (*Mustela putorius*). *Hearing Research*, 24(2):111–5.
- Kelly, J. B. and Kavanagh, G. L. (1994). Sound localization after unilateral lesions of inferior colliculus in the ferret (*Mustela putorius*). *Journal of Neurophysiology*, 71(3):1078–1087.

- Kelly, J. B., Kavanagh, G. L., and Dalton, J. C. (1986b). Hearing in the ferret (*Mustela putorius*): Thresholds for pure tone detection. *Hearing Research*, 24(3):269–275. Publisher: Elsevier.
- Kelly, J. B., Rooney, B. J., and Phillips, D. P. (1996). Effects of bilateral auditory cortical lesions on gap-detection thresholds in the ferret (*Mustela putorius*). *Behavioral Neuroscience*, 110(3):542–550.
- Keniston, L. P., Allman, B. L., Meredith, M. A., and Clemo, H. R. (2009). Somatosensory and multisensory properties of the medial bank of the ferret rostral suprasylvian sulcus. *Experimental Brain Research*, 196(2):239–251.
- Kepecs, A. and Fishell, G. (2014). Interneuron cell types are fit to function. *Nature*, 505(7483):318–326.
- Kerlin, A. M., Andermann, M. L., Berezovskii, V. K., and Reid, R. C. (2010). Broadly tuned response properties of diverse inhibitory neuron subtypes in mouse visual cortex. *Neuron*, 67(5):858–871.
- Kerr, J. N. D., de Kock, C. P. J., Greenberg, D. S., Bruno, R. M., Sakmann, B., and Helmchen, F. (2007). Spatial organization of neuronal population responses in layer 2/3 of rat barrel cortex. *Journal of Neuroscience*, 27(48):13316–28.
- Keshishian, M., Akbari, H., Khalighinejad, B., Herrero, J. L., Mehta, A. D., and Mesgarani, N. (2020). Estimating and interpreting nonlinear receptive field of sensory neural responses with deep neural network models. *eLife*, 9:e53445.
- Kettner, R., Feng, J., and Brugge, J. (1985). Postnatal development of the phase-locked response to low frequency tones of auditory nerve fibers in the cat. *The Journal of Neuroscience*, 5(2):275–283.
- Keuroghlian, A. S. and Knudsen, E. I. (2007). Adaptive auditory plasticity in developing and adult animals. *Progress in Neurobiology*, 82(3):109–121.
- Khalighinejad, B., Herrero, J. L., Mehta, A. D., and Mesgarani, N. (2019). Adaptation of the human auditory cortex to changing background noise. *Nature Communications*, 10(1):2509.
- Khan, A. G., Poort, J., Chadwick, A., Blot, A., Sahani, M., Mrsic-Flogel, T. D., and Hofer, S. B. (2018). Distinct learning-induced changes in stimulus selectivity and interactions of GABAergic interneuron classes in visual cortex. *Nature Neuroscience*, 21(6):851–859.
- Kikuchi, Y., Attaheri, A., Wilson, B., Rhone, A. E., Nourski, K. V., Gander, P. E., Kovach, C. K., Kawasaki, H., Griffiths, T. D., Howard, M. A., and Petkov, C. I. (2017). Sequence learning modulates neural responses and oscillatory coupling in human and monkey auditory cortex. *PLoS biology*, 15(4):e2000219.
- Kim, C. K., Adhikari, A., and Deisseroth, K. (2017). Integration of optogenetics with complementary methodologies in systems neuroscience. *Nature Reviews Neuroscience*, 18(4):222–235.
- Kim, D. O., Zahorik, P., Carney, L. H., Bishop, B. B., and Kuwada, S. (2015). Auditory Distance Coding in Rabbit Midbrain Neurons and Human Perception: Monaural Amplitude Modulation Depth as a Cue. *Journal of Neuroscience*, 35(13):5360–5372.

- Kim, H. and Bao, S. (2013). Experience-dependent overrepresentation of ultrasonic vocalization frequencies in the rat primary auditory cortex. *Journal of Neurophysiology*, 110(5):1087–1096.
- King, A. J., Teki, S., and Willmore, B. D. B. (2018). Recent advances in understanding the auditory cortex. *F1000Research*, 7.
- Kinoshita, K., Delcroix, M., Gannot, S., P. Habets, E. A., Haeb-Umbach, R., Kellermann, W., Leutnant, V., Maas, R., Nakatani, T., Raj, B., Sehr, A., and Yoshioka, T. (2016). A summary of the REVERB challenge: state-of-the-art and remaining challenges in reverberant speech processing research. *EURASIP Journal on Advances in Signal Processing*, 2016(1):7.
- Kisvárdy, Z. F., Beaulieu, C., and Eysel, U. T. (1993). Network of GABAergic large basket cells in cat visual cortex (area 18): implication for lateral disinhibition. *The Journal of Comparative Neurology*, 327(3):398–415.
- Kloc, M. and Maffei, A. (2014). Target-Specific Properties of Thalamocortical Synapses onto Layer 4 of Mouse Primary Visual Cortex. *The Journal of Neuroscience*, 34(46):15455.
- Knudsen, V. O. (1929). The hearing of speech in auditoriums. *The Journal of the Acoustical Society of America*, 1(1):56.
- Kobayashi, R., Kurita, S., Kurth, A., Kitano, K., Mizuseki, K., Diesmann, M., Richmond, B. J., and Shinomoto, S. (2019). Reconstructing neuronal circuitry from parallel spike trains. *Nature Communications*, 10(1):4468.
- Koenig, A. H., Allen, J. B., Berkley, D. A., and Curtis, T. H. (1977). Determination of masking-level differences in a reverberant environment. *The Journal of the Acoustical Society of America*, 61(5):1374–1376.
- Kogo, N. and Trengove, C. (2015). Is predictive coding theory articulated enough to be testable? *Frontiers in Computational Neuroscience*, 9.
- Kondo, H. M. and Kashino, M. (2009). Involvement of the Thalamocortical Loop in the Spontaneous Switching of Percepts in Auditory Streaming. *Journal of Neuroscience*, 29(40):12695–12701.
- Kowalski, N., Versnel, H., and Shamma, S. A. (1995). Comparison of responses in the anterior and primary auditory fields of the ferret cortex. *Journal of Neurophysiology*, 73(4):1513–1523.
- Kozlov, A. S. and Gentner, T. Q. (2014). Central auditory neurons display flexible feature recombination functions. *Journal of Neurophysiology*, 111(6):1183–1189.
- Kozlov, A. S. and Gentner, T. Q. (2016). Central auditory neurons have composite receptive fields. *Proceedings of the National Academy of Sciences*, 113(5):1441–1446.
- Kuchibhotla, K. and Bathellier, B. (2018). Neural encoding of sensory and behavioral complexity in the auditory cortex. *Current Opinion in Neurobiology*, 52:65–71.
- Kuchibhotla, K. V., Gill, J. V., Lindsay, G. W., Papadoyannis, E. S., Field, R. E., Sten, T. A. H., Miller, K. D., and Froemke, R. C. (2017). Parallel processing by cortical inhibition enables context-dependent behavior. *Nature Neuroscience*, 20(1):62–71.
- Kudo, M. and Niimi, K. (1980). Ascending projections of the inferior colliculus in the cat: An autoradiographic study. *The Journal of Comparative Neurology*,

- 191(4):545–556.
- Kumar, S., Sedley, W., Nourski, K. V., Kawasaki, H., Oya, H., Patterson, R. D., Howard, M. A., Friston, K. J., and Griffiths, T. D. (2011). Predictive coding and pitch processing in the auditory cortex. *Journal of Cognitive Neuroscience*, 23(10):3084–3094.
- Kuttruff, H. (2017). *Room acoustics*. CRC Press/Taylor & Francis Group, Boca Raton, sixth edition edition.
- Kuttruff, H. (2019). *Room Acoustics*. CRC Press, 6th edition.
- Kuwada, S., Bishop, B., and Kim, D. O. (2012). Approaches to the study of neural coding of sound source location and sound envelope in real environments. *Frontiers in Neural Circuits*, 6.
- Köppl, C. (1997). Phase Locking to High Frequencies in the Auditory Nerve and Cochlear Nucleus Magnocellularis of the Barn Owl, *Tyto alba*. *The Journal of Neuroscience*, 17(9):3312–3321.
- Körding, K. P. and Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, 427(6971):244–247. Number: 6971 Publisher: Nature Publishing Group.
- Kügler, S., Kilic, E., and Bähr, M. (2003). Human synapsin 1 gene promoter confers highly neuron-specific long-term transgene expression from an adenoviral vector in the adult rat brain depending on the transduced area. *Gene Therapy*, 10(4):337–347.
- Langner, G. (1992). Periodicity coding in the auditory system. *Hearing Research*, 60(2):115–142.
- Las, L., Stern, E. A., and Nelken, I. (2005). Representation of tone in fluctuating maskers in the ascending auditory system. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 25(6):1503–1513.
- Lashley, K. S. (1950). *In search of the engram*. Publisher: Academic Press.
- Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.
- Lee, K. S., Huang, X., and Fitzpatrick, D. (2016). Topology of on and off inputs in visual cortex enables an invariant columnar architecture. *Nature*, 533(7601):90–94. Publisher: Nature Publishing Group.
- Lee, S., Hjerling-Leffler, J., Zagha, E., Fishell, G., and Rudy, B. (2010). The Largest Group of Superficial Neocortical GABAergic Interneurons Expresses Ionotropic Serotonin Receptors. *Journal of Neuroscience*, 30(50):16796–16808.
- Lee, S., Kruglikov, I., Huang, Z. J., Fishell, G., and Rudy, B. (2013). A disinhibitory circuit mediates motor integration in the somatosensory cortex. *Nature Neuroscience*, 16(11):1662–1670.
- Li, L.-y., Xiong, X. R., Ibrahim, L. A., Yuan, W., Tao, H. W., and Zhang, L. I. (2015). Differential Receptive Field Properties of Parvalbumin and Somatostatin Inhibitory Neurons in Mouse Auditory Cortex. *Cerebral Cortex*, 25(7):1782–1791.
- Liang, F., Li, H., Chou, X.-l., Zhou, M., Zhang, N. K., Xiao, Z., Zhang, K. K., Tao, H. W., and Zhang, L. I. (2019). Sparse Representation in Awake Auditory Cortex: Cell-type Dependence, Synaptic Mechanisms, Developmental Emergence,

- and Modulation. *Cerebral Cortex*, 29(9):3796–3812.
- Linden, J. F., Liu, R. C., Sahani, M., Schreiner, C. E., and Merzenich, M. M. (2003). Spectrotemporal Structure of Receptive Fields in Areas AI and AAF of Mouse Auditory Cortex. *Journal of Neurophysiology*, 90(4):2660–2675. Publisher: American Physiological Society.
- Linden, J. F. and Schreiner, C. E. (2003). Columnar transformations in auditory cortex? A comparison to visual and somatosensory cortices. *Cerebral Cortex*, 13(1):83–89. ISBN: 1047-3211 (Print) 1047-3211 (Linking).
- Lindsay, G. W. (2020). Convolutional Neural Networks as a Model of the Visual System: Past, Present, and Future. *Journal of Cognitive Neuroscience*, pages 1–15.
- Litovsky, R. Y. and Yin, T. C. T. (1998). Physiological Studies of the Precedence Effect in the Inferior Colliculus of the Cat. I. Correlates of Psychophysics. *Journal of Neurophysiology*, 80(3):1285–1301.
- Loeb, J. (1900). *Comparative physiology of the brain and comparative psychology*. G P Putnam’s Sons, New York, NY, US. Pages: x, 309.
- Lohse, M., Bajo, V. M., King, A. J., and Willmore, B. D. B. (2020a). Neural circuits underlying auditory contrast gain control and their perceptual implications. *Nature Communications*, 11(1):324.
- Lohse, M., Dahmen, J. C., Bajo, V. M., and King, A. J. (2020b). Subcortical Circuits Mediate Communication Between Primary Sensory Cortical Areas. preprint, Neuroscience.
- Lovett-Barron, M., Turi, G. F., Kaifosh, P., Lee, P. H., Bolze, F., Sun, X.-H., Nicoud, J.-F., Zemelman, B. V., Sternson, S. M., and Losonczy, A. (2012). Regulation of neuronal input transformations by tunable dendritic inhibition. *Nature Neuroscience*, 15(3):423–430, S1–3.
- Lu, K. and Vicario, D. S. (2014). Statistical learning of recurring sound patterns encodes auditory objects in songbird forebrain. *Proceedings of the National Academy of Sciences*, 111(40):14553–14558.
- Ludwig, K. A., Miriani, R. M., Langhals, N. B., Joseph, M. D., Anderson, D. J., and Kipke, D. R. (2009). Using a Common Average Reference to Improve Cortical Neuron Recordings From Microelectrode Arrays. *Journal of Neurophysiology*, 101(3):1679–1689.
- MacLean, P. D. (1982). On the Origin and Progressive Evolution of the Triune Brain. In Armstrong, E. and Falk, D., editors, *Primate Brain Evolution: Methods and Concepts*, pages 291–316. Springer US, Boston, MA. Type: 10.1007/978-1-4684-4148-2\_19.
- Malmierca, M. S. (2013). Anatomy and Physiology of the Mammalian Auditory System. In Jaeger, D. and Jung, R., editors, *Encyclopedia of Computational Neuroscience*, pages 1–36. Springer New York, New York, NY.
- Malmierca, M. S., Merchán, M. A., Henkel, C. K., and Oliver, D. L. (2002). Direct projections from cochlear nuclear complex to auditory thalamus in the rat. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 22(24):10891–10897.

- Manger, P. R., Engler, G., Moll, C. K. E., and Engel, A. K. (2005). The anterior ectosylvian visual area of the ferret: a homologue for an enigmatic visual cortical area of the cat? *The European Journal of Neuroscience*, 22(3):706–714.
- Mann, Z. F. and Kelley, M. W. (2011). Development of tonotopy in the auditory periphery. *Hearing Research*, 276(1-2):2–15.
- Maor, I., Shalev, A., and Mizrahi, A. (2016). Distinct Spatiotemporal Response Properties of Excitatory Versus Inhibitory Neurons in the Mouse Auditory Cortex. *Cerebral Cortex*, 26(11):4242–4252.
- Marguet, S. L. and Harris, K. D. (2011). State-Dependent Representation of Amplitude-Modulated Noise Stimuli in Rat Auditory Cortex. *Journal of Neuroscience*, 31(17):6414–6420.
- Marr, D. (1982). *Vision: a computational investigation into the human representation and processing of visual information*. W.H. Freeman, San Francisco.
- May, B. J., Prell, G. S. L., and Sachs, M. B. (1998). Vowel Representations in the Ventral Cochlear Nucleus of the Cat: Effects of Level, Background Noise, and Behavioral State. *Journal of Neurophysiology*, 79(4):1755–1767.
- May, B. J. and Sachs, M. B. (1992). Dynamic range of neural rate responses in the ventral cochlear nucleus of awake cats. *Journal of Neurophysiology*, 68(5):1589–1602.
- McAlpine, D., Jiang, D., and Palmer, A. R. (2001). A neural code for low-frequency sound localization in mammals. *Nature Neuroscience*, 4(4):396–401.
- McCormick, D. A., Connors, B. W., Lighthall, J. W., and Prince, D. A. (1985). Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex. *Journal of Neurophysiology*, 54(4):782–806.
- McMahon, S. M. and Jackson, M. B. (2018). An Inconvenient Truth: Calcium Sensors Are Calcium Buffers. *Trends in Neurosciences*, 41(12):880–884.
- Medrihan, L., Sagi, Y., Inde, Z., Krupa, O., Daniels, C., Peyrache, A., and Greengard, P. (2017). Initiation of Behavioral Response to Antidepressants by Cholecystokinin Neurons of the Dentate Gyrus. *Neuron*, 95(3):564–576.e4.
- Mehaffey, W. H. (2005). Deterministic Multiplicative Gain Control with Active Dendrites. *Journal of Neuroscience*, 25(43):9968–9977.
- Mendoza, G., Peyrache, A., Gámez, J., Prado, L., Buzsáki, G., and Merchant, H. (2016). Recording extracellular neural activity in the behaving monkey using a semichronic and high-density electrode system. *Journal of Neurophysiology*, 116(2):563–574.
- Meng, X., Winkowski, D. E., Kao, J. P., and Kanold, P. O. (2017). Sublaminar subdivision of mouse auditory cortex layer 2/3 based on functional translaminar connections. *Journal of Neuroscience*, 37(42):10200–10214. Publisher: Society for Neuroscience.
- Meredith, M. A., Allman, B. L., Keniston, L. P., and Clemo, H. R. (2009). Auditory influences on non-auditory cortices. *Hearing Research*, 258(1-2):64–71.
- Merzenich, M. M. and Brugge, J. F. (1973). Representation of the cochlear partition on the superior temporal plane of the macaque monkey. *Brain Research*, 50(2):275–296.
- Merzenich, M. M. and Reid, M. D. (1974). Representation of the cochlea within the inferior colliculus of the cat. *Brain Research*, 77(3):397–415.

- Mesgarani, N., David, S. V., Fritz, J. B., and Shamma, S. A. (2008). Phoneme representation and classification in primary auditory cortex. *The Journal of the Acoustical Society of America*, 123(2):899–909.
- Mesgarani, N., David, S. V., Fritz, J. B., and Shamma, S. A. (2009). Influence of context and behavior on stimulus reconstruction from neural activity in primary auditory cortex. *Journal of Neurophysiology*, 102(6):3329–3339.
- Mesgarani, N., David, S. V., Fritz, J. B., and Shamma, S. A. (2014). Mechanisms of noise robust representation of speech in primary auditory cortex. *Proceedings of the National Academy of Sciences*, 111(18):6792–6797.
- Mesik, L., Ma, W.-p., Li, L.-y., Ibrahim, L. A., Huang, Z. J., Zhang, L. I., and Tao, H. W. (2015). Functional response properties of VIP-expressing inhibitory neurons in mouse visual and auditory cortex. *Frontiers in Neural Circuits*, 09.
- Metherate, R., Kaur, S., Kawai, H., Lazar, R., Liang, K., and Rose, H. J. (2005). Spectral integration in auditory cortex: Mechanisms and modulation. *Hearing Research*, 206(1-2):146–158.
- Meyer, H. S., Schwarz, D., Wimmer, V. C., Schmitt, A. C., Kerr, J. N. D., Sakmann, B., and Helmstaedter, M. (2011). Inhibitory interneurons in a cortical column form hot zones of inhibition in layers 2 and 5A. *Proceedings of the National Academy of Sciences*, 108(40):16807.
- Miller, K. D., Pinto, D. J., and Simons, D. J. (2001a). Processing in layer 4 of the neocortical circuit: new insights from visual and somatosensory cortex. *Current Opinion in Neurobiology*, 11(4):488–497.
- Miller, L. M., Escabí, M. A., Read, H. L., and Schreiner, C. E. (2001b). Functional convergence of response properties in the auditory thalamocortical system. *Neuron*, 32(1):151–160.
- Mizrahi, A., Shalev, A., and Nelken, I. (2014). Single neuron and population coding of natural sounds in auditory cortex. *Current Opinion in Neurobiology*, 24(1):103–110. ISBN: 1873-6882 (Electronic) 0959-4388 (Linking).
- Moore, A. K. and Wehr, M. (2013). Parvalbumin-Expressing Inhibitory Interneurons in Auditory Cortex Are Well-Tuned for Frequency. *Journal of Neuroscience*, 33(34):13713–13723.
- Moore, R. C., Lee, T., and Theunissen, F. E. (2013). Noise-invariant Neurons in the Avian Auditory Cortex: Hearing the Song in Noise. *PLoS Computational Biology*, 9(3):e1002942.
- Natan, R. G., Briguglio, J. J., Mwilambwe-Tshilobo, L., Jones, S. I., Aizenberg, M., Goldberg, E. M., and Geffen, M. N. (2015). Complementary control of sensory adaptation by two types of cortical interneurons. *eLife*, 4:e09868.
- Natan, R. G., Rao, W., and Geffen, M. N. (2017). Cortical Interneurons Differentially Shape Frequency Tuning following Adaptation. *Cell Reports*, 21(4):878–890.
- Nayagam, D. A. X., Clarey, J. C., and Paolini, A. G. (2005). Powerful, Onset Inhibition in the Ventral Nucleus of the Lateral Lemniscus. *Journal of Neurophysiology*, 94(2):1651–1654.

- Nelken, I., Bizley, J., Shamma, S. A., and Wang, X. (2014). Auditory Cortical Processing in Real-World Listening: The Auditory System Going Real. *Journal of Neuroscience*, 34(46):15135–15138.
- Nelken, I., Bizley, J. K., Nodal, F. R., Ahmed, B., King, A. J., and Schnupp, J. W. (2008). Responses of auditory cortex to complex stimuli: functional organization revealed using intrinsic optical signals. *Journal of Neurophysiology*, 99(4):1928–1941.
- Nelken, I., Bizley, J. K., Nodal, F. R., Ahmed, B., Schnupp, J. W., and King, A. J. (2004). Large-scale organization of ferret auditory cortex revealed using continuous acquisition of intrinsic optical signals. *Journal of Neurophysiology*, 92(4):2574–2588.
- Nelken, I., Fishbach, A., Las, L., Ulanovsky, N., and Farkas, D. (2003). Primary auditory cortex of cats: feature detection or something else? *Biological Cybernetics*, 89(5):397–406.
- Nelson, A. and Mooney, R. (2016). The Basal Forebrain and Motor Cortex Provide Convergent yet Distinct Movement-Related Inputs to the Auditory Cortex. *Neuron*, 90(3):635–648.
- Niell, C. M. and Stryker, M. P. (2008). Highly Selective Receptive Fields in Mouse Visual Cortex. *Journal of Neuroscience*, 28(30):7520–7536.
- Nielsen, J. B. and Dau, T. (2010). Revisiting perceptual compensation for effects of reverberation in speech identification. *The Journal of the Acoustical Society of America*, 128(5):3088–3094.
- Nishiura, T., Hirano, Y., Denda, Y., and Nakayama, M. (2007). Investigations into early and late reflections on distant-talking speech recognition toward suitable reverberation criteria. In *INTERSPEECH*.
- Niwa, M., Johnson, J. S., O’Connor, K. N., and Sutter, M. L. (2012). Activity related to perceptual judgment and action in primary auditory cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 32(9):3193–3210.
- Noda, T. and Takahashi, H. (2015). Anesthetic effects of isoflurane on the tonotopic map and neuronal population activity in the rat auditory cortex. *The European Journal of Neuroscience*, 42(6):2298–2311.
- Nodal, F. R., Bajo, V. M., and King, A. J. (2012). Plasticity of spatial hearing: behavioural effects of cortical inactivation. *The Journal of Physiology*, 590(16):3965–3986.
- Nodal, F. R., Bajo, V. M., Parsons, C. H., Schnupp, J. W., and King, A. J. (2008). Sound localization behavior in ferrets: comparison of acoustic orientation and approach-to-target responses. *Neuroscience*, 154(1):397–408.
- Nodal, F. R. and King, A. J. (2014). Hearing and Auditory Function in Ferrets. In Fox, J. G. and Marini, R. P., editors, *Biology and Diseases of the Ferret*, pages 685–710. John Wiley & Sons, Inc., Ames, USA.
- Norman-Haignere, S. V. and McDermott, J. H. (2018). Neural responses to natural and model-matched stimuli reveal distinct computations in primary and nonprimary auditory cortex. *PLOS Biology*, 16(12):e2005127.
- Nábélek, A. K. (1978). Influence of the precedence effect on word identification by normally hearing and hearing-impaired subjects. *The Journal of the Acoustical*

- Society of America*, 63(1):187.
- Nábělek, A. K., Letowski, T. R., and Tucker, F. M. (1989). Reverberant overlap- and self-masking in consonant identification. *The Journal of the Acoustical Society of America*, 86(4):1259–65.
- Oakley, D. A. (1981). Brain Mechanisms of Mammalian Memory. *British Medical Bulletin*, 37(2):175–180.
- Ohki, K., Chung, S., Ch'ng, Y. H., Kara, P., and Reid, R. C. (2005). Functional imaging with cellular resolution reveals precise microarchitecture in visual cortex. *Nature*, 433(7026):597–603.
- Olsen, T. (2020). *Current source density (CSD)*. Publication Title: MATLAB Central File Exchange.
- Olshausen, B. A. and Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607–609. Number: 6583 Publisher: Nature Publishing Group.
- Oviedo, H. V. (2017). Connectivity motifs of inhibitory neurons in the mouse Auditory Cortex. *Scientific Reports*, 7(1):16987.
- Oviedo, H. V., Bureau, I., Svoboda, K., and Zador, A. M. (2010). The functional asymmetry of auditory cortex is reflected in the organization of local cortical circuits. *Nature Neuroscience*, 13(11):1413–1420.
- Pachitariu, M., Steinmetz, N., Kadir, S., Carandini, M., and Harris, K. D. (2016). Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels. *bioRxiv*, page 061481.
- Pachitariu, M., Stringer, C., and Harris, K. D. (2018). Robustness of spike deconvolution for neuronal calcium imaging. *Journal of Neuroscience*, 38(37):7976–7985. Publisher: Society for Neuroscience.
- Palmer, A. and Russell, I. (1986). Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells. *Hearing Research*, 24(1):1–15.
- Panniello, M., King, A., Dahmen, J., and Walker, K. (2018). Local and global spatial organization of interaural level difference and frequency preferences in auditory cortex. *Cerebral Cortex*, 28(1).
- Parras, G. G., Nieto-Diego, J., Carbajal, G. V., Valdés-Baizabal, C., Escera, C., and Malmierca, M. S. (2017). Neurons along the auditory pathway exhibit a hierarchical organization of prediction error. *Nature Communications*, 8(1):2148.
- Pavlov, I. P. (1927). *Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex*. Oxford Univ. Press, Oxford, England. Pages: xv, 430.
- Pecka, M., Zahn, T. P., Saunier-Rebori, B., Siveke, I., Felmy, F., Wiegrebe, L., Klug, A., Pollak, G. D., and Grothe, B. (2007). Inhibiting the Inhibition: A Neuronal Network for Sound Localization in Reverberant Environments. *Journal of Neuroscience*, 27(7):1782–1790.
- Pettersen, K. H., Devor, A., Ulbert, I., Dale, A. M., and Einevoll, G. T. (2006). Current-source density estimation based on inversion of electrostatic forward solution: effects of finite extent of neuronal activity and conductivity discontinuities. *Journal of*

- Neuroscience Methods*, 154(1-2):116–33.
- Peyrache, A., Dehghani, N., Eskandar, E. N., Madsen, J. R., Anderson, W. S., Donoghue, J. A., Hochberg, L. R., Halgren, E., Cash, S. S., and Destexhe, A. (2012). Spatiotemporal dynamics of neocortical excitation and inhibition during human sleep. *Proceedings of the National Academy of Sciences*, 109(5):1731–1736.
- Peyrache, A. and Destexhe, A. (2019). Electrophysiological monitoring of inhibition in mammalian species, from rodents to humans. *Neurobiology of Disease*, 130:104500.
- Phillips, C. G., Zeki, S., and Barlow, H. B. (1984). Localization of Function in the Cerebral Cortex: Past, Present and Future. *Brain*, 107(1):328–361.
- Phillips, D. and Irvine, D. (1981). Responses of single neurons in physiologically defined area AI of cat cerebral cortex: sensitivity to interaural intensity differences. *Hearing Research*, 4(3):299–307.
- Phillips, E. A. and Hasenstaub, A. R. (2016). Asymmetric effects of activating and inactivating cortical interneurons. *eLife*, 5:e18383.
- Pi, H.-J., Hangya, B., Kvitsiani, D., Sanders, J. I., Huang, Z. J., and Kepecs, A. (2013). Cortical interneurons that specialize in disinhibitory control. *Nature*, 503(7477):521–524.
- Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E. J., and Simoncelli, E. P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207):995–999.
- Poissant, S. F., Whitmal, N. A., and Freyman, R. L. (2006). Effects of reverberation and masking on speech intelligibility in cochlear implant simulations. *The Journal of the Acoustical Society of America*, 119(3):1606–15.
- Polley, D. B., Read, H. L., Storace, D. A., and Merzenich, M. M. (2007). Multiparametric auditory receptive field organization across five cortical fields in the albino rat. *Journal of Neurophysiology*, 97(5):3621–3638.
- Polterovich, A., Jankowski, M. M., and Nelken, I. (2018). Deviance sensitivity in the auditory cortex of freely moving rats. *PLOS ONE*, 13(6):e0197678.
- Portfors, C. V., Mayko, Z. M., Jonson, K., Cha, G. F., and Roberts, P. D. (2011). Spatial organization of receptive fields in the auditory midbrain of awake mouse. *Neuroscience*, 193:429–439.
- Poulet, J. F. A. and Crochet, S. (2019). The Cortical States of Wakefulness. *Frontiers in Systems Neuroscience*, 12:64.
- Pressnitzer, D., Sayles, M., Micheyl, C., and Winter, I. M. (2008). Perceptual organization of sound begins in the auditory periphery. *Current biology: CB*, 18(15):1124–1128.
- Purves, D., Augustine, G. J., Fitzpatrick, D., Hall, W. C., LaMantia, A.-S., McNamara, J. O., and Williams, S. M., editors (2004). *Neuroscience*, 3rd ed. Neuroscience, 3rd ed. Sinauer Associates, Sunderland, MA, US. Pages: xix, 773.
- Puvvada, K. C. and Simon, J. Z. (2017). Cortical Representations of Speech in a Multitalker Auditory Scene. *The Journal of Neuroscience*, 37(38):9189–9196.
- Puvvada, K. C., Villafañe-Delgado, M., Brodbeck, C., and Simon, J. Z. (2017). Neural Coding of Noisy and Reverberant Speech in Human Auditory Cortex. preprint,

- Neuroscience.
- Qin, M. K. and Oxenham, A. J. (2005). Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification. *Ear and hearing*, 26(5):451–60.
- Rabinowitz, N., Willmore, B., Schnupp, J., and King, A. (2011). Contrast Gain Control in Auditory Cortex. *Neuron*, 70(6):1178–1191. Publisher: Elsevier Inc.
- Rabinowitz, N. C., Willmore, B. D. B., King, A. J., and Schnupp, J. W. H. (2013). Constructing Noise-Invariant Representations of Sound in the Auditory Pathway. *PLoS Biology*, 11(11):e1001710.
- Rabinowitz, N. C., Willmore, B. D. B., Schnupp, J. W. H., and King, A. J. (2012). Spectrotemporal Contrast Kernels for Neurons in Primary Auditory Cortex. *Journal of Neuroscience*, 32(33):11271–11284.
- Rahman, M., Willmore, B. D. B., King, A. J., and Harper, N. S. (2019). A dynamic network model of temporal receptive fields in primary auditory cortex. *PLOS Computational Biology*, 15(5):e1006618. Publisher: Public Library of Science.
- Rahman, M., Willmore, B. D. B., King, A. J., and Harper, N. S. (2020). Simple transformations capture auditory input to cortex. *Proceedings of the National Academy of Sciences*, 117(45):28442–28451.
- Rakerd, B. and Hartmann, W. M. (2005). Localization of noise in a reverberant environment. In Pressnitzer, D., de Cheveigné, A., McAdams, S., and Collet, L., editors, *Auditory Signal Processing*, pages 413–421. Springer New York, New York, NY.
- Ramsay, A. M. and Meredith, M. A. (2004). Multiple sensory afferents to ferret pseudosylvian sulcal cortex. *Neuroreport*, 15(3):461–465.
- Rao, R. P. N. and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1):79–87.
- Rauschecker, J. P. (2018). Where, When, and How: Are they all sensorimotor? Towards a unified view of the dorsal pathway in vision and audition. *Cortex*, 98:262–268.
- Read, H. L., Winer, J. A., and Schreiner, C. E. (2002). Functional architecture of auditory cortex. *Current Opinion in Neurobiology*, 12(4):433–440.
- Recanzone, G. H., Schreiner, C. E., Sutter, M. L., Beitel, R. E., and Merzenich, M. M. (1999). Functional organization of spectral receptive fields in the primary auditory cortex of the owl monkey. *Journal of Comparative Neurology*, 415(4):460–481. Publisher: John Wiley & Sons, Ltd.
- Robinson, B. L., Harper, N. S., and McAlpine, D. (2016). Meta-adaptation in the auditory midbrain under cortical influence. *Nature Communications*, 7(1):13442.
- Rodriguez, M. Z., Comin, C. H., Casanova, D., Bruno, O. M., Amancio, D. R., Costa, L. d. F., and Rodrigues, F. A. (2019). Clustering algorithms: A comparative approach. *PLOS ONE*, 14(1):e0210236.
- Roe, A. W., Pallas, S. L., Kwon, Y. H., and Sur, M. (1992). Visual projections routed to the auditory pathway in ferrets: receptive fields of visual neurons in primary auditory cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 12(9):3651–3664.

- Romani, G. L., Williamson, S. J., and Kaufman, L. (1982). Tonotopic organization of the human auditory cortex. *Science*, 216(4552):1339–40. Publisher: American Association for the Advancement of Science.
- Romero, S., Hight, A. E., Clayton, K. K., Resnik, J., Williamson, R. S., Hancock, K. E., and Polley, D. B. (2019). Cellular and widefield imaging of sound frequency organization in primary and higher-order fields of the mouse auditory cortex. *Cerebral Cortex*, page Epub.
- Rothschild, G., Nelken, I., and Mizrahi, A. (2010). Functional organization and population dynamics in the mouse primary auditory cortex. *Nature Neuroscience*, 13(3):353–360. Publisher: Nature Publishing Group.
- Roux, L. and Buzsáki, G. (2015). Tasks for inhibitory interneurons in intact brain circuits. *Neuropharmacology*, 88:10 – 23.
- Roux, L., Stark, E., Sjulson, L., and Buzsáki, G. (2014). In vivo optogenetic identification and manipulation of GABAergic interneuron subtypes. *Current Opinion in Neurobiology*, 26:88–95.
- Royer, S., Zemelman, B. V., Losonczy, A., Kim, J., Chance, F., Magee, J. C., and Buzsáki, G. (2012). Control of timing, rate and bursts of hippocampal place cells by dendritic and somatic inhibition. *Nature Neuroscience*, 15(5):769–775.
- Rubel, E. W. and Fritsch, B. (2002). Auditory System Development: Primary Auditory Neurons and Their Targets. *Annual Review of Neuroscience*, 25(1):51–101.
- Rubin, J., Ulanovsky, N., Nelken, I., and Tishby, N. (2016). The Representation of Prediction Error in Auditory Cortex. *PLOS Computational Biology*, 12(8):e1005058.
- Rudy, B., Fishell, G., Lee, S., and Hjerling-Leffler, J. (2011). Three groups of interneurons account for nearly 100% of neocortical GABAergic neurons. *Developmental Neurobiology*, 71(1):45–61.
- Runyan, C. A., Piasini, E., Panzeri, S., and Harvey, C. D. (2017). Distinct timescales of population coding across cortex. *Nature*, 548(7665):92–96.
- Russell, J. T. (2011). Imaging calcium signals in vivo: a powerful tool in physiology and pharmacology. *British Journal of Pharmacology*, 163(8):1605–1625.
- Sadagopan, S. and Wang, X. (2009). Nonlinear spectrotemporal interactions underlying selectivity for complex sounds in auditory cortex. *Journal of Neuroscience*, 29(36):11192–202. Publisher: Society for Neuroscience.
- Sahani, M. and Linden, J. F. (2003). How Linear are Auditory Cortical Responses? In Becker, S., Thrun, S., and Obermayer, K., editors, *Advances in Neural Information Processing Systems 15*, pages 125–132. MIT Press.
- Sakai, H., Sato, S.-i., and Ando, Y. (1998). Orthogonal acoustical factors of sound fields in a forest compared with those in a concert hall. *The Journal of the Acoustical Society of America*, 104(3):1491–1497.
- Sakata, S. and Harris, K. D. (2012). Laminar-dependent effects of cortical state on auditory cortical spontaneous activity. *Frontiers in Neural Circuits*, 6.
- Sally, S. L. and Kelly, J. B. (1988). Organization of auditory cortex in the albino rat: sound frequency. *Journal of Neurophysiology*, 59(5):1627–1638.

- Sanes, D. (1990). An in vitro analysis of sound localization mechanisms in the gerbil lateral superior olive. *The Journal of Neuroscience*, 10(11):3494–3506.
- Sanes, D. and Constantine-Paton, M. (1983). Altered activity patterns during development reduce neural tuning. *Science*, 221(4616):1183–1185.
- Sato, T. R., Gray, N. W., Mainen, Z. F., and Svoboda, K. (2007). The functional microarchitecture of the mouse barrel cortex. *PLoS Biology*, 5(7):e189. Publisher: Public Library of Science.
- Sayles, M. and Winter, I. M. (2008). Reverberation Challenges the Temporal Representation of the Pitch of Complex Sounds. *Neuron*, 58(5):789–801.
- Schiff, M. L. and Reyes, A. D. (2012). Characterization of thalamocortical responses of regular-spiking and fast-spiking neurons of the mouse auditory cortex in vitro and in silico. *Journal of Neurophysiology*, 107(5):1476–1488.
- Schnupp, J., Nelken, I., and King, A. (2011). *Auditory neuroscience: Making sense of sound*. Auditory neuroscience: Making sense of sound. MIT Press, Cambridge, MA, US. Pages: x, 356.
- Schnupp, J. W., Mrsic-Flogel, T. D., and King, A. J. (2001). Linear processing of spatial cues in primary auditory cortex. *Nature*, 414(6860):200–204.
- Schofield, B. R., Motts, S. D., Mellott, J. G., and Foster, N. L. (2014). Projections from the dorsal and ventral cochlear nuclei to the medial geniculate body. *Frontiers in Neuroanatomy*, 8:10.
- Scholl, B., Pattadkal, J., Dilly, G., Priebe, N., and Zemelman, B. (2015). Local Integration Accounts for Weak Selectivity of Mouse Neocortical Parvalbumin Interneurons. *Neuron*, 87(2):424–436.
- Schreiner, C. E. and Winer, J. A. (2007). Auditory cortex mapmaking: principles, projections, and plasticity. *Neuron*, 56(2):356–365.
- Schroeder, M. R. (1961). Natural Sounding Artificial Reverberation. In *Audio Engineering Society Convention 13*.
- Schroeder, M. R. (1962). Frequency-Correlation Functions of Frequency Responses in Rooms. *The Journal of the Acoustical Society of America*, 34(12):1819–1823.
- Schroeder, M. R. (1965). New Method of Measuring Reverberation Time. *The Journal of the Acoustical Society of America*, 37(3):409–412.
- Schwartz, O. and Simoncelli, E. P. (2001). Natural signal statistics and sensory gain control. *Nature Neuroscience*, 4(8):819–825.
- Schweitzer, H. C. (2003). Reducing the Negative Effects of Reverberation in Hearing Aid Processing. pages 1–5.
- Sedley, W., Gander, P. E., Kumar, S., Kovach, C. K., Oya, H., Kawasaki, H., Howard, M. A., and Griffiths, T. D. (2016). Neural signatures of perceptual inference. *eLife*, 5:e11476.
- Seybold, B., Phillips, E., Schreiner, C., and Hasenstaub, A. (2015). Inhibitory Actions Unified by Network Integration. *Neuron*, 87(6):1181–1192.
- Shamma, S. A. (2000). Physiological basis of timbre perception. In Gazzaniga, M. S., editor, *The new cognitive neurosciences*, pages 411–23. MIT Press, Cambridge, Mass.

- Shamma, S. A., Elhilali, M., and Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends in Neurosciences*, 34(3):114–123.
- Sharma, J., Angelucci, A., and Sur, M. (2000). Induction of visual orientation modules in auditory cortex. *Nature*, 404(6780):841–847.
- Sharpee, T., Rust, N. C., and Bialek, W. (2004). Analyzing neural responses to natural signals: maximally informative dimensions. *Neural Computation*, 16(2):223–250.
- Sheikhhattar, A., Miran, S., Liu, J., Fritz, J. B., Shamma, S. A., Kanold, P. O., and Babadi, B. (2018). Extracting neuronal functional network dynamics via adaptive Granger causality analysis. *Proceedings of the National Academy of Sciences*, 115(17):E3869–E3878.
- Shepherd, G. M. G. and Svoboda, K. (2005). Laminar and columnar organization of ascending excitatory projections to layer 2/3 pyramidal neurons in rat barrel cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 25(24):5670–5679.
- Sherman, S. M. and Guillery, R. W. (2002). The role of the thalamus in the flow of information to the cortex. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 357(1428):1695–1708.
- Shinn-Cunningham, B. (2000). Learning Reverberation: Considerations for Spatial Auditory Displays. *Proceedings of the 2000 International Conference on Auditory Display*, (April):126–134.
- Shinn-Cunningham, B., Desloge, J., and Kopco, N. (2001). Empirical and modeled acoustic transfer functions in a simple room: effects of distance and direction. In *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No.01TH8575)*, pages 183–186, New Platz, NY, USA. IEEE.
- Shinn-Cunningham, B. and Kawakyu, K. (2003). Neural representation of source direction in reverberant space. In *2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (IEEE Cat. No.03TH8684)*, pages 79–82, New Paltz, NY, USA. IEEE.
- Shinn-Cunningham, B. G., Lin, I.-F., and Streeter, T. (2005). Trading Directional Accuracy for Realism in a Virtual Auditory Display.
- Shoham, S., O’Connor, D. H., and Segev, R. (2006). How silent is the brain: is there a “dark matter” problem in neuroscience? *Journal of Comparative Physiology A*, 192(8):777–784.
- Silberberg, G. and Markram, H. (2007). Disynaptic Inhibition between Neocortical Pyramidal Cells Mediated by Martinotti Cells. *Neuron*, 53(5):735–746.
- Silver, R. A. (2010). Neuronal arithmetic. *Nature Reviews Neuroscience*, 11(7):474–489.
- Singer, Y., Teramoto, Y., Willmore, B. D., Schnupp, J. W., King, A. J., and Harper, N. S. (2018). Sensory cortex is optimized for prediction of future input. *eLife*, 7.
- Sirota, A., Montgomery, S., Fujisawa, S., Isomura, Y., Zugaro, M., and Buzsáki, G. (2008). Entrainment of Neocortical Neurons and Gamma Oscillations by the Hippocampal Theta Rhythm. *Neuron*, 60(4):683–697.

- Slama, M. C. C. and Delgutte, B. (2015). Neural Coding of Sound Envelope in Reverberant Environments. *Journal of Neuroscience*, 35(10):4452–4468.
- Smith, E. C. and Lewicki, M. S. (2006). Efficient auditory coding. *Nature*, 439(7079):978–982.
- Smith, S. L. and Häusser, M. (2010). Parallel processing of visual space by neighboring neurons in mouse visual cortex. *Nature Neuroscience*, 13(9):1144–1149.
- Snyder, A. C., Morais, M. J., and Smith, M. A. (2016). Dynamics of excitatory and inhibitory networks are differentially altered by selective attention. *Journal of Neurophysiology*, 116(4):1807–1820.
- Somogyi, P., Kisvárdy, Z. F., Martin, K. A., and Whitteridge, D. (1983). Synaptic connections of morphologically identified and physiologically characterized large basket cells in the striate cortex of cat. *Neuroscience*, 10(2):261–294.
- South, D. A. and Weinberger, N. M. (1995). A comparison of tone-evoked response properties of 'cluster' recordings and their constituent single cells in the auditory cortex. *Brain research*, 704(2):275–88.
- Spitzer, M. W., Bala, A. D. S., and Takahashi, T. T. (2004). A Neuronal Correlate of the Precedence Effect Is Associated With Spatial Selectivity in the Barn Owl's Auditory Midbrain. *Journal of Neurophysiology*, 92(4):2051–2070.
- Spratling, M. (2017). A review of predictive coding algorithms. *Brain and Cognition*, 112:92–97.
- Stabler, S. E., Palmer, A. R., and Winter, I. M. (1996). Temporal and mean rate discharge patterns of single units in the dorsal cochlear nucleus of the anesthetized guinea pig. *Journal of Neurophysiology*, 76(3):1667–1688.
- Stark, E., Eichler, R., Roux, L., Fujisawa, S., Rotstein, H., and Buzsáki, G. (2013). Inhibition-Induced Theta Resonance in Cortical Circuits. *Neuron*, 80(5):1263–1276.
- Steffen, H., Simonis, C., Thomas, H., Tillein, J., and Scheich, H. (1988). Auditory cortex: Multiple fields, their architectonics and connections in the Mongolian Gerbil. In *Auditory Pathway*, pages 223–228. Springer US, Boston, MA.
- Steinmetz, N. A., Aydin, C., Lebedeva, A., Okun, M., Pachitariu, M., Bauza, M., Beau, M., Bhagat, J., Böhm, C., Broux, M., Chen, S., Colonell, J., Gardner, R. J., Karsh, B., Kostadinov, D., Mora-Lopez, C., Park, J., Putzeys, J., Sauerbrei, B., van Daal, R. J. J., Vollan, A. Z., Welkenhuysen, M., Ye, Z., Dudman, J., Dutta, B., Hantman, A. W., Harris, K. D., Lee, A. K., Moser, E. I., O'Keefe, J., Renart, A., Svoboda, K., Häusser, M., Haesler, S., Carandini, M., and Harris, T. D. (2020). Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings. preprint, *Neuroscience*.
- Stiebler, I., Neulist, R., Fichtel, I., and Ehret, G. (1997). The auditory cortex of the house mouse: left-right differences, tonotopic organization and quantitative analysis of frequency representation. *Journal of Comparative Physiology A: Sensory, Neural, and Behavioral Physiology*, 181(6):559–571.
- Sturgill, J. F. and Isaacson, J. S. (2015). Somatostatin cells regulate sensory response fidelity via subtractive inhibition in olfactory cortex. *Nature Neuroscience*, 18(4):531–535.

- Sumner, C. J. and Palmer, A. R. (2012). Auditory nerve fibre responses in the ferret: Auditory nerve fibre responses in ferrets. *European Journal of Neuroscience*, 36(4):2428–2439.
- Sur, M., Garraghty, P., and Roe, A. (1988). Experimentally induced visual projections into auditory thalamus and cortex. *Science*, 242(4884):1437–1441.
- Sutter, M. L. and Schreiner, C. E. (1991). Physiology and topography of neurons with multi-peaked tuning curves in cat primary auditory cortex. *Journal of Neurophysiology*, 65(5):1207–1226.
- Szymanski, F. D., Garcia-Lazaro, J. A., and Schnupp, J. W. H. (2009). Current Source Density Profiles of Stimulus-Specific Adaptation in Rat Auditory Cortex. *Journal of Neurophysiology*, 102(3):1483–1490.
- Szymanski, F. D., Rabinowitz, N. C., Magri, C., Panzeri, S., and Schnupp, J. W. H. (2011). The laminar and temporal structure of stimulus information in the phase of field potentials of auditory cortex. *Journal of Neuroscience*, 31(44):15787–15801.
- Tamás, G., Buhl, E. H., and Somogyi, P. (1997). Fast IPSPs elicited via multiple synaptic release sites by different types of GABAergic neurone in the cat visual cortex. *The Journal of Physiology*, 500 ( Pt 3):715–738.
- Tasic, B., Menon, V., Nguyen, T. N., Kim, T. K., Jarsky, T., Yao, Z., Levi, B., Gray, L. T., Sorensen, S. A., Dolbeare, T., Bertagnolli, D., Goldy, J., Shapovalova, N., Parry, S., Lee, C., Smith, K., Bernard, A., Madisen, L., Sunkin, S. M., Hawrylycz, M., Koch, C., and Zeng, H. (2016). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nature Neuroscience*, 19(2):335–346.
- Teki, S., Barascud, N., Picard, S., Payne, C., Griffiths, T. D., and Chait, M. (2016). Neural Correlates of Auditory Figure-Ground Segregation Based on Temporal Coherence. *Cerebral Cortex*, 26(9):3669–3680.
- Teki, S. and Griffiths, T. D. (2016). Brain Bases of Working Memory for Time Intervals in Rhythmic Sequences. *Frontiers in Neuroscience*, 10.
- Terashima, H. and Okada, M. (2012). The topographic unsupervised learning of natural sounds in the auditory cortex. *Advances in Neural Information Processing Systems*, pages 1–9. ISBN: 9781627480031.
- Theunissen, F. E., David, S. V., Singh, N. C., Hsu, A., Vinje, W. E., and Gallant, J. L. (2001). Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. *Network: Computation in Neural Systems*, 12(3):289–316. Publisher: Taylor & Francis \_eprint: <https://doi.org/10.1080/net.12.3.289.316>.
- Theunissen, F. E. and Elie, J. E. (2014). Neural processing of natural sounds. *Nature Reviews. Neuroscience*, 15(6):355–366.
- Thompson, J. A. F. (2020). *Characterizing and comparing acoustic representations in convolutional neural networks and the human auditory system*. PhD Thesis, Université de Montréal.
- Tischbirek, C. H., Noda, T., Tohmi, M., Birkner, A., Nelken, I., and Konnerth, A. (2019). In vivo functional mapping of a cortical column at single-neuron resolution. *Cell Reports*.

- Tollin, D. J., Populin, L. C., and Yin, T. C. T. (2004). Neural Correlates of the Precedence Effect in the Inferior Colliculus of Behaving Cats. *Journal of Neurophysiology*, 92(6):3286–3297.
- Toyama, K., Matsunami, K., Ono, T., and Tokashiki, S. (1974). An intracellular study of neuronal organization in the visual cortex. *Experimental Brain Research*, 21(1):45–66.
- Traer, J. and McDermott, J. H. (2016). Statistics of natural reverberation enable perceptual separation of sound and space. *Proceedings of the National Academy of Sciences*, 113(48):E7856–E7865.
- Tremblay, R., Lee, S., and Rudy, B. (2016). GABAergic Interneurons in the Neocortex: From Cellular Properties to Circuits. *Neuron*, 91(2):260–292.
- Trivedi, U., Dieckman, E., and Xiang, N. (2009). Reciprocal maximum-length and related sequences in the generation of natural, spatial sounding reverberation. *The Journal of the Acoustical Society of America*, 125(4):2735–2735.
- Trussell, L. O. (1999). Synaptic mechanisms for coding timing in auditory neurons. *Annual Review of Physiology*, 61(1):477–496.
- Tsunada, J., Liu, A. S. K., Gold, J. I., and Cohen, Y. E. (2016). Causal contribution of primate auditory cortex to auditory perceptual decision-making. *Nature Neuroscience*, 19(1):135–142.
- Turner, J. G., Hughes, L. F., and Caspary, D. M. (2005). Divergent response properties of layer-V neurons in rat primary auditory cortex. *Hearing Research*.
- Turner, R. E. (2010). *Statistical Models for Natural Sounds*. PhD Thesis, Gatsby Computational Neuroscience Unit, UCL.
- Vasquez-Lopez, S. A., Weissenberger, Y., Lohse, M., Keating, P., King, A. J., and Dahmen, J. C. (2017). Thalamic input to auditory cortex is locally heterogeneous but globally tonotopic. *eLife*, 6:1–18.
- Vu, E. and Krasne, F. (1992). Evidence for a computational distinction between proximal and distal neuronal inhibition. *Science*, 255(5052):1710–1712.
- Walker, K. M. M., Bizley, J. K., King, A. J., and Schnupp, J. W. H. (2011). Multiplexed and Robust Representations of Sound Features in Auditory Cortex. *Journal of Neuroscience*, 31(41):14565–14576.
- Walker, K. M. M., Schnupp, J. W. H., Hart-Schnupp, S. M. B., King, A. J., and Bizley, J. K. (2009). Pitch discrimination by ferrets for simple and complex sounds. *The Journal of the Acoustical Society of America*, 126(3):1321–1335.
- Wang, X. (2013). The harmonic organization of auditory cortex. *Frontiers in Systems Neuroscience*, 7.
- Wang, Y., Toledo-Rodriguez, M., Gupta, A., Wu, C., Silberberg, G., Luo, J., and Markram, H. (2004). Anatomical, physiological and molecular properties of Martinotti cells in the somatosensory cortex of the juvenile rat. *The Journal of Physiology*, 561(Pt 1):65–90.
- Wehr, M. and Zador, A. M. (2003). Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. *Nature*, 426(6965):442–446.

- Weinberger, N. M. (2011). The medial geniculate, not the amygdala, as the root of auditory fear conditioning. *Hearing Research*, 274(1-2):61–74.
- Werthat, F., Alexandrova, O., Grothe, B., and Koch, U. (2008). Experience-dependent refinement of the inhibitory axons projecting to the medial superior olive. *Developmental Neurobiology*, 68(13):1454–1462.
- Wever, E. G. and Bray, C. W. (1937). The Perception of Low Tones and the Resonance-Volley Theory. *The Journal of Psychology*, 3(1):101–114.
- Williamson, R. S., Ahrens, M. B., Linden, J. F., and Sahani, M. (2016). Input-Specific Gain Modulation by Local Sensory Context Shapes Cortical and Thalamic Responses to Complex Sounds. *Neuron*, 91(2):467–481.
- Willmore, B. D. and King, A. (2009). Auditory Cortex: Representation through Sparsification? *Current Biology*, 19(24):R1123–R1125.
- Willmore, B. D. B., Schoppe, O., King, A. J., Schnupp, J. W. H., and Harper, N. S. (2016). Incorporating Midbrain Adaptation to Mean Sound Level Improves Models of Auditory Cortical Processing. *Journal of Neuroscience*, 36(2):280–289.
- Wilson, B., Kikuchi, Y., Sun, L., Hunter, D., Dick, F., Smith, K., Thiele, A., Griffiths, T. D., Marslen-Wilson, W. D., and Petkov, C. I. (2015). Auditory sequence processing reveals evolutionarily conserved regions of frontal cortex in macaques and humans. *Nature Communications*, 6(1):8901.
- Wilson, D. E., Scholl, B., and Fitzpatrick, D. (2018). Differential tuning of excitation and inhibition shapes direction selectivity in ferret visual cortex. *Nature*, 560(7716):97–101.
- Wilson, D. E., Smith, G. B., Jacob, A. L., Walker, T., Dimidschstein, J., Fishell, G., and Fitzpatrick, D. (2017). GABAergic neurons in ferret visual cortex participate in functionally specific networks. *Neuron*, 93(5):1058–1065.e4. Publisher: Elsevier Inc.
- Wilson, N. R., Runyan, C. A., Wang, F. L., and Sur, M. (2012). Division and subtraction by distinct cortical inhibitory networks in vivo. *Nature*, 488(7411):343–348.
- Winer, J. A., Miller, L. M., Lee, C. C., and Schreiner, C. E. (2005). Auditory thalamocortical transformation: structure and function. *Trends in Neurosciences*, 28(5):255–263.
- Winkowski, D. E. and Kanold, P. O. (2013). Laminar transformation of frequency organization in auditory cortex. *Journal of Neuroscience*, 33(4):1498–1508.
- Wood, K. C., Blackwell, J. M., and Geffen, M. N. (2017). Cortical inhibitory interneurons control sensory processing. *Current Opinion in Neurobiology*, 46:200–207.
- Yamins, D. L. K., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., and DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, 111(23):8619–8624.
- Yao, J. D., Bremen, P., and Middlebrooks, J. C. (2015). Emergence of Spatial Stream Segregation in the Ascending Auditory Pathway. *Journal of Neuroscience*, 35(49):16199–16212.
- Yavorska, I. and Wehr, M. (2016). Somatostatin-Expressing Inhibitory Interneurons in Cortical Circuits. *Frontiers in Neural Circuits*, 10.

- Yin, T. (1994). Physiological correlates of the precedence effect and summing localization in the inferior colliculus of the cat. *The Journal of Neuroscience*, 14(9):5170–5186.
- Yoshioka, T., Sehr, A., Delcroix, M., Kinoshita, K., Maas, R., Nakatani, T., and Kellermann, W. (2012). Making Machines Understand Us in Reverberant Rooms: Robustness Against Reverberation for Automatic Speech Recognition. *IEEE Signal Processing Magazine*, 29(6):114–126.
- Young, E. D. and Oertel, D. (2004). Cochlear Nucleus. In Shepherd, G. M., editor, *The Synaptic Organization of the Brain*, pages 125–164. Oxford University Press.
- Yu, J., Hu, H., Agmon, A., and Svoboda, K. (2019). Principles Governing the Dynamics of GABAergic Interneurons in the Barrel Cortex. preprint, Neuroscience.
- Zahorik, P. and Wightman, F. L. (2001). Loudness constancy with varying sound source distance. *Nature Neuroscience*, 4(1):78–83.
- Zeisel, A., Muñoz-Manchado, A. B., Codeluppi, S., Lönnerberg, P., La Manno, G., Jureus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., Rolny, C., Castelo-Branco, G., Hjerling-Leffler, J., and Linnarsson, S. (2015). Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science (New York, N.Y.)*, 347(6226):1138–1142.
- Zeng, H. and Sanes, J. R. (2017). Neuronal cell-type classification: challenges, opportunities and the path forward. *Nature Reviews. Neuroscience*, 18(9):530–546.
- Zeng, H.-h., Huang, J.-f., Chen, M., Wen, Y.-q., Shen, Z.-m., and Poo, M.-m. (2019). Local homogeneity of tonotopic organization in the primary auditory cortex of marmosets. *Proceedings of the National Academy of Sciences*, 116(8):3239–3244.
- Zhang, L. I., Bao, S., and Merzenich, M. M. (2001). Persistent and specific influences of early acoustic environments on primary auditory cortex. *Nature Neuroscience*, 4(11):1123–1130.
- Zhang, Q., Hu, X., Hong, B., and Zhang, B. (2019). A hierarchical sparse coding model predicts acoustic feature encoding in both auditory midbrain and cortex. *PLOS Computational Biology*, 15(2):e1006766.
- Zhang, S. Q., Sun, X. D., and Jen, P. H. (1987). Anatomical study of neural projections to the superior colliculus of the big brown bat, *Eptesicus fuscus*. *Brain Research*, 416(2):375–380.
- Zoltowski, D. and Pillow, J. W. (2018). Scaling the Poisson GLM to massive neural datasets through polynomial approximations. In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R., editors, *Advances in Neural Information Processing Systems 31*, pages 3517–3527. Curran Associates, Inc.