Spectrotemporal receptive fields of auditory cortical neurons measured with two-photon calcium imaging



Arev A. Melikyan B.Sc

Exeter College, University of Oxford

Thesis submitted for the degree of

Master of Science by Research

Trinity 2025

For Mum, Dad and Lusin

Acknowledgments

A year is not particularly long, especially if you're working on the brain. But when faced with such a challenge it's those around you that get you through. I'd like to begin by thanking Prof. Kerry Walker for her supervision and mentorship. Kerry, I couldn't be more grateful to you for taking me under your wing despite a hectic schedule and making me feel at home. Our many meetings, zoom calls and hallway discussions have taught me so much about being a scientist, and your ability to put a smile on my face at the most challenging of times will not be forgotten.

Thanks also go to Dr Johannes Dahmen. Johannes, I admire your friendliness, patience and mystifying ability to head-fix mice. From teaching me how to use a two-photon microscope to fielding the most unexpected of questions, your pragmatism and level-headedness kept me grounded and made my project possible.

For countless discussions which kept inspiring me, I'd like to thank Dr Ben Willmore. Ben, working together I've come to realise that the devil *really is* in the detail. Thank you for being an incredible teacher and encouraging me to keep trying. I'd also like to thank Prof. Andy King for his leadership and making this year possible, and the rest of the Auditory Neuroscience Group for making me feel ever so welcome.

I am grateful to Exeter College, in which I found a home away from home. The year would not have been complete without the whirlwind social calendar, camaraderie and excitement that college life brings.

Finally, I'd like to thank my family for their unwavering support and belief in me, for consoling me in times of stress and encouraging me to pursue my passion. I wouldn't be here without you.

Scientific acknowledgments

Surgical procedures, widefield imaging and pre-processing of two-photon calcium imaging data were carried out by Dr Johannes Dahmen (*Chapter 3*). I would also like to thank Dr Ben Willmore for providing code to calibrate speakers (*Chapter 3*) and construct dynamic random chord stimuli (*Chapter 2*, *Chapter 3*).

Abstract

Arev A. Melikyan B.Sc

Exeter College, University of Oxford

Thesis submitted for the degree of Master of Science by Research

Trinity 2025

The spectrotemporal receptive field (STRF) describes the tuning of auditory neurons to complex sounds. Traditionally fit to electrical recordings, STRFs capture response properties in the form of a linear filter. Two-photon calcium imaging provides a complementary approach to electrophysiology by measuring activity with calcium indicators which fluoresce when neurons fire action potentials. Calcium imaging can be targeted to anatomically or functionally defined neurons by targeted expression of genetically-encoded calcium sensors. Despite successive indicator optimisation, it is unclear whether STRFs can be imaged. To answer this, I first modelled cortical neurons to understand the contribution of cellular (e.g. spike rate), indicator (e.g. kinetics) and experimental (e.g. noise) factors. Using simulated receptive fields to generate spiking and fluorescence responses to complex stimulus input, I found that fluorescence STRFs had clear structure resembling spiking STRFs, and slower temporal tuning reflecting indicator kinetics rather than neuronal properties. Importantly, STRFs were not hindered by indicator nonlinearity, noisy fluorescence traces or slow spiking. Extending this *in vivo*, STRFs could be obtained in stimulus-driven neurons and had tuning in agreement with frequency response areas. STRFs had slow temporal structure as predicted but were substantially noisier than in simulations. The main limitation was the ability to drive consistent responses as many neurons exhibited little to no evoked activity and high trial-to-trial variability. These results are an important first step in developing this new methodology, which could contribute to our understanding of complex auditory processing at cellular, circuit and global scales.

List of abbreviations

A1	Primary auditory cortex
AAF	Anterior auditory field
ANOVA	Analysis of variance
DRC	Dynamic random chord
FRA	Frequency response area
RF	Receptive field
S2F	Spike-to-fluorescence
STRF	Spectrotemporal receptive field
TORC	Temporally orthogonal ripple combination

Contents

1 General introduction **2**

1.1 Modelling neuronal function 2

1.1.1 Why modelling neurons is challenging 3

1.1.2 Encoding models 5

1.2 Spectrotemporal receptive fields 6

1.2.1 Stimulus choice 7

1.2.2 Fitting procedures 9

1.2.3 Model evaluation 11

1.2.4 Utility of the linear STRF 12

1.3 The auditory pathway 14

1.3.1 Auditory cortex 17

1.4 Imaging neural activity 18

1.4.1 Two-photon calcium imaging in auditory cortex 20

1.5 Thesis overview 22

2 Spike-to-fluorescence modelling of spectrotemporal receptive fields 25

2.1 Introduction 25

2.2 Methods 29

2.2.1 Stimuli 29

2.2.2 Simulated receptive fields 30
2.2.3 Cochleagrams 32
2.2.4 Linear-Nonlinear-Poisson spiking model 32
2.2.5 Spike-to-fluorescence forward model 33
2.2.6 STRF estimation 36
2.2.7 Tuning properties of STRFs 37
2.3 Results 38
2.3.1 Fluorescence STRFs 38
2.3.2 Indicator kinetics 40
2.3.3 Tuning properties 42
2.3.4 Excitatory-inhibitory balance 47
2.3.5 Indicator noise, nonlinearity and spike rate 49
2.3.6 Stimulus class 53
2.4 Discussion 55

3 *In vivo* two-photon calcium imaging of spectrotemporal receptive fields 60

3.1 Introduction 60
3.2 Methods 62
3.2.1 Animals 62
3.2.2 Surgical procedures 62
3.2.3 Widefield imaging 63
3.2.4 *In vivo* two-photon imaging 63
3.2.5 Stimuli 64
3.2.6 Presentation of stimuli 65
3.2.7 Isolating neural responses 65
3.2.8 Analysis of pure tone responses 66

3.2.9 Neuron inclusion . . . 66
3.2.10 Cochleagrams . . . 69
3.2.11 STRF estimation . . . 69
3.2.12 Tuning properties of STRFs . . . 70
3.2.13 Model evaluation . . . 70
3.3 Results . . . 71
3.3.1 STRFs can be fit to fluorescence traces . . . 73
3.3.2 Effect of regularisation . . . 76
3.3.3 Tuning properties of fluorescence-based STRFs . . . 80
3.3.4 Spike deconvolution . . . 82
3.3.5 Tuning of deconvolved spike STRFs . . . 86
3.3.6 Effect of stimulus density . . . 88
3.3.7 Fluorescence responses driven by complex stimuli . . . 91
3.3.8 Slow TORC STRFs . . . 96
3.4 Discussion . . . 100
3.4.1 Stimulus-evoked fluorescence and inclusion criteria . . . 100
3.4.2 Predictive power . . . 101
3.4.3 Choice of estimation procedure . . . 103
3.4.4 Conclusions . . . 105

4 General discussion **107**

4.1 Fluorescence responses to complex stimuli . . . 107
4.2 Stimulus and model optimisation for calcium imaging . . . 111
4.3 The future of imaging STRFs . . . 113
4.4 Conclusions . . . 114

List of Figures

2.1 Complex stimuli used to model spike trains 29

2.2 Construction of simulated receptive fields. 31

2.3 Linear-nonlinear-Poisson and spike-to-fluorescence cascade 34

2.4 STRF fits to simulated spiking and fluorescence 39

2.5 Effect of indicator kinetics on $\mathrm{STRF_f}$ 41

2.6 Comparison of tuning in spiking and fluorescence STRFs 43

2.7 Temporal tuning of fluorescence STRFs 45

2.8 Excitatory-inhibitory balance modelled STRFs 48

2.9 Spike-to-fluorescence with indicator nonlinearity, noise and different spike rates 50

2.10 STRFs at different neuronal spike rates 52

2.11 STRFs obtained using different stimuli 54

3.1 Experimental design for *in vivo* two-photon imaging 72

3.2 STRF fits to two-photon imaging data 75

3.3 Effect of regularisation on STRF structure 78

3.4 Tuning properties of fluorescence STRFs 81

3.5 Spike deconvolution in an exemplar field 83

3.6 STRFs fit to deconvolved spiking 85

3.7 Tuning of properties deconvolved spike STRFs 87
3.8 STRFs estimated at different stimulus densities 89
3.9 DRC-driven fluorescence responses 93
3.10 Accuracy of STRFs on held-out test data. 94
3.11 STRFs fit using slow TORCs 97
3.12 TORC-driven fluorescence responses 99

Chapter 1

General introduction

1.1 Modelling neuronal function

There are some fundamental questions which remain unanswered in modern neuroscience and attract considerable interest. One such question asks how information from the environment is captured, transmitted and transformed to generate behaviour, cognition and memory. Indeed, such higher functions are reliant on sensory input, so a unified neurobiological account of behaviour cannot be given without understanding sensory systems. In the auditory system, pressure waves arriving at the eardrum are transformed into the rich sonic tapestry that we *hear*. Whether it is the laughter of a child, the groove of a samba or the screeching of tyres against the road, all sounds begin as physical oscillations which are captured and processed to build a sense of the acoustic world around us. Not only is this done unconsciously, but our auditory system accomplishes it constantly and in real time. The complexity and beauty of this process is far from understood.

Our approach to studying sensory systems has evolved conceptually and technologically over the past century. Early work explored subjective reports of perception (Wundt,

1887), with more recent work emphasising internal representations and information processing (Boone and Piccinini, 2016). The mid-20th century saw technical and conceptual developments, with the ability to record from individual neurons uncovering details of the tonotopic organisation in auditory cortex (Merzenich et al., 1975; Walzl and Woolsey, 1946). At present, systems neuroscience employs an array of techniques allowing us to pose questions at various biological levels. Understanding a sensory system requires an appreciation of its anatomical organisation, cellular properties and encoding principles. It is likely that many unanswered questions would benefit from approaches which unify these aspects.

This thesis focuses on the encoding problem (Rieke et al., 1999), which asks whether we can predict the activity of neurons given some sensory input. This problem is particularly challenging since the space encompassed by all possible inputs is large and it is unclear what information is maintained in the neural code. By a combination of computational modelling and *in vivo* experiments, I investigate whether a popular model of auditory encoding can be applied to data acquired by two-photon calcium imaging. By validating this new methodology in mice, this thesis aims to broaden the scope of calcium imaging and further our understanding of how auditory cortex processes complex sounds.

1.1.1 Why modelling neurons is challenging

Modelling can take on different forms depending on the context. A useful classification of models into descriptive, mechanistic and interpretive types is provided by Dayan and Abbott, 2001. Descriptive models are largely phenomenological, attempting to summarise experimental findings compactly and accurately. In doing so, they answer the question of 'what' the nervous system accomplishes. Mechanistic models address the question of 'how', by making use of known anatomy and physiology. Interpretive models draw on theoretical frameworks to investigate the functional relevance of certain

features of neural activity, helping to explain why the nervous system operates as it does. Many models of the auditory system are strongly motivated by signal processing methods, and describe neurons in terms of filters (Aertsen et al., 1981; De Boer, 1968; Thorson et al., 2015). While different approaches to modelling should ideally converge, this lies ahead of most modelling paradigms at present (Rudnicki et al., 2015).

Describing the behaviour of single neurons accurately is difficult for various reasons. Neuronal responses are noisy *in vitro* and even more unpredictable *in vivo*, with successive trials of identical stimuli driving considerably different spike trains (Gerstner et al., 2014). For instance, integrate-and-fire neurons spike regularly when driven by a constant current, but real neurons behave far less regularly (Gerstner et al., 2014, however see Pospischil et al., 2011). A significant portion of noise in neural networks arises from stochasticity at the pre-synapse (e.g. vesicle release) and the post-synapse (e.g. variability in changes in synaptic conductance following binding of neurotransmitter)(Rusakov et al., 2020). Additional sources of noise include intrinsic factors such as stochastic spike generation and extraneous variables such as thermal and photonic noise (Gerstner et al., 2014). Importantly, irregularities in firing patterns may in fact be part of an elaborate neural code beyond our comprehension (Stein et al., 2005). For instance, 'noise' may arise from neurons multiplexing processing and participating in distinct subpopulations simultaneously to accomplish different tasks (for recent examples in auditory cortex see Hu et al., 2024; See et al., 2018, 2021). Regardless, trial-to-trial variability complicates the interpretation of responses but may represent a computationally advantageous phenomenon (Wiesenfeld and Moss, 1995).

Finally, the choice of *which* measure of activity is to be modelled raises additional questions. Many formulations assume information is contained in the spike rate and hence model this. However, the nervous system does not always adhere to a simple rate code (Gerstner et al., 1997). For instance, the spatial location of auditory stimuli

is represented by spike latency in some auditory cortical neurons (Mrsic-Flogel et al., 2005) . In this case, a peristimulus time histogram-based representation could overlook this phenomenon. The design of a model should therefore consider known coding mechanisms. Unfortunately, this is often the question that the model is attempting to answer.

1.1.2 Encoding models

Sensory systems have the unenviable task of transforming a barrage of input into useful representations. The spaces in which a system operates depend on the sensory modality – for instance, the visual system operates in the spatial domain comprising the visual field. In contrast, the auditory system operates within a stimulus space comprised of the audible frequencies and their respective intensities. For both systems, a temporal dimension is added since input is dynamic. Sensory systems carry out diverse tasks including: (i) real-time processing of continuous input; (ii) integration across numerous inputs; and (iii) countering significant internal and external noise.

To explain the transformations underlying this, encoding models find a mapping between sensory input and neural output known as a stimulus-response function (Triplett and Goodhill, 2019). Encoding models are fit to various response measures including spiking (Chichilnisky, 2001), fluorescence (Runyan et al., 2017), membrane potential (Machens et al., 2004) and fMRI blood-oxygen-level-dependent signals (Venezia et al., 2021). The response measure to be predicted constraints the aspects of encoding which a model may hope to capture. For example, a model fit to fMRI signals is unlikely to predict fine spatiotemporal activity within a cortical circuit but can explain changes in functional connectivity and regional activity across the brain.

Importantly, encoding models do not currently encapsulate subcellular functionality including channel dynamics and synaptic transmission but provide abstracted descrip-

tions of processing (Meyer et al., 2017). The power of such models lies in their phenomenological approach which enables intuitive descriptions of processing, such as the description of an auditory neuron as a filter. Further, it is possible that universally optimal encoding models describing the activity of neurons are not feasible. Regardless, the generality of certain models makes them widely applicable to processing along the hierarchical stages of a sensory pathway. Receptive field-based models are one such example which provide a mapping between stimulus space, defined by the dimensions along which sensory input varies, and the activity of a neuron. In the auditory system, these models relate the spectral and temporal features of sound to neural output and so are termed spectrotemporal receptive fields (STRFs).

1.2 Spectrotemporal receptive fields

The behaviour of auditory neurons can be described in different ways reflecting their spectral and temporal preferences. Frequency tuning curves represent the simple end of this continuum and describe the response to tones of different frequencies. Frequency-response areas extend this by jointly considering frequency and intensity tuning. However, neither of these models can provide a quantitative prediction of time-varying activity in response to dynamic sounds (Aertsen and Johannesma, 1981). Furthermore, it is well documented in theoretical (V. Z. Marmarelis, 2004) and experimental (Rotman et al., 2001; Theunissen et al., 2000) literature that models fit using simple stimuli fail to predict responses to complex sounds.

The STRF attempts to address this issue by jointly describing the spectral and temporal tuning properties of a neuron estimated using complex stimuli (Eggermont, 1993; Klein et al., 2000; Kowalski et al., 1996a, 1996b). Following subcortical processing, cortical neurons are better thought of as acting on a frequency domain representation (Chi et al., 2005; Rahman et al., 2020). The STRF, $k(f, h)$ is a linear filter which

provides a mapping between the dynamic spectrum, $s(f,t)$, and the spike rate $r(t)$:

$$r(t) = \sum k(f,h)s(f,t-h) + e(t) \tag{1.1}$$

STRF weights describe the effect of stimulus energy at frequency f and lag h on the mean firing rate, and $e(t)$ is the part of the response which is not predicted by the model (e.g. noise). Positive weights indicate parts of the stimulus correlated with increased firing whereas negative weights correspond to parts which are anticorrelated with firing and the theoretical maximum response is evoked by a stimulus whose dynamic spectrum matches the STRF. This simple model can capture a range of processing characteristics including simple excitation or inhibition, spectral/temporal sidebands and selectivity to the direction of frequency sweeps (Elhilali et al., 2004).

1.2.1 Stimulus choice

The choice of an appropriate stimulus to estimate STRFs is a challenging task. STRFs in auditory cortex have been measured with both natural sounds (e.g. animal vocalisations and speech) and synthetic stimuli (Blake and Merzenich, 2002; David et al., 2009; deCharms et al., 1998; Machens et al., 2004; Willmore et al., 2016). Limited recording times necessitate stimulus ensembles which tile stimulus space efficiently, that is, probing the various axes along which sounds vary such as carrier structure and frequency/amplitude modulations (C. Atencio and Schreiner, 2013). While natural stimuli relevant to the animal under study can elicit strong, naturalistic patterns of neural activity, synthetic stimuli grant total control to the experimenter who can vary parameters as desired. This work makes use of two well-established types of synthetic stimuli discussed below.

Dynamic random chords (DRCs) are complex tonal stimuli comprised of randomly interleaved pure tone pips at various frequencies (deCharms et al., 1998). They can

be considered an generalisation of the two-tone paradigm where pure tones of different frequencies are presented closely in time to identify inhibitory sidebands (Abbas and Sachs, 1976). Various parameters including the spacing of frequencies, temporal density and duration of tone pips are specified. Additionally, tone pip amplitude can be constant (Blake and Merzenich, 2002; deCharms et al., 1998) or drawn from a distribution (Linden et al., 2003; Rabinowitz et al., 2012; Willmore et al., 2016). Properties of DRC tone pips including amplitude contrast and temporal density influence receptive field estimates (Blake and Merzenich, 2002; Cooke et al., 2018; Rabinowitz et al., 2012).

Temporally orthogonal ripple combination (TORC) stimuli are created by summing several moving ripples, the auditory corollary of drifting gratings (Klein et al., 2000). Moving ripples are created by applying sinusoidal modulations, which drift up or down the frequency axis, to the spectrotemporal envelope of a broadband noise carrier (Kowalski et al., 1996a). The spectrotemporal envelope, $S(t, x)$, of a moving ripple can be described by the following (C. Atencio and Schreiner, 2013):

$$S(t, x) = D[1 + A \sin(2\pi\omega t + 2\pi\Omega x + \Phi)] \tag{1.2}$$

where x is the frequency in octaves relative to the lowest carrier frequency carrier, t is time and ω (Hz) and Ω (cycles/octave) are the temporal and spectral modulation frequencies. D sets the modulation depth, A sets the amplitude of the sinusoid and Φ randomises the phase of the moving ripple. The sign of ω determines the direction of drift along the frequency axis. Superimposing the envelopes of moving ripples with different temporal modulation rates gives TORCs, hence temporal orthogonality. An important motivation behind the use of sinusoids is the Fourier theorem which posits that an arbitrary dynamic spectrum can be decomposed into a series of moving ripples by a two-dimensional Fourier transform (Klein et al., 2000). Assuming that neural

responses are linearly dependent upon the spectrogram of the stimulus, it follows that the response to an arbitrary stimulus can be approximated by the sum of responses to its constituent moving ripples (Kowalski et al., 1996b).

Both DRC and TORC stimuli are synthetic sounds that have the advantage that their acoustical components are minimally correlated (in a relevant space) and can be specified parametrically as desired. Importantly, these stimuli lack natural spectrotemporal structure that would be ecologically relevant to the listening animal (allowing me to control for learning and to avoid top-down arousal effects) and evoke fundamentally different responses than natural sounds (David et al., 2009; Theunissen et al., 2001). The gold-standard for an encoding model is to describe stimulus transformations under natural conditions. STRFs estimated using synthetic stimuli generalise poorly to natural stimuli (though DRC stimuli outperform TORCs in this respect) and STRFs fit using natural sounds are better at cross-prediction (Willmore et al., 2016). Despite these limitations, DRC and TORC stimuli are used in this work for their advantages listed above and because they are known to be potential drivers of auditory cortical activity.

1.2.2 Fitting procedures

STRFs have historically been estimated by *spectrotemporal reverse correlation*, a procedure which identifies stimulus features eliciting neural responses by correlating the time-reversed spectrogram with the peristimulus-time histogram (De Boer, 1968; De Boer and Kuyper, 1968). Over an adequate duration, reverse correlation captures features of interest and averages out other patterns, yielding a kernel (the STRF) representing the average stimulus feature preceding a response. When action potentials are being measured, the STRF estimated by reverse correlation is equivalent to the spike-triggered average (i.e. the average spectrogram segment preceding an action potential; P. Z. Marmarelis and Marmarelis, 1978). Importantly, reverse correlation

requires that stimuli be uncorrelated in frequency and time to avoid fitting biased STRFs with spurious structure. This places a constraint on stimulus choice, ruling out natural sounds which are highly correlated but desirable to fit ethologically valid STRFs. Additionally, reverse correlation is an unregularised method which often fits noisy kernels with poor predictions on test data (David et al., 2007).

Such issues have prompted the development of alternative fitting procedures allowing greater flexibility. A closely related method, normalised reverse correlation, accounts for stimulus correlations by multiplying the STRF by the inverse of the stimulus autocorrelation matrix (Theunissen et al., 2001). This step effectively 'divides' stimulus correlations out of the filter estimate. Various other methods have been proposed which differ in their algorithm and regulariser but ultimately yield a linear STRF (David et al., 2007; Depireux et al., 2001; Park and Pillow, 2011). Other models such as maximally informative dimensions analysis (Nagel and Doupe, 2008) and spike-triggered covariance (Paninski, 2003) extend the STRF model to multiple linear filters, representing different stimulus features driving individual neurons. STRFs can also be fit by linear regression with regularisation (a type of generalised linear model) where filter coefficients are estimated via gradient descent/ascent-based methods (Calabrese et al., 2011). These algorithms seek to minimise an error term, E, which is a function of the difference between model predictions, $\hat{y}[t]$, and the true neural response, $y[t]$:

$$E_{\mathrm{L1}} = \sum_t \left(y[t] - \hat{y}[t]\right)^2 + \lambda \sum_{fh} |W_{fh}| \tag{1.3}$$

$$E_{\mathrm{L2}} = \sum_t \left(y[t] - \hat{y}[t]\right)^2 + \lambda \sum_{fh} {W_{fh}}^2 \tag{1.4}$$

Since hundreds to thousands of coefficients must be estimated, the possible parameter space is very large. Stimuli attempt to sample numerous dimensions of this space but are unable cover each one sufficiently. Response variability along poorly-sampled di-

mensions is not eliminated by averaging across presentations and affects STRF weights (Meyer et al., 2017). Since random variability is captured as stimulus dependence, the model is said to be overfitting to noise and requires regularisation. This is achieved by the inclusion of a hyperparameter included in the error function. A penalty on the L1- ($|W_{fh}|$) or L2-norm ($W_{fh}{}^2$) with weighting λ enforces a prior of sparsity or smoothness on the receptive field, respectively.

Alternative techniques avoid strict priors by optimising their distributions with respect to the data – such *evidence optimisation* algorithms exist to estimate STRFs (M. Sahani and Linden, 2002). Biologically plausible kernels are likely to be globally sparse but locally smooth (Jones and Palmer, 1987; Park and Pillow, 2011), thus fitting with these constraints improve estimates on limited and/or noisy datasets. The question addressed by this thesis is whether STRFs can be fit to two-photon calcium imaging data. Since the form of fluorescence signals does not purely reflect cellular properties but also those of the indicator, algorithms which make biologically-motivated assumptions on kernel structure are not used. Thus, linear regression with L1- and L2-regularisation (Eq. 1.3 and 1.4) are the main method of fitting STRFs in this thesis.

1.2.3 Model evaluation

Once a model is fit, evaluation is necessary to assess whether it provides an acceptable approximation of the underlying process. The correlation coefficient, r, between the model prediction and measured response is an intuitive starting point as it is bounded between -1 and 1 and increases as the model generates more accurate predictions. A related measure, R^2, describes the proportion of response variance accounted for by the model. However, encoding models often predict complex, highly variable stimulus response relationships (Hsu et al., 2004; Sahani and Linden, 2003; Schoppe et al., 2016), which complicates the interpretation of simple metrics like the correlation coefficient. Neural responses are fundamentally variable and exceedingly large numbers of

stimuli and repeats are required to eliminate what is assumed to be noise. Importantly, response variability can be stimulus independent (e.g. recording noise) or generated by an inherently variable stimulus-dependent process (Carandini, 2004; Stein et al., 2005). Regardless, this places an upper bound on the correlation coefficient as only part of the signal is linearly-predictable (Schoppe et al., 2016).

A variant of the correlation coefficient, known as CC_{norm}, seeks to normalise the correlation coefficient based on the signal power in the measured response (Hsu et al., 2004; Schoppe et al., 2016). This is an important step to contextualise the performance of a model if it cannot be expected to produce perfect predictions, as is often the case. CC_{norm} values are bounded between -1 and 1 (except in cases of overfitting) and are obtained by dividing the correlation coefficient by a maximum achievable value given response variability (see *Section 3.2.13*). Signal power explained (Sahani and Linden, 2003) is another measure which seeks to normalise model performance based on response variability, but does not have a lower bound like CC_{norm}. Ultimately, the type of metric used will depend on the context and type of model, with implications on how performance is interpreted.

1.2.4 Utility of the linear STRF

STRFs have been fit in a variety of animals including primates (deCharms et al., 1998), ferrets (Depireux et al., 2001), rodents (Homma et al., 2021; Laudanski et al., 2012; Linden et al., 2003) and recently to humans (Holdgraf et al., 2016; Hullett et al., 2016). In ferrets, STRFs have been used to study task-dependent plasticity in auditory cortex and midbrain across a battery of behavioural tasks including tone detection (J. Fritz et al., 2003), frequency discrimination (J. Fritz et al., 2005), gap detection and click rate discrimination (J. B. Fritz et al., 2005). Across tasks, changes in STRFs facilitate the detection of behaviourally relevant stimuli. For instance, conditioning animals to stop licking in response to a target tone leads to new-onset excitation at the tone

frequency (J. Fritz et al., 2003). The exact nature and mechanism of rapid plasticity is not understood but is likely to include top-down contributions from frontal areas (Francis et al., 2018; J. B. Fritz et al., 2010; Winkowski et al., 2013, 2018) as well as changes in bottom-up processing (J. D. Chambers et al., 2019).

The principal disadvantage of the STRF is that it provides a linear approximation of processing. If the underlying stimulus-response function is linear within the probed stimulus subspace, the estimated weight provides a sensible description whereas other e.g. quadratic dependencies cannot be captured (David, 2018; Rust et al., 2005). Auditory neurons are known to exhibit a range of nonlinearities related to spike generation (e.g. thresholding, saturating), stimulus statistics (e.g. divisive inhibition, gain control) and spectrotemporal integration (Ahrens, Linden, and Sahani, 2008; Brosch and Schreiner, 1997; Dean et al., 2005; Sutter and Schreiner, 1991). Not only are nonlinearities not captured in linear STRFs, but they lead to biased or spurious structure due to interactions with stimulus statistics and the estimation procedure (Christianson et al., 2008; David et al., 2009; Theunissen et al., 2000).

Including input/output nonlinearities improves model performance compared to the standard STRF but requires more data and elaborate model architectures to be successful. At the simplest, a static sigmoidal function following a linear STRF can improve model predictions by accounting for instantaneous nonlinearities such as spike thresholding and saturation (Chichilnisky, 2001). Dedicated input/output nonlinearities can also be added to reflect specific aspects of auditory processing. For instance, discrepancies between the linear STRF estimated using TORCs and speech stimuli can be partially accounted for by including an input nonlinearity representing synaptic depression (David et al., 2009). Input nonlinearities acting on the stimulus at different combinations of delays (Ahrens, Paninski, and Sahani, 2008), or on the stimulus spectrogram to represent cochlear processing (Elhilali et al., 2004; Rahman et al., 2020)

have also demonstrated improvements in prediction accuracy. In this way, the STRF can be used as a substrate for more elaborate models of auditory processing.

1.3 The auditory pathway

Our sense of hearing is critical because the acoustic environment is a rich source of information invaluable for survival. Although the auditory pathway is often presented as a single sequence of stations, this is an oversimplification. Auditory information takes numerous distinct paths to cortex, each of which transform it to serve different purposes such as spectral or temporal processing. Additionally, retrograde projections target points along the pathway down to the cochlea (Elgueda and Delano, 2020; Terreros and Delano, 2015). This short review will provide a whistle-stop tour of subcortical processing in the mammalian auditory system (for a detailed treatment see Webster et al., 1992 and Pickles, 2015).

The sounds we hear and the percepts they evoke are ultimately due to the vibration of particles. Oscillations of an object cause longitudinal pressure waves in air which are transmitted to the ear canal and the tympanic membrane. Vibrations of the tympanic membrane are transmitted and amplified via a sequence of three bones, known as the ossicular chain. At the end of this chain, vibrations are concentrated over a very small area called the oval window. Displacement of the oval window then causes pressure waves to propagate throughout the cochlea, a fluid-filled spiral-shaped structure in the inner ear.

The cochlea contains the coiled basilar membrane which has resonant frequencies that are logarithmically spaced from high to low along its length. This establishes a scheme of cochlear tonotopy. The Organ of Corti runs along the basilar membrane and contains *hair cells* responsible for the conversion of the vibrations of the membrane into electrical signals. Two types of hair cell exist, *inner* and *outer* hair cells, and both

contain hair-like projections called stereocilia positioned against the basilar membrane. Movement of the basilar membrane causes a shear force against stereocilia and therefore rhythmic deflections. Mechanoreceptors within hair cells register these deflections and open potassium channels leading to cellular depolarisation which forms the basis of encoding at the auditory periphery. Outer hair cells employ a similar mechanism but are thought to amplify acoustic information via a mechanical positive feedback mechanism (Schnupp et al., 2010).

Hair cells synapse with auditory nerve fibres, each of which are most sensitive to a specific *characteristic* frequency and responds by firing action potentials at a specific phase of the sinusoid. Such phase-locked responses have been observed for frequencies up to 4-5 kHz, especially if population phase locking is invoked, above which sustained responses predominate (Johnson, 1980; Verschooten et al., 2019). Auditory nerve fibres preserve the cochlear tonotopy which is maintained up to the level of cortex and hence follow a place code when encoding sounds. After entering the brainstem, auditory nerve fibres synapse in the cochlear nucleus, a group of three distinct, tonotopic regions. The cochlear nucleus contains over 20 types of cells, which vary in morphology, physiology and response properties (Merchán et al., 1993; Schnupp et al., 2010). Auditory information from the cochlear nucleus is transmitted directly to the inferior colliculus or via the superior olivary complex, depending on the projecting cell type.

Bilateral projections from the cochlear nuclei innervate the superior olivary complex (Tremblay and Burkhard, 2012), making it the initial site of significant binaural interaction (Kulesza, 2007). The superior olivary complex integrates binaural cues to achieve azimuthal localisation (Heffner and Masterton, 1990; Illing et al., 2000), has roles in temporal processing (Kulesza, 2007), and provides descending modulation to outer hair cells (Tremblay and Burkhard, 2012). Ascending projections from the superior olivary complex then synapse in the inferior colliculus along with descending cor-

tical projections (Ito et al., 2016), commissural connections from contralateral inferior colliculus and projections from superior colliculus (Coleman and Clerici, 1987). The inferior colliculus has a complex anatomical organisation, including multiple classes of interneuron whose roles in auditory processing are poorly understood (Ito et al., 2016).

Neurons in inferior colliculus exhibit receptive fields with complex tuning properties including inhibitory sidebands and sensitivity to frequency sweeps (Escabí and Schreiner, 2002), however their stimulus preference can be described reasonably well using a single filter (C. A. Atencio et al., 2012). This differs to cortical neurons where a moderately good description generally requires two linear filters with a joint nonlinearity (C. A. Atencio et al., 2008, 2009). Additionally, many neurons in inferior colliculus are nonlinear: 40% of neurons have significantly different STRFs when obtained using different stimuli (Escabí and Schreiner, 2002). Indeed, incorporating an input nonlinearity representing midbrain adaptation to mean sound level improves the predictions of STRFs fit to auditory cortex (Willmore et al., 2016).

Projections from inferior colliculus synapse next in the medial geniculate body, an obligatory relay station before auditory cortex (Purves, 2001). STRFs measured in the auditory thalamus are like cortex except tuning to shorter latencies and higher temporal modulation rates (Miller et al., 2002). A minority of thalamic STRFs exhibit atypical structure including sensitivity to frequency sweeps and multiple disjoint excitatory/inhibitory components. The emergence of complex spectrotemporal tuning suggests extensive integration of convergent inputs prior to the thalamus. Interestingly, simultaneous recordings from auditory cortex and thalamus reveals that functionally connected units often have largely non-overlapping STRF features (Miller et al., 2001). How cortical receptive fields are formed is not fully understood, although direct inheritance and construction from numerous inputs constitute possible mechanisms.

1.3.1 Auditory cortex

Auditory cortex is comprised of anatomically and functionally distinct fields which vary across species. Mouse auditory cortex can be subdivided into several areas including primary auditory cortex (A1) and anterior auditory field (AAF), which are tonotopically arranged with mirrored dorsoventral gradients for best frequency (Guo et al., 2012; Joachimsthaler et al., 2014). Secondary auditory cortex is found ventral to the primary fields (A1 and AAF) and additionally involved in higher-order functions such as sensorimotor learning (Cook et al., 2022) and movement-related modulation (Morandell et al., 2023).

Like other sensory cortices, A1 contains six layers which differ in cell type, afferent/efferent projections and functional properties. Topographic organisation is found in supragranular (2/3) and thalamorecipient layers (3b/4), however it is more heterogeneous in the former (Winkowski and Kanold, 2013). Layer 2/3 neurons receive input from other supragranular neurons, corticocortical projections and thalamorecipient neurons and represent a further step in cortical processing compared to infragranular layers (Chang and Kawai, 2018). Neurons in A1 code for attributes beyond just frequency and level, including interaural differences (Panniello et al., 2018), frequency/amplitude modulations, and frequency sweeps (Kuchibhotla and Bathellier, 2018). A diversity of response patterns including bursting, sustained and other types are observed dependent on the stimulus, neuron, and state of anaesthesia (Wang, 2007). Cells in anaesthetised animals show transient onset responses to sounds (DeWeese et al., 2003), whereas sustained firing to preferred stimuli is seen in the awake state (Wang et al., 2005). During continuous stimuli, a large population of onset-responsive neurons is whittled down quickly to a small, well-driven group of neurons, resulting in the emergence of a population- and time-selective representation (Wang, 2007). Differences in spiking responses based on the state of anaesthesia and nature

of stimulus must be considered when comparing findings.

Auditory cortex is likely to support a transformation from a neural code for acoustic properties to one for perceptual properties (Kuchibhotla and Bathellier, 2018; Wang, 2007). There is evidence that cortex encodes auditory objects, which are perceptual constructs formed by parsing and assigning specific spectrotemporal patterns of sounds in the acoustic environment, such as tweeting, to distinct sources – in this case, a bird (Bizley and Cohen, 2013; Griffiths and Warren, 2004; Winkler et al., 2009). Firstly, representations appear to be invariant to the exact structure of the waveform of a particular type of sound (Carruthers et al., 2015), and discrete, showing rapid switches from one object to another (Bathellier et al., 2012). Additionally, activity in primary auditory areas is modulated based on contextual factors including behavioural state, locomotor activity and stimulus statistics (David, 2018). The signalling of such variables is not fully understood, however it is known to involve a combination of bottom-up adaptation and top-down intracortical modulation from frontal and other auditory areas recruited as part of an attentional circuit (J. B. Fritz et al., 2013).

1.4 Imaging neural activity

Imaging neural activity using voltage and calcium indicators provides a complementary approach to electrophysiology. The former method uses sensors engineered to emit fluorescence based on the membrane potential and is capable of reporting subthreshold activity at high temporal resolution (Peterka et al., 2011). At present, voltage imaging is hampered by biological and logistical limitations. Indicator molecules must be localised to the cell membrane and compete with significant noise from molecules on intracellular membranes, resulting in low signal-to-noise ratios (Bando et al., 2019). Furthermore, most current microscopy systems are not appropriate for voltage imaging – commonly used galvo-based and acousto-optical scanning systems are not capable of

covering more than a few cells at the high acquisition rates required to capture action potentials.

Two-photon calcium imaging is an alternative technique which uses intracellular calcium-sensitive indicators to report neural activity by a proxy measure – cytosolic calcium concentration (Grienberger and Konnerth, 2012). The opening of voltage-gated calcium channels during action potentials leads to an influx of the calcium which binds to indicator throughout the cell. Calcium-bound indicator molecules fluorescence due to excitation by femtosecond-pulsed laser beams directed over the tissue in a two-dimensional pattern (Helmchen and Denk, 2005; Malik et al., 2011). Early indicators such as the bioluminescence photoprotein aequorin extracted from jellyfish (Ashley and Ridgway, 1968; Shimomura et al., 1962) had poor sensitivities and dynamic ranges and were superseded by synthetic indicators (Paredes et al., 2008; Tsien, 1980). With manual dye delivery posing a significant challenge, the development of genetically-encoded calcium indicators proved to be revolutionary (Miyawaki et al., 1997). The typical genetically-encoded indicator, GCaMP, has been iteratively optimised to have higher sensitivity, and can be targeted to specific regions, cells, and even subcellular compartments (Zhang et al., 2023). Stable expression levels enable long-term *in vivo* imaging through a cranial window (Grienberger and Konnerth, 2012). Over the preceding decades, calcium imaging has grown in popularity and become a standard method for tracking neural activity *in vivo*.

An important limitation of calcium imaging is the speed of the signal relative to action potentials. There are two contributors to this slowness: the first relates to the speed of underlying calcium transients, the second to the indicator kinetics. When a neuron spikes, intracellular calcium concentration rises within milliseconds and fall over tens of milliseconds (Maravall et al., 2000; Sabatini et al., 2002; Zhang and Looger, 2024). Even if one had a perfect indicator which fluoresces based only on intracellular calcium,

the temporal resolution of the signal would still be an order of magnitude slower than action potentials. However, GCaMP fluorescence traces are often ten-fold slower than calcium transients, limited by the speed of conformational changes at subunit interfaces within indicator molecules (Zhang and Looger, 2024). Thus, indicator kinetics describe the rise and decay times of transients and represent an important variable with regards to selection of an appropriate indicator and the temporal resolution one may hope to achieve. The fast variant of GCaMP8, from the latest generation of GCaMP indicators, has a half-rise time of ~ 2 milliseconds and a half-decay time of ~ 50 milliseconds in response to single action potentials, making it the most accurate yet in capturing the true dynamics of calcium signals (Zhang et al., 2023).

Beyond kinetics, linearity and sensitivity importantly guide the choice of indicator. Linearity refers to the ability of a sensor to sum action potentials and is requisite for accurate spiking inference. There exists a trade-off between linearity and sensitivity; sensitive indicators report single action potentials with larger transients however have a poorer dynamic range and exhibit sublinearity at fewer spikes (Zhang and Looger, 2024). GCaMP8 is reported to integrate bursts of at least 5 action potentials linearly (Zhang et al., 2023), however this may be insufficient in fast-spiking neurons or when continuous, complex stimuli evoke robust patterns of activity. Interdependencies between intrinsic indicator properties (e.g. affinity, kinetics, brightness) and the cellular properties (e.g. expression levels, calcium buffering) further complicate the interpretation of neural activity from fluorescence data (Grienberger and Konnerth, 2012).

1.4.1 Two-photon calcium imaging in auditory cortex

Understanding the anatomical and functional organisation of sensory cortices is a central endeavour of systems neuroscience. Mapping the spatial arrangement of functionally distinct or coordinated processing helps to form a coherent picture of the entire

system and allows comparisons between organisms, healing versus injury states, and during development (Bednar and Wilson, 2016; White and Fitzpatrick, 2007). Traditionally, electrophysiology has been the mainstay for defining such maps and has revealed smooth tonotopicity along the A1 dorsoventral axis in multiple species (Kelly et al., 1986; Merzenich and Brugge, 1973; Sally and Kelly, 1988). However, findings from imaging experiments have contested these results, revealing fractured tonotopicity on a local scale. For instance, neighbouring units can show best frequencies separated by up to 4 octaves (Bandyopadhyay et al., 2010; Rothschild et al., 2010). Imaging also uncovers broad and complex frequency tuning less likely to be observed with electrophysiology (Issa et al., 2014), and that units with multipeaked tuning are less spatially organised than single-peaked units (Gaucher et al., 2020).

Numerous explanations for this disparity exist, including the bias of electrophysiology towards strongly driven units while imaging captures activity from sparsely active neurons. Such units may appear to have less defined tuning, at least to pure tone stimuli. Additionally, recordings of multiunit activity could smooth the local heterogeneity observed with imaging. Interestingly, including subthreshold fluorescence changes reveals clearer topographic organisation in A1 than suprathreshold activity alone (Bandyopadhyay et al., 2010), suggesting that nearby units receive overlapping inputs but perform distinct operations.

STRFs measured in auditory cortex are often smooth in spectrotemporal space and exhibit stereotyped structure such as simple excitation or inhibitory sidebands. Whether disparities exist between spectrotemporal tuning measured by electrophysiology and imaging, as they do for pure tones, is unknown. Additionally, most work measuring STRFs in auditory cortex has been carried out in ferrets, but interspecies differences in spectrotemporal tuning likely exist (deCharms et al., 1998; Homma et al., 2024). Neurons are specialised to process naturalistic sounds and STRFs exhibit significant,

largely redundant support for spectrotemporal dimensions relevant to an animal (Woolley et al., 2005). It is unclear whether calcium imaging can uncover species-specific features in STRFs. If this is possible, we could use this approach to much better understand the cortical circuits underlying auditory plasticity, by comparing STRF changes during behaviour across different genetically targeted cell types and neural projections between regions (e.g. auditory cortical cells receiving direct feedback from specific regions of frontal cortex).

1.5 Thesis overview

My objective is to investigate whether neuronal responses measured using two-photon calcium imaging can be used to fit STRFs. Previously, only fast-varying signals (e.g. spiking, local field potential, magnetoencephalography) have been used and it is unclear whether slow signals such as fluorescence are adequate. Since the answer to this question is likely to depend on a multitude of variables related to neuronal properties (e.g. spike rate), indicator properties (e.g. kinetics, sensitivity) and stimulus choice, a two-pronged approach is taken. The first experimental chapter uses a computational model to simulate spike trains and fluorescence traces of auditory cortical neurons stimulated by complex sounds. By manipulation of the above parameters, the model seeks to identify optimal conditions under which STRFs can be imaged. The second experimental chapter builds on these results by imaging auditory cortex in awake mice.

- **Chapter 2**: This chapter uses an end-to-end computational model to ask whether STRFs can feasibly be fit to fluorescence. Auditory cortical neurons are simulated with tuning described by receptive fields which form the linear stage of a linear-nonlinear-Poisson spiking model. Spike trains are generated using DRC, TORC and natural stimuli and converted to fluorescence traces by convolution with a double exponential kernel. STRFs are fit to spiking and fluorescence using

different spike rates, indicator kinetics and noise levels, and are compared to the original receptive fields.

- **Chapter 3**: Building on the results of *Chapter 2*, *Chapter 3* describes two-photon calcium imaging experiments performed in awake mice. A combination of complex stimuli and pure tones are presented to obtain STRFs and frequency response areas for comparison. The structure, tuning properties and prediction on held-out data of calcium STRFs are assessed, with explicit comparisons drawn to previous results from electrophysiology. Finally, STRFs are fit to deconvolved spikes to assess whether this approach reveals tuning properties on faster, biologically plausible timescales.

Chapter 2

Spike-to-fluorescence modelling of spectrotemporal receptive fields

2.1 Introduction

Auditory neurons integrate acoustical information over time and frequency to form representations of complex sounds. Information encoded in the dynamic activity of neurons is processed in multiple hierarchical and parallel feedforward stations within the auditory system, from the cochlea to auditory cortex, including feedback projections (Elgueda and Delano, 2020; Terreros and Delano, 2015). The stimulus preference of individual neurons along this neural pathway can be described using receptive field (RF)-based models, which are widely employed across sensory modalities (Ramirez et al., 2014; Rust et al., 2005). In the visual system, spatiotemporal receptive fields describe neuronal tuning in terms of the spatial location and timing of points of light. In the auditory system, the spectrotemporal receptive field (STRF) describes how

acoustic energy at different frequencies and latencies drives activity in neurons.

STRFs have proved successful as first-order approximations of stimulus processing in the peripheral and central auditory system. STRF tuning properties recapitulate spectral preferences obtained by pure tones, while revealing preferences towards complex stimulus features that cannot be probed with simple stimuli, such as frequency sweep selectivity and sideband inhibition (deCharms et al., 1998; Linden et al., 2003). The STRF has also been used as a substrate for more elaborate encoding models (Ahrens, Linden, and Sahani, 2008; Rabinowitz et al., 2012; Williamson et al., 2016), as well as in neurophysiology in behaving animals that has uncovered task-dependent neural plasticity in auditory cortex (J. Fritz et al., 2003, 2005). Despite it being clear that predictions of linear STRFs only explain a fraction of response variance in auditory cortex (Machens et al., 2004), they continue to provide useful, compact descriptors of neuronal tuning.

STRFs are classically fit to microelectrode recordings obtained in animals during the presentation of various types of complex stimuli. Changes in spiking rate associated with specific features in the stimulus are used to infer a linear mapping between the stimulus and neural activity. Crucially, electrophysiology has the high degree of temporal resolution required to confidently attribute individual spiking events to specific time points within the stimulus.

The introduction of two-photon calcium imaging to measure auditory cortical has provided unparalleled access to the activity of a large number of individual neurons over large cortical areas. Action potentials (APs) produce large, sudden increases in cytosolic calcium which then fall over tens of milliseconds (Zhang and Looger, 2024). Calcium indicators (such as the genetically-encoded GCaMP sensors) bind to calcium and undergo conformational changes leading to fluorescence which can be imaged by two-photon microscopy. In this manner, increases in fluorescence due to calcium tran-

sients associated with APs provide a readout of neural activity. Calcium transients have a stereotyped structure, consisting of a sharp rise from baseline (the rapid influx of calcium through voltage-gated calcium channels) followed by a slower, exponential-like decay process (the slow buffering of calcium back out of the neuron following repolarization). Importantly, imaging reduces bias towards neurons with high firing rates, which are more likely to be isolated during spike sorting in electrophysiology. Additionally, imaging achieves superior spatial resolution by localising fluorescence changes to specific regions of interest whose position and depth are known. Thus, signals can be attributed more confidently to single neurons at well-defined locations.

So far, the temporal resolution of fluorescence signals has limited their ability to track intricate patterns of activity. However, new indicators with improved kinetics and sensitivity make it possible to image neural activity on biologically relevant timescales (Zhang et al., 2023). By allowing access to anatomically- and genetically-defined neural populations and subcellular compartments (e.g. dendrites) imaging neurons can uncover a richness of processing at the cellular and circuit level. Moreover, the ability to image neurons stably over long durations and chronically through cranial windows makes imaging useful for behavioural studies. Therefore, obtaining STRFs from imaging represents a particularly desirable methodology.

While imaging has been used widely to study auditory cortex (Issa et al., 2014; Kang and Kanold, 2024; Kline et al., 2023), it has not been used to fit STRFs. One potential reason is related to the low temporal resolution of fluorescence signals. Although cytosolic calcium levels return to baseline within 10-100 ms (Grienberger and Konnerth, 2012), indicator fluorescence decays over a longer timescale (half-decay times commonly > 250 ms) and is insufficient to resolve spikes reliably (Maravall et al., 2000; Sabatini et al., 2002; Zhang and Looger, 2024). Indicators are also subject to varying degrees of nonlinearity when integrating multiple spikes. GCaMP8 can linearly report

pairs of APs occurring at up to 50 Hz (Zhang et al., 2023), potentially allowing temporally precise spiking patterns to be reflected in the fluorescence trace. It is unclear whether such linearity is possible during sustained activity, such as when long-duration stimuli are presented. Steady-state cells may exhibit different calcium dynamics altogether, placing a further constraint on the information that can be contained in the fluorescence trace (Grienberger and Konnerth, 2012). Additionally, the relationship between spiking and fluorescence is less clear during typical population imaging where hundreds of neurons are imaged simultaneously. Under such conditions, low signal-to-noise ratio of individual neurons can mean that single or double AP events are not reflected in the fluorescence trace (Huang et al., 2021). Therefore, it is unclear whether STRFs can be derived reliably from two-photon calcium imaging data.

This chapter describes an end-to-end computational model designed to assess the feasibility of inferring STRFs from two-photon calcium imaging data. There exist well tested and reliable forward models that allow us to approximately simulate neural spiking activity from a prototypical receptive field using a Poisson process. Furthermore, spike trains can be reliably converted to synthetic fluorescence traces using a spike-to-fluorescence (S2F) model (Wei et al., 2020). We use both these approaches to simulate 'ground truth' neurons with known STRFs, spiking responses and calcium indicator traces. This allowed us to then test the validity of STRFs derived from the calcium indicator data, and compare them to STRFs derived from spiking data in the same neuron. Furthermore, we use this model to explore the reliability of STRFs derived from calcium florescence when we manipulate key variables that may be difficult to control experimentally, such as indicator kinetics, neural firing rate, stimulus type, and stimulus duration.

2.2 Methods

2.2.1 Stimuli

DRC stimuli (Fig. 2.1A) were constructed by randomly interleaving pure tone pips at 31 log-spaced frequencies between 2-64 kHz (1/6th octave spacing). Tone pips (25 ms duration, 5 ms half-cosine onset/offset ramps) were of constant amplitude and generated using a Poisson process with a rate of 3.5 tone pips per frequency band per second (Blake and Merzenich, 2002). The total duration of the DRC stimulus was 4 minutes.

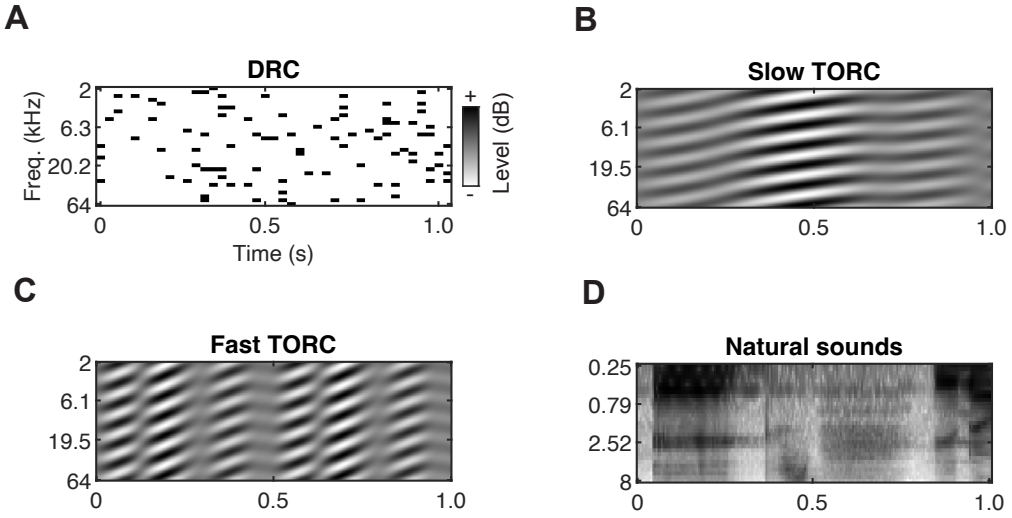


Figure 2.1: Spectrotemporal envelopes of stimuli used to model neuronal responses. A. DRC. **B.** Slow TORC. **C.** Fast TORC. **D.** Natural sounds. Note that synthetic stimuli range from 2 to 64 kHz and natural sounds range from 0.25 to 8 kHz. Both stimuli contain energy over a 5-octave range.

TORCs belong to the family of ripple stimuli and have been used extensively to characterise STRFs (Klein et al., 2000). Moving ripples were created by applying sinusoidal modulations along the spectrotemporal axes of an envelope composed of 100 random phase carrier frequencies per octave. Slow TORCs (Fig. 2.1B) contained 6 moving ripples with temporal modulation frequencies between 2 and 7 Hz (1 Hz spacing) and

spectral modulation frequencies up to 1.4 cycles/octave (0.2 cycles/octave spacing). A total of 30 TORCs each lasting 8 s were created and concatenated to form a single stimulus lasting 4 minutes. Fast TORCs (Fig. 2.1C) were created as above except moving ripples had temporal modulations between 4 and 24 Hz spaced by 4 Hz.

For the natural sound stimulus, sections from 24 recordings including speech from male and female speakers, vocalisations of different animal species and environmental sounds (e.g. water, chatter) were concatenated to form a single stimulus lasting approximately 4 minutes (frequency range 0.25-8 kHz, 44.1 kHz sample rate; Fig. 2.1D). Recordings were obtained from the BBC Sound Effects library[1] and The People's Speech Dataset (Galvez et al., 2021).

2.2.2 Simulated receptive fields

Spectrotemporal tuning of model neurons was described by synthetic RFs created by summing excitatory (Fig. 2.2A) and inhibitory subfields (Fig. 2.2B). To create subfields, spectral and temporal marginal functions were defined using Gaussians:

$$\mathrm{G}_x(x) = \frac{A_x}{\sigma_x\sqrt{2\pi}} \exp -\frac{1}{2}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right] \tag{2.1}$$

$$\mathrm{G}_h(h) = \frac{A_h}{\sigma_h\sqrt{2\pi}} \exp -\frac{1}{2}\left[\left(\frac{h-\mu_h}{\sigma_h}\right)^2\right]. \tag{2.2}$$

Peak response latency and best frequency were controlled by setting the means, μ_x and μ_h. Temporal and spectral response bandwidths were controlled using the variances, σ_x^2 and σ_h^2. The heights of spectral and temporal peaks were set using A_x and A_h, respectively. Gaussians $\mathrm{G}_x(x)$ and $\mathrm{G}_h(h)$ were sampled at discrete evenly spaced points to create spectral and temporal marginal vectors F (1×31) and H (1×40), respectively,

[1] https://sound-effects.bbcrewind.co.uk/

whose outer product gave 31×40 excitatory and inhibitory subfields ($\mathbf{R_e}$ and $\mathbf{R_i}$, respectively):

$$\mathbf{R_e} = \mathrm{F_e} \otimes \mathrm{H_e} \tag{2.3}$$

$$\mathbf{R_i} = \mathrm{F_i} \otimes \mathrm{H_i} \tag{2.4}$$

Subfields were then summed elementwise to produce the simulated RF, $\mathbf{R}$ (Fig. 2.2C):

$$\mathbf{R} = \mathbf{R_e} + \mathbf{R_i}. \tag{2.5}$$

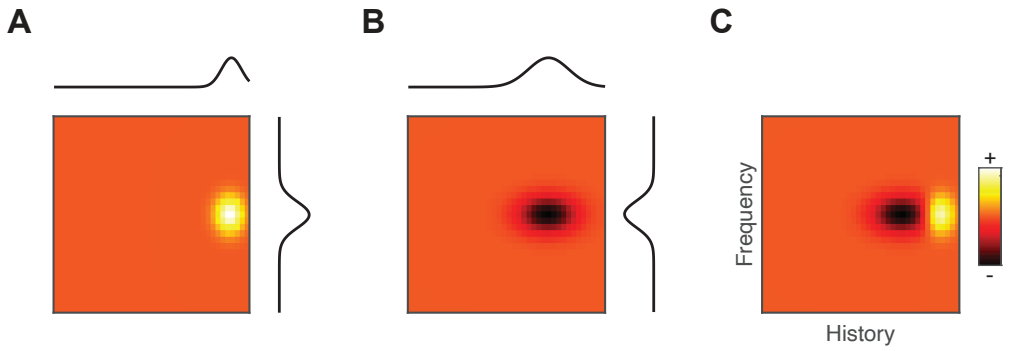


Figure 2.2: Construction of simulated RFs. A. Excitatory subfield, $\mathbf{R}_\mathrm{e}$. **B.** Inhibitory subfield, $\mathbf{R}_\mathrm{i}$. Subfields were formed by the outer product of spectral and temporal marginal vectors, $\mathrm{F_e}$ and $\mathrm{H_e}$ (annotated to the right of and above **A.**) and $\mathrm{F_i}$ and $\mathrm{H_i}$ (above and to right of **B.**) **C.** Excitatory and inhibitory subfields were pointwise-summed to produce $\mathbf{R}$.

Peak excitatory and inhibitory amplitudes were equal in synthetic RFs, unless otherwise stated. The frequency and history axes were arbitrary and dependent on the stimulus cochleagram with which the simulated RF was convolved. Here, stimulus cochleagrams contained energy between 2-64 kHz (DRCs and TORCs) or 0.25-8 kHz (natural sounds) and were binned at 5 ms resolution (see *Section 2.2.3*). Therefore, simulated RFs covered a 5 octave spectral range and 200 ms of stimulus history.

2.2.3 Cochleagrams

Input to the simulated RF was a frequency domain representation of the stimulus. This representation approximates processing of the waveform at the cochlea and is known as a cochleagram. A log-spaced, log-valued variety of the cochleagram was used, aiming to recapitulate the logarithmic spacing of different frequencies along the cochlear basilar membrane (Rahman et al., 2020). First, spectrograms were created using Hanning windows of 10 ms width and 5 ms overlap. Next, overlapping triangular windows with 31 log-spaced centre frequencies between 2 and 64 kHz were used to sum the amplitude of adjacent frequency bands (code adapted from melbank.m[2]). Finally, the power of each bin was calculated by converting the amplitude using $20\log(\cdot)$ and thresholding all values to a minimum of -40.

2.2.4 Linear-Nonlinear-Poisson spiking model

Spike trains were generated using a linear-nonlinear-Poisson model (Fig. 2.3A). In discrete time, the linear activation of the model at time t is given by a weighted sum over frequency channels and temporal lags:

$$r[t] = \sum_{f} \sum_{h} X_{\mathrm{f[t-h]}} R_{\mathrm{fh}} \tag{2.6}$$

where $X_{\mathrm{f[t-h]}}$ denotes the cochleagram evaluated at frequency channel f, time t, and time lag h, and R_{fh} denotes the corresponding element of the spectrotemporal receptive field matrix R.

The linear activation, $r[t]$, was then normalised and scaled to vary between $r_0 \pm r_{\max}$, where r_0 represented was spontaneous rate and $r_{\max}$ the maximum deviation in firing

[2]http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html

rate. This vector was then passed through a positive half-wave rectifying nonlinearity to give the instantaneous rate function, $\lambda[t]$:

$$\lambda[t] = \left| r_0 + r_{\max} \left[\frac{r[t] - \langle r[t] \rangle}{\max |r[t] - \langle r[t] \rangle|} \right] \right|^+ \tag{2.7}$$

where $\langle \cdot \rangle$ denoted an average over time. For 'high' spike rate neurons, $r_0 = 5$ and $r_{\max} = 60$ Hz, for 'medium' spiking neurons, $r_0 = 5$ and $r_{\max} = 30$ Hz and for 'low' spiking neurons $r_0 = 2$ and $r_{\max} = 30$ Hz. Spike rates were chosen to reflect neurophysiological findings in auditory cortex (David et al., 2009; deCharms et al., 1998; Hromádka et al., 2008). Assuming that spikes are independent of each other and driven exclusively by the stimulus, spike rates can be generated by an inhomogeneous Poisson process with rate $\lambda[t]$. For time bins of size Δt, spike counts, $z[t]$, are Poisson distributed (Meyer et al., 2017):

$$P\left(z[t] \,|\lambda[t]\right) = \frac{1}{z[t]!} e^{-\lambda[t]\Delta t} (\lambda[t]\Delta t)^{z[t]}. \tag{2.8}$$

While most applications of the linear-nonlinear-Poisson model are to fit recorded spiking data, the use of a forward model here meant that individual spikes had to be generated. The rate vector $\lambda[t]$ was defined at 5 ms resolution and described time-varying firing rates in Hz. Assuming only one spike can occur during a 1 ms interval, spikes were generated for each 1 ms bin with probability $\lambda[t]/1000$. Therefore, the resulting spike times, t_k, reflected a Poisson process with instantaneous probability proportional to the local rate. Unless otherwise specified, 10 stimulus repetitions were used to generate spiking responses.

2.2.5 Spike-to-fluorescence forward model

Two-photon calcium imaging captures fluorescence transients generated by calcium-sensitive indicators in response to neuronal activity (Grienberger and Konnerth, 2012).

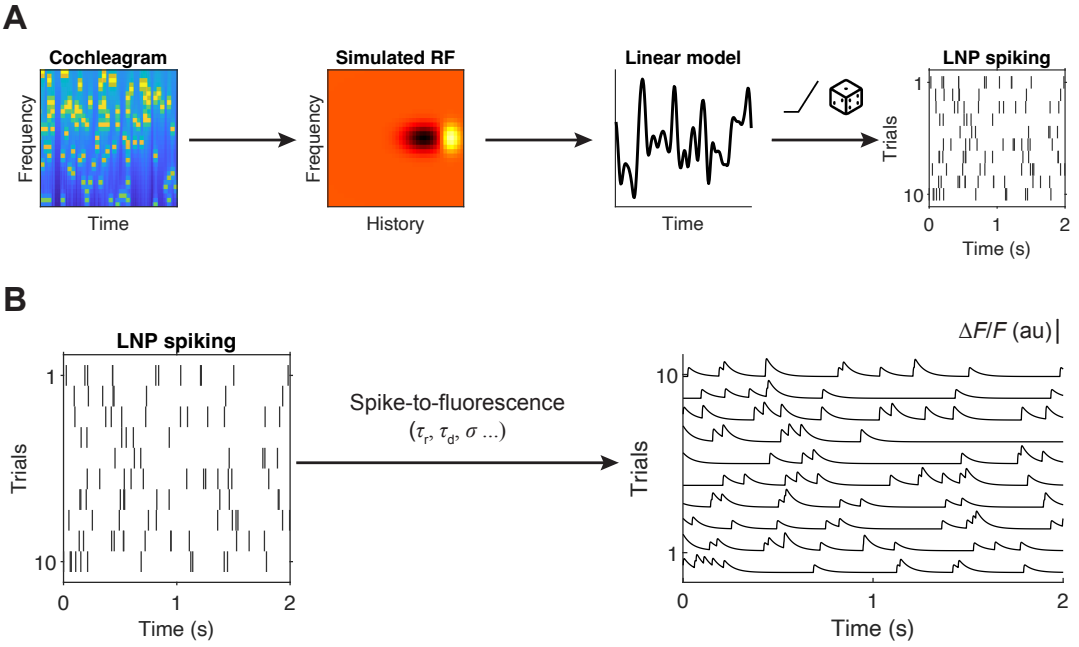


Figure 2.3: Linear-nonlinear-Poisson and spike-to-fluorescence model cascade. A. Stimulus cochleagrams were convolved with a simulated RF to give linear model output, $r[t]$, which was scaled and positive half-wave rectified to form the input to an inhomogeneous Poisson process, $\lambda[t]$. $n = 10$ trials of binned spike counts were sampled and converted to spike times. **B.** Spike trains were converted to synthetic fluorescence, $\Delta F/F$, using the spike-to-fluorescence forward model. $\Delta F/F$ is reported in arbitrary units (au) throughout. LNP = linear-nonlinear-Poisson.

APs are associated with a rise in intracellular calcium concentration leading to a sharp increase in fluorescence as per the emission spectrum of the indicator. Transients then decay in a stereotyped manner over the course of hundreds of milliseconds to seconds. Single APs can cause detectable excursions from baseline under optimal imaging conditions however the summation and dynamic range of overlapping transients is unclear, often exhibiting significant nonlinearity (Huang et al., 2021).

Here, I made use of a S2F forward model (Fig. 2.3B; Wei et al., 2020) which models fluorescence transients using a double exponential kernel:

$$c[t] = \sum_{t>t_k} \exp\left(-\frac{t-t_k}{\tau_\mathrm{d}}\right)\left[1-\exp\left(-\frac{t-t_k}{\tau_\mathrm{r}}\right)\right]. \tag{2.9}$$

For spikes occurring at times t_k, a latent variable $c[t]$ was generated by convolving the spike train with the double exponential kernel. τ_d and τ_r represented decay and rise times, respectively. $c[t]$ was then converted to synthetic fluorescence through a link function, as detailed below.

Linear: The following linear transformation generated the synthetic fluorescence trace, $\Delta F/F$:

$$\Delta F/F[t] = F_\mathrm{m} c[t] + F_0 + n[t]. \tag{2.10}$$

F_m is the maximum increase of fluorescence above baseline and F_0 is baseline fluorescence. Initial simulations (not shown) revealed no effect of a constant baseline fluorescence on estimated kernels, hence F_0 was set to 0 for subsequent simulations. $n \sim N(0, \sigma^2)$ is Gaussian distributed noise included in some simulations (see *Section 2.3.5*).

Sigmoidal: A sigmoidal nonlinearity was used to relate the latent variable $c[t]$ to

synthetic fluorescence:

$$\Delta F/F[t] = \frac{F_{\mathrm{m}}}{1 + \exp\left[-k\left(c[t] - c_{1/2}\right)\right]} + n[t] \tag{2.11}$$

where k is a steepness parameter and $c_{1/2}$ determines the half-activation level. Non-linearity parameters are usually fit to neurophysiological data consisting of spike times and accompanying fluorescence traces. Here, no such data was available so parameters were specified a priori. Values were selected to reflect the linearity of GCaMP8 for which linear integration is reported for up to 5-10 simultaneous APs (Zhang et al., 2023).

2.2.6 STRF estimation

STRFs provide a linear approximation of the mapping between the stimulus cochleagram, X_{fh}, and the trial-averaged neural response, $y[t]$. The STRF, k_{fh}, predicts a response, $\hat{y}[t]$ by the following:

$$\hat{y}[t] = \sum_{f,h} \mathrm{X}_{fht} k_{fh}. \tag{2.12}$$

Individual stimulus cochleagrams were first tensorised then concatenated to form a stimulus set tensor, X_{fht}. To obtain spiking STRFs, spike trains were binned at 5 ms and summed to give PSTHs to which kernels were fit. For fluorescence STRFs, the trial-averaged $\Delta F/F$ was used along with the stimulus cochleagrams binned at the corresponding 25 ms temporal resolution. Filter weights were estimated by minimising a sum-of-squares error term, E, which is a measure of the difference between $\hat{y}[t]$ and $y[t]$:

$$E = \sum_{t} \left(y[t] - \hat{y}[t]\right)^2 + \lambda \sum_{fh} {k_{fh}}^2 \tag{2.13}$$

Inseparable kernels were fit by linear regression with L2 regularisation (penalisation with the L2 norm, ${W_{fh}}^2$, with weighting λ) using a generalised linear model (Qian et al., 2013). Initial simulations (not shown) found that L2 regularisation was more resilient to added noise in fluorescence traces and yielded STRFs with more biologically plausible structure (i.e. smoother spectrotemporal modulations, fewer disjoint excitatory/inhibitory fields).

2.2.7 Tuning properties of STRFs

Tuning properties of kernels including best frequency, spectral bandwidth, peak response latency and temporal integration bandwidth were measured to enable comparisons between simulated RFs, spiking STRFs and fluorescence STRFs. Kernels were interpolated (interp2.m with 'spline' interpolation, MATLAB) and averaged along temporal and spectral axes to generate spectral and temporal tuning profiles, respectively (see Fig. 2.6A). The frequency band corresponding to the maximum value of the spectral profile was assigned best frequency, and the frequency bandwidth was defined as the range of the spectral profile greater than half the maximum. Temporal tuning was calculated similarly, with peak latency assigned to the time bin containing the maximum value of the temporal profile. Temporal bandwidth was defined as the range for which the temporal profile exceeded half the maximum.

To investigate the effect of the S2F transformation on the balance between excitatory and inhibitory regions of the RF, I calculated the excitatory/inhibitory index (EI) for each STRF (Harper et al., 2016). First, excitatory weight was calculated by thresholding the kernel to ≥ 0 and summing the remaining coefficients. Inhibitory weight was calculated similarly except a threshold of ≤ 0 was used and the modulus of the sum was taken. EI was then calculated as follows:

$$EI = \frac{E - I}{E + I} \tag{2.14}$$

2.3 Results

To understand how well STRFs can be estimated using two-photon calcium imaging data, I simulated action potential responses of model neurons to a variety of sound stimuli using an linear-nonlinear-Poisson spiking model. The stimulus selectivity of model neurons was pre-defined with linear STRFs that model those commonly observed in auditory cortical neurons. Each STRF spanned a 5 octave range and integrated over 200 ms of stimulus history. The convolution product between stimulus cochleagrams and simulated RF was normalised, scaled and half-wave rectified to form the input to a Poisson process generating spikes (Fig. 2.3A). I then used a S2F forward model (Wei et al., 2020) to convert these simulated spike trains to the type of florescence calcium traces recorded with two-photon imaging. Spike times were convolved with a double exponential kernel to generate synthetic fluorescence traces (Fig 2.3B). I estimated each model neuron's STRF independently using the spike post-stimulus time histograms (PSTHs) and from trial-averaged $\Delta F/F$. Comparing the STRF fits based on spiking and florescence data to each other and to the 'ground truth' pre-defined RF for each model neuron allowed me to understand how well the spectrotemporal tuning of single neurons can be estimated from two-photon calcium imaging data, and how they depend on a number of experimental variables such as indicator kinetics, spike rate, receptive field shapes, and stimulus parameters

2.3.1 Fluorescence STRFs

The simulated receptive field of a model neuron, along with the STRFs fit to trial-averaged simulated spiking responses ($\mathrm{STRF_s}$) and convolved $\Delta F/F$ traces ($\mathrm{STRF_f}$) are shown (Fig. 2.4A). This example is a simple excitatory RF, and results in corresponding simple excitatory $\mathrm{STRF_s}$ and $\mathrm{STRF_f}$. A single excitatory subfield with a best frequency of $\sim$ 10 kHz was seen in both STRFs, however this subfield ex-

tended significantly further into stimulus history in STRF_f compared to STRF_s. The florescence-based STRF_f resembled a version of STRF_s with longer peak latency and broader temporal bandwidth. Since kernels fit to $\Delta F/F$ contained energy at points earlier in stimulus history compared to spiking kernels, they were fit to 1000 ms of history.

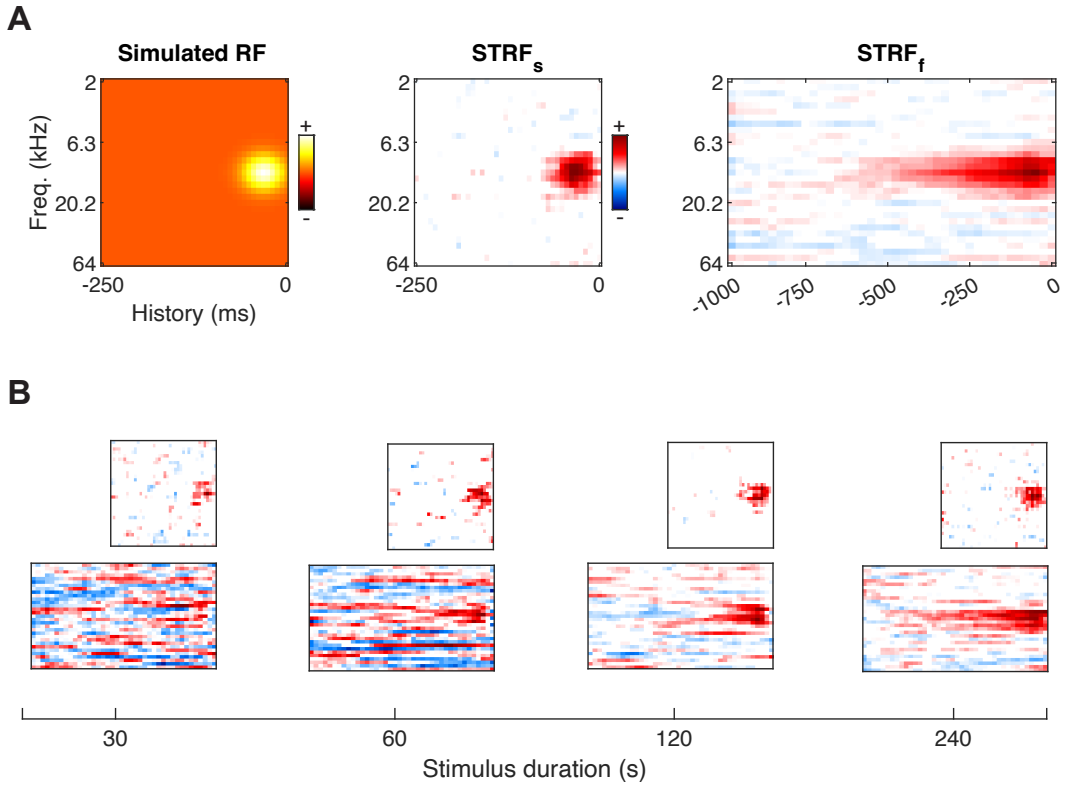


Figure 2.4: STRF_f (fit to trial-averaged simulated fluorescence) structure resembles STRF_s (fit to trial-averaged simulated spiking) and simulated RF. A. Simple excitatory simulated RF (left), corresponding STRF_s (middle) and STRF_f (right). Note history axis for STRF_f extends to -1000 ms to account for broader temporal tuning. **B.** STRF_s fit to 1 repeat of a variable duration stimulus. Shorter DRCs were obtained by truncating the full-length stimulus (240 s) to preserve 0 to x seconds (x = 30, 60 or 120). STRFs in **A** were fit to 10 stimulus repeats.

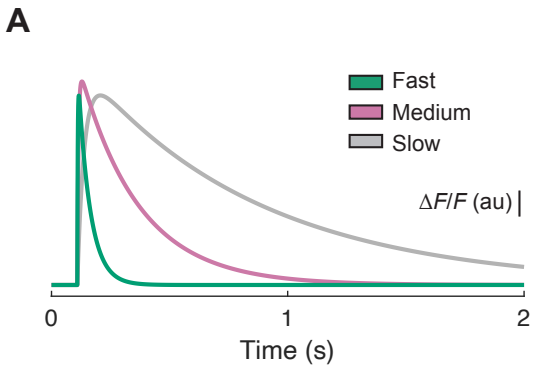
Next, I investigated the minimum stimulus duration required to fit this STRF reliably by modelling responses to a single repeat of a DRC that varied in duration (Fig. 2.4B). While the excitatory component of the simulated RF could be observed in STRF_s from

a 30 s DRC, $STRF_f$ required 60-120 s of stimulus to recover it. $STRF_f$ fit to 1 repeat of a 60 s DRC captured excitation at the BF of the neuron, however it included more noise, visible as a scattering of small and weak excitatory and inhibitory regions outside the single excitatory region of the simulated RF. Compared to spiking, $\Delta F/F$ required a significantly longer stimulus duration to yield an STRF. Since the decay time of GCaMP indicators is long compared to the temporal structure of the DRC, features following the feature of interest can be captured. Therefore, noisy structure in $STRF_f$ may be a result of stimulus structure captured during the decay of the transient. Inclusion of temporally correlated features necessitates a longer DRC exploring more of stimulus space to average out random patterns which follow the feature of interest.

2.3.2 Indicator kinetics

Iterative optimisation of the GCaMP family has yielded indicators with biologically relevant, millisecond-order response times (Chen et al., 2013; Zhang et al., 2023). Since fine patterns of temporal activity are paramount in estimating STRFs, indicator kinetics are likely to represent a critical variable contributing to the STRFs observed for simulated florescence data in Fig 2.4. To understand how the kinetics influence STRF estimates, I modelled fluorescence traces using 'fast', 'medium' and 'slow' kinetics chosen to approximate the properties of GCaMP8f, GCaMP8s and GCaMP6s, respectively (Fig. 2.5A, B).

Five simulated RFs representing canonical primary auditory cortex STRFs were used to simulate responses, and STRFs were then fit to the resulting spiking and florescent traces (Fig. 2.5C. As predicted, the discrepancy between temporal tuning of $STRF_f$ and the simulated RF was related to kinetics. This was demonstrated clearly for RFs containing simple excitatory and frequency sidebands (first and second rows of Fig. 2.5C). Slower kinetics led to greater peak latencies and temporal bandwidths, with $STRF_f$ estimated using the slow indicator still containing significant energy at -1000



B

Indicator	τ_r, ms	τ_d, ms
Fast	2	50
Medium	5	250
Slow	30	750

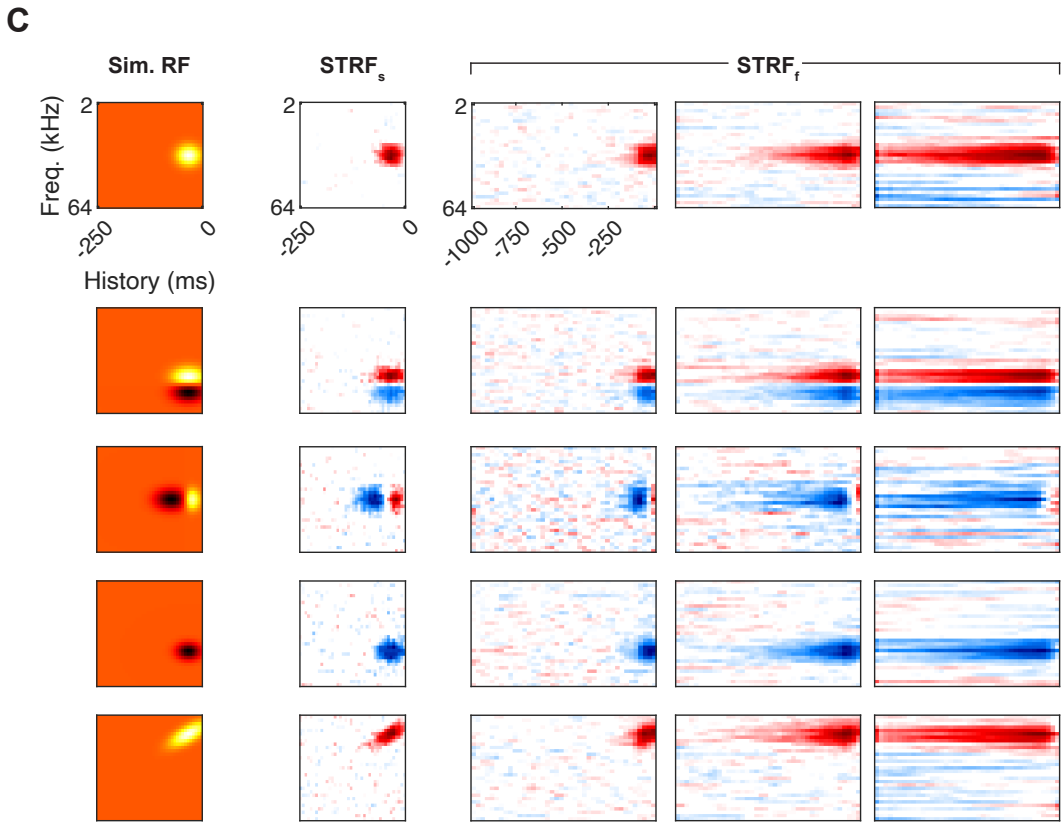


Figure 2.5: Indicator kinetics determine the extent of STRF smoothing. A. Simulated fluorescence transients evoked by a single AP at 100 ms. 'Fast', 'medium' and 'slow' indicator types were modelled to reflect GCaMP8f, 8 s and 6 s, respectively. **B.** Half-rise, τ_r], and half-decay times, τ_d, used in S2F transformation (Chen et al., 2013; Zhang et al., 2023). **C.** Five types of simulated RF are shown: (from top to bottom) simple excitation, frequency sideband inhibition, temporal sideband inhibition, simple inhibition and frequency sweep.

ms. In fact, non-zero weights in this time bin suggest that instantaneous $\Delta F/F$ integrates over even longer than 1000 ms. Considering this impact on temporal tuning, I wanted to understand how RFs containing more temporally complex elements might be affected by indicator kinetics. To do so, I generated a temporal sideband RF (Fig. 2.5C, third row) and found that the component more recent in history (in this case, the excitatory lobe) was attenuated and temporally shortened in $\mathrm{STRF_f}$ while the less recent component (the inhibitory lobe) showed a longer duration with slower indicators. Interestingly, the inhibitory region was temporally broadened and fragmented in the slow indicator condition, becoming indistinguishable from noise further back in stimulus history. Whether fragmentation of STRF components arose from temporally correlated structure in the DRC stimulus or some other artefact is unclear. Some degree of broadening along the frequency axis of $\mathrm{STRF_f}$ was also observed for the simple inhibitory RF in addition to lengthening along the history axis (Fig. 2.5C, fourth row). The frequency sweep RF (Fig. 2.5C, fifth row) was also temporally elongated, leading to sweep selectivity not being captured in $\mathrm{STRF_f}$ even for the fastest indicator.

2.3.3 Tuning properties

STRFs are commonly used to derive neuronal tuning properties including best frequency, spectral bandwidth, response latency, and temporal bandwidth. To quantify the effect of the S2F transformation, 60 simple excitatory RFs with different tuning properties were created. The mean and variance of the spectral and temporal Gaussian functions defining the RF were sampled from bounded uniform distributions (see *Section 2.2.2*). Possible values ranged between approximately 20-70 ms for peak latency, 20-80 ms for temporal bandwidth, 2-64 kHz for best frequency and 0.5-1.5 octaves for spectral bandwidth.

Tuning properties of $\mathrm{STRF_f}$ were compared to those of $\mathrm{STRF_s}$, as I have shown that the latter reliably recovers properties of the simulated RF (Fig. 2.5C). Representa-

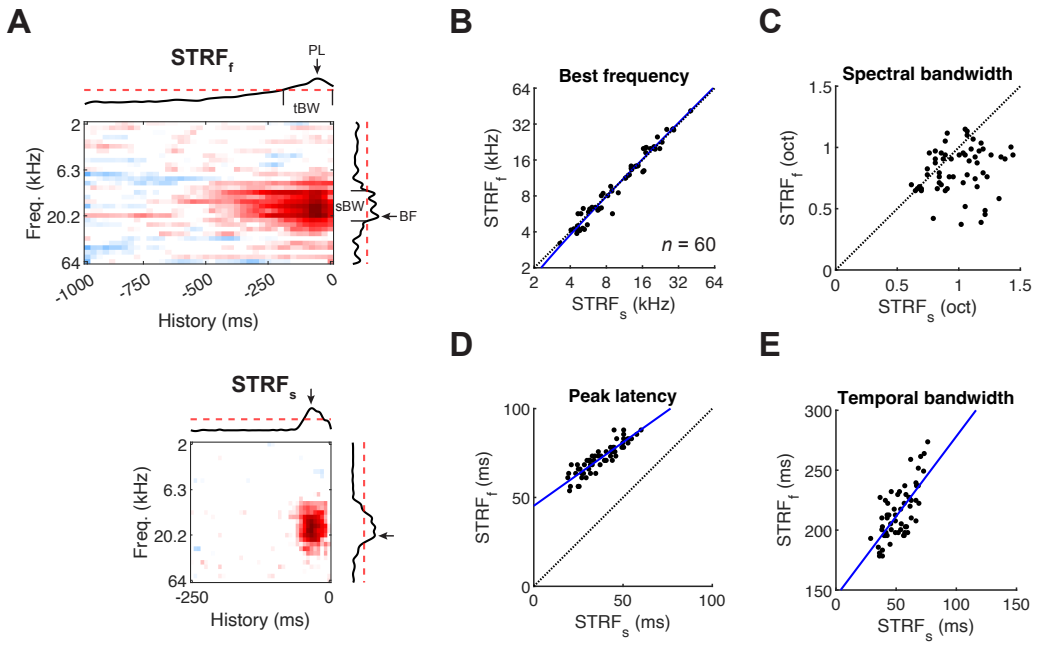


Figure 2.6: Tuning properties of fluorescence and spiking STRFs. Simple excitatory simulated RFs ($n = 60$) were created and tuning properties compared between resulting $\mathrm{STRF_s}$ and $\mathrm{STRF_f}$. **A.** Spectrotemporal tuning shown on exemplar $\mathrm{STRF_f}$ (top) and $\mathrm{STRF_s}$ (bottom). Temporal and spectral tuning profiles are plotted above and next to kernels, respectively. Best frequency and peak latency (black arrows) and half-maxima of profiles (red dashed lines) are annotated. Spectral and temporal bandwidths are defined by the points of intersection between the half-max line and tuning profile. BF = best frequency, sBW = spectral bandwidth, PL = peak latency, tBW = temporal bandwidth. Comparison of tuning properties between STRFs and STRFf: **B.** best frequency, **C.** bandwidth, **D.** latency and **C.** temporal bandwidth. Best fit by linear regression is shown by solid blue lines. Note axes **E** are not equal, with the best fit substantially above the line of unity.

tive $STRF_f$ and $STRF_s$ for a simple excitatory RF are shown with tuning properties annotated (Fig. 2.6A). $STRF_f$ best frequency was strongly correlated with $STRF_s$ (Spearman's rank correlation coefficient, $\rho = 0.97$, $p = 1.6 \times 10^{-38}$), with most neurons showing the same best frequency in these two fits (Fig 2.6B). Median $STRF_f$ best frequency (11.3 kHz, IQR 6.0-19.9) did not differ significantly from $STRF_s$ (11.6 kHz, IQR 5.7-17.1; Wilcoxon signed rank test, $p = 0.75$). Spectral bandwidth was weakly correlated between spike-based and florescence-based STRFs ($\rho = 0.26$, $p = 0.04$). However, $STRF_f$ bandwidths (median 0.85 oct, IQR 0.68-0.95) were biased towards lower values than $STRF_s$ (1.0 oct, IQR 0.84-1.2; Wilcoxon signed rank test, $p = 1.1 \times 10^{-5}$).

Temporal tuning properties were also affected by the S2F transformation. The median latency of peak excitation was significantly lower at 36.2 ms (IQR 28.1-45.2) in $STRF_s$ compared to 70.9 ms (IQR 66.0-78.2) in $STRF_f$ (Wilcoxon signed rank test, $p = 1.6 \times 10^{-11}$). Although different, the peak latency was correlated across model neurons (Spearman's rank correlation coefficient, $\rho = 0.91$, $p = 5.1 \times 10^{-24}$). $STRF_f$ was binned at a 25 ms resolution so latencies considerably less than a single bin width were not theoretically possible. However, given the half-rise time of the indicator ($\tau_r = 5$ ms), it is surprising that a mean difference of over 30 ms in peak latency was observed between STRFs. Additionally, the slope of the best fit was less than 1 ($\beta_1 = 0.71$, $p = 8.8 \times 10^{-22}$), suggesting that predictions of latency may converge for arbitrarily high values and that kernels fit to fluorescence data diverge most for shorter latencies. The point of convergence is, however, beyond the limits of temporal integration reported in primary auditory cortical neurons.

Finally, temporal bandwidth provides a measure of the temporal precision of the main excitatory component of a simple STRF. The median temporal bandwidth was significantly shorter in $STRF_s$ (52.3 ms, IQR 42.0-61.3) fits than in $STRF_f$ fits (210.1 ms, IQR 197.9-227.3; Wilcoxon signed rank test, $p = 1.6 \times 10^{-11}$). Temporal bandwidths

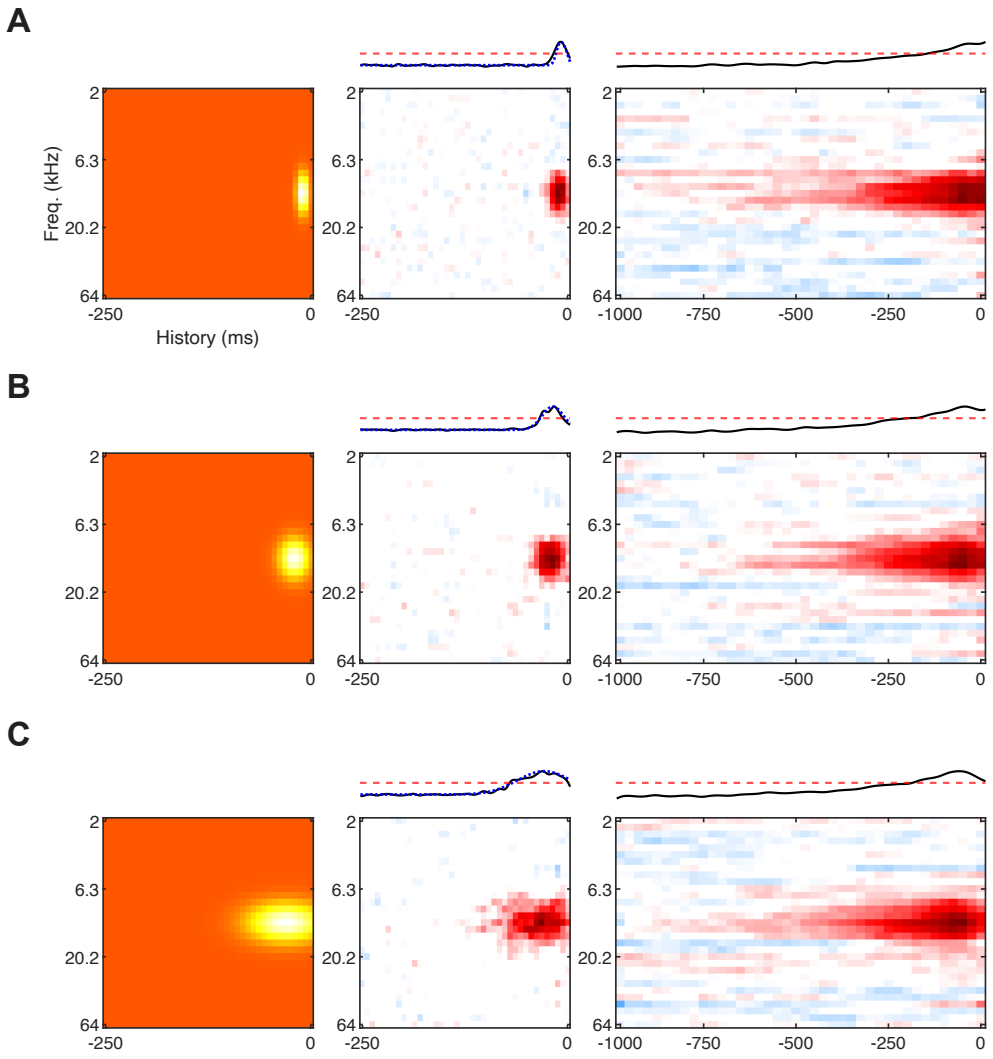


Figure 2.7: $\mathbf{STRF_s}$ and $\mathbf{STRF_f}$ temporal tuning. A. low latency and narrow temporal bandwidth, **B.** medium latency and temporal bandwidth and **C.** long latency and broad temporal bandwidth simulated RFs. Each row contains the simulated RF (left), $\mathrm{STRF_s}$ (middle) and $\mathrm{STRF_f}$ (right). Curves show temporal profiles of simulated RFs (blue dotted) and STRFs (black). Red dashed lines are plotted at half-maximum of temporal profiles.

were positively correlated between $STRF_s$ and $STRF_f$ (Spearman's rank correlation coefficient, $\rho = 0.67$, $p = 5.3 \times 10^{-9}$), despite $STRF_s$ bandwidths being shorter by over 150 ms on average. To aid visualisation of how temporal tuning compares across $STRF_s$ and $STRF_f$, I created simulated RFs which varied systematically in their latency and temporal integration (Fig. 2.7). These simulations show that STRFs captured millisecond-order differences in best frequency tuning, while this information was lost in $STRF_f$. This almost certainly occurs because $STRF_f$ fits to the slower temporal properties of fluorescence transients.

Although simulated fluorescence responses were obtained by a linear convolution of the spike train (produced using a linear STRF) with a double exponential kernel (see *Section 2.2.5*), the structure of $STRF_f$ was not equivalent to the result of the same convolution applied along the history axis of $STRF_s$. This implies that the introduction of variability at some point in the model cascade was a source of these differences. This could stem from the spiking process, the spike-to-fluorescence transformation, or the STRF fitting procedure. While the spike-to-fluorescence convolution was a deterministic process, it used an infinite impulse response filter truncated to match the length of the simulated response. Following such a convolution, the value of $\Delta F/F$ in later time bins had a small contribution from the earliest spikes and spiking noise in the response. This enables past as well as present spike train variability to influence the fluorescence response, in effect introducing a degree of history-dependence not present in the simulated spike train. Additionally, the random process for weight initialisation introduces variability at the fitting procedure.

While it is unclear how the combination of these sources of noise and variability may systematically bias $STRF_f$, my simulations showed that $STRF_f$ tended to exhibit narrower spectral bandwidths than $STRF_s$ (Fig. 2.6C). While this could be related to noise as described, it may also be a result of the temporal elongation of receptive field

features in $\mathrm{STRF_f}$. Briefly, spectral bandwidths were obtained by taking the mean average of the STRF along the temporal axis, and measuring the width of the resulting spectral tuning profile above its half-maximum value (Fig. 2.6A). In the case of $\mathrm{STRF_f}$, excitatory regions extended further along the history axis compared to $\mathrm{STRF_s}$ (e.g. compare second and third column in Fig. 2.7). This may have resulted in 'peakier' tuning profiles obtained by averaging along the history axis, in turn yielding narrower tuning bandwidths.

2.3.4 Excitatory-inhibitory balance

As significant changes in structure were observed in temporal sideband STRFs (Fig. 2.5C, row 3), I wanted to further characterise how temporally apposed antagonistic RF components are affected by the S2F transformation. To do so, two types of temporal sideband RF were simulated: inhibition preceding excitation, and excitation preceding inhibition (Fig. 2.8A). Simulated RFs were varied by sampling excitatory and inhibitory component amplitudes from a bounded uniform distribution. To quantify excitatory-inhibitory balance within a kernel, I calculated the excitatory-inhibitory index, EI (Harper et al., 2016). EI varies between -1 and 1 with the lower and upper bounds representing kernels with purely negative or positive coefficients, respectively. An EI of 0 represents a kernel with equally balanced excitatory and inhibitory weights. Across simulated RFs, EI ranged between approximately -0.8 to 0.8, dependent on the relative amplitude and size of excitatory and inhibitory components (see *Section 2.2.7*).

$\mathrm{STRF_s}$ closely correlated with the excitatory-inhibitory properties of simulated RFs, for both delayed inhibition (Fig. 2.8B; Spearman's rank correlation coefficient, $\rho = 0.96$, $p = 0$) and delayed excitation (Fig. 2.8D; $\rho = 0.96$, $p = 0$). Note that p-values reported as zero were below the floating-point threshold in MATLAB implying a true value of $p < 2 \times 10^{-308}$. In delayed inhibition RFs, median EI_s (-0.39, IQR -0.56 to

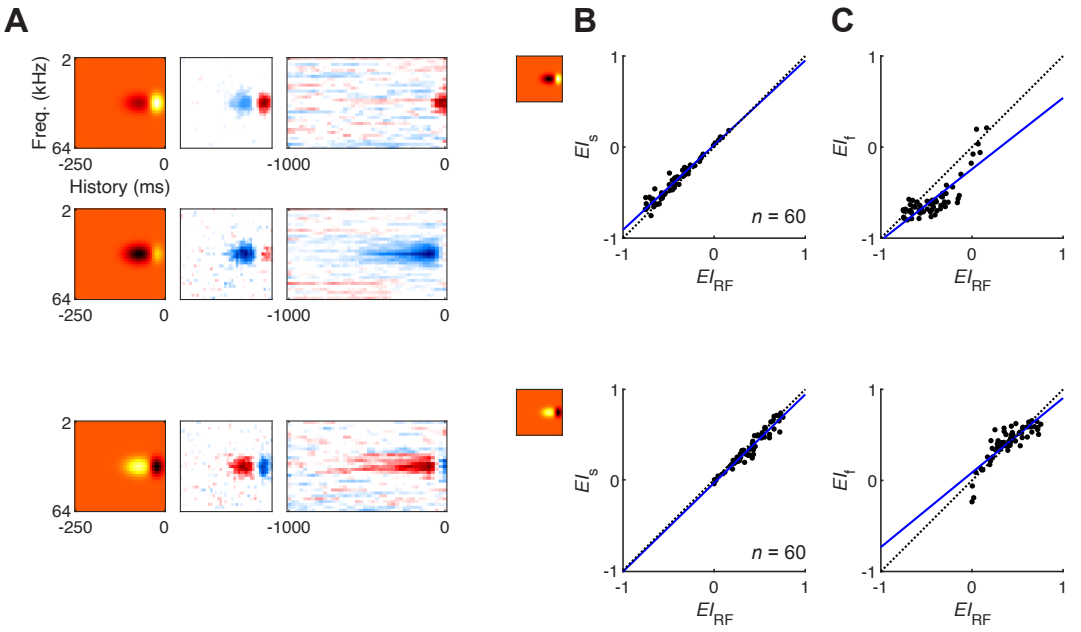


Figure 2.8: STRF excitatory/inhibitory balance is affected by spike-to-fluorescence transformation. A. Temporal sideband RFs were simulated excitatory and inhibitory component amplitudes sampled from a uniform distribution. **B.** Scatter plots of EI in $\mathrm{STRF_s}$ compared to simulated RF ($n = 60$; **B**, **D**) and $\mathrm{STRF_f}$ compared to simulated RF ($n = 60$; **C**, **E**). Insets: delayed inhibition (**B**, **C**) and delayed excitation RFs (**D**, **E**; note reversal of subfield order).

-0.24) was significantly greater than EI_{RF} (-0.44, IQR -0.61 to -0.30; Wilcoxon sign rank test, $p = 2.48 \times 10^{-9}$). In contrast, median EI_{s} for delayed excitation (0.34, IQR 0.23 to 0.50) was less than EI_{RF} (0.42, IQR 0.28 to 0.55; $p = 4.25 \times 10^{-7}$). One may expect inhibitory components to be underrepresented in $\mathrm{STRF_s}$ due to a floor effect arising from the non-negativity of spiking, however this was only observed in delayed inhibition where EI_{s} was more positive than EI_{RF}.

$\mathrm{STRF_f}$ excitatory-inhibitory balance also correlated well with both simulated RFs (Fig. 2.8C, E): delayed inhibition ($\rho = 0.73$, $p = 0$) and delayed excitation ($\rho = 0.83$, $p = 0$). Despite positive correlations, visual inspection of the scatter plots indicated that the mapping between EI_{RF} and EI_{f} was not linear. Each plot revealed half of what appeared to be a sigmoidal dependency, where the stronger component in the simulated RF was amplified in $\mathrm{STRF_f}$ (see exemplar kernels in Fig. 2.8A). For delayed inhibition, mean EI_{f} (-0.63, IQR -0.70 to -0.57) was significantly smaller than EI_{RF} (-0.44, IQR -0.61 to -0.30; $p = 5.29 \times 10^{-9}$), indicating that kernels fit to fluorescence overestimate the negative component in this RF type. The opposite was observed in delayed excitation (mean $EI_{\mathrm{f}} = 0.45$, IQR 0.36 to 0.56, $EI_{\mathrm{RF}} = 0.42$, IQR 0.28 to 0.55) but was not significant ($p = 0.27$). Therefore, $\mathrm{STRF_s}$ and $\mathrm{STRF_f}$ behaved in opposite ways: $\mathrm{STRF_s}$ overestimated the excitatory component in delayed inhibition RFs and overestimated the inhibitory component in delayed excitation RFs, and vice versa for $\mathrm{STRF_f}$.

2.3.5 Indicator noise, nonlinearity and spike rate

While simulations using a noiseless, linear S2F transformation are valuable to characterise fundamental differences between spiking and fluorescence STRFs, they do not represent the realities of imaging neurons *in vivo*. Importantly, indicator nonlinearities, measurement noise and neuronal firing rates complicate the interpretation of fluorescence data. To account for the dynamic range of GCaMP indicators, I introduced a

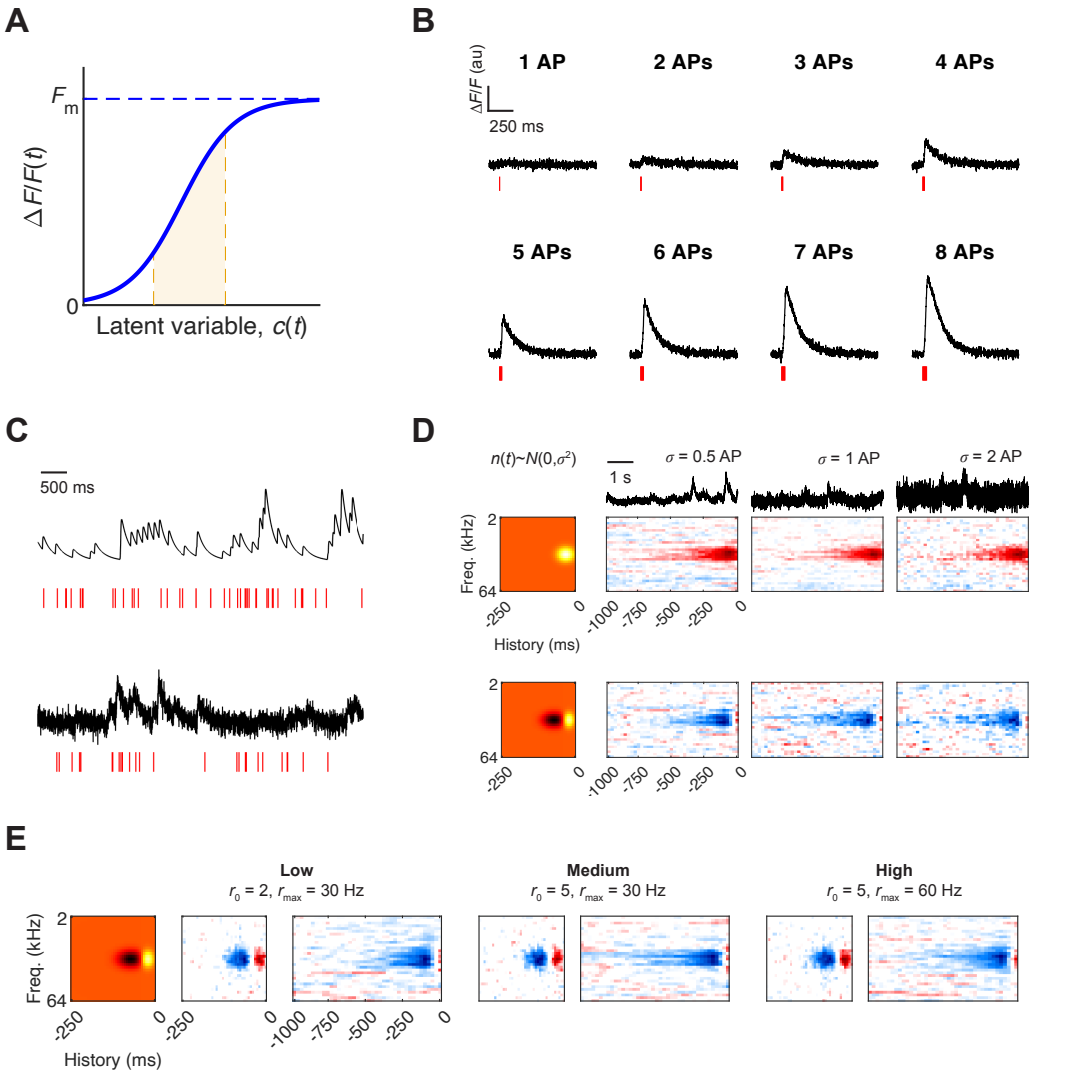


Figure 2.9: Incorporation of indicator nonlinearity, measurement noise and neuronal spike rate into spike-to-fluorescence model. A. S2F sigmoidal nonlinearity. Input is formed by $c(t)$, the convolution product of the model spike train with the fluorescence kernel. F_{m} (blue dotted line) is the maximum fluorescence, orange shading represents linear summation. **B.** Zero-mean Gaussian white noise added to $\Delta F/F$ generated by a nonlinear S2F model made 1 AP events undetectable. Red ticks = APs occurring at 3 ms intervals starting at 100 ms. **C.** Single AP events and couplets are obscured in noisy traces. **D.** $\mathrm{STRF_f}$ are resilient to large amplitude noise. Noise standard deviation, σ, is given in multiples of peak $\Delta F/F$ evoked by a single AP. **E.** Neuronal spike rate has little effect on $\mathrm{STRF_s}$ and $\mathrm{STRF_f}$ with similar structure observed across low, medium and high spike rate conditions. Vertical scale bar in **B** applies to **C** and **D**.

static nonlinearity in the form of a sigmoidal link function in the S2F transformation (Fig. 2.9A). This had a weak thresholding effect at 1-2 APs and saturated at ~ 10 APs, with largely linear summation in between, more closely approximating the non-linear indicator response observed experimentally with increased spike rates (Huang et al., 2021).

Additionally, STRF estimation must overcome significant noise which, in imaging, may be photonic, autofluorescence or related to shifts in cytosolic calcium unrelated to neural activity. Electrophysiology data are typically thresholded to identify APs and smoothed by construction of a PSTH. By contrast, two-photon data are only trial-averaged following pre-processing and thus may retain noisy deviations which occur on timescales faster than transients themselves. There is evidence that a majority of 1 AP events are missed in fluorescence traces under imaging conditions typical of large scale microscopy (Huang et al., 2021). Therefore, defining detectability as a rise to > 2 standard deviations above the noise floor in response to an AP, I simulated noisy fluorescence traces by adding zero-mean Gaussian noise with standard deviation equal to half the peak $\Delta F/F$ evoked by a single AP ($\sigma = 0.5$ AP). By definition, a single AP caused an average peak $\Delta F/F$ of 2 standard deviations above baseline and was undetectable. To gauge the effect of imaging noise alone, no additional noise was added to the spike-generating process.

The effect of measurement noise combined with the sigmoidal nonlinearity on the detection of spiking events is shown (Fig. 2.9B, C). Interestingly, $\mathrm{STRF_f}$ was resistant under significantly noisy conditions and continued to exhibit clear structure (Fig. 2.9D). Unsurprisingly, kernels became increasingly noisy at higher noise levels but not to the extent that tuning properties were completely obscured. These results demonstrate that STRFs estimated by regularised linear regression can overcome significant noise, provided neuronal responses are stationary and linear. It follows that significantly

noisy STRFs *in vivo* are unlikely to be purely related to imaging noise.

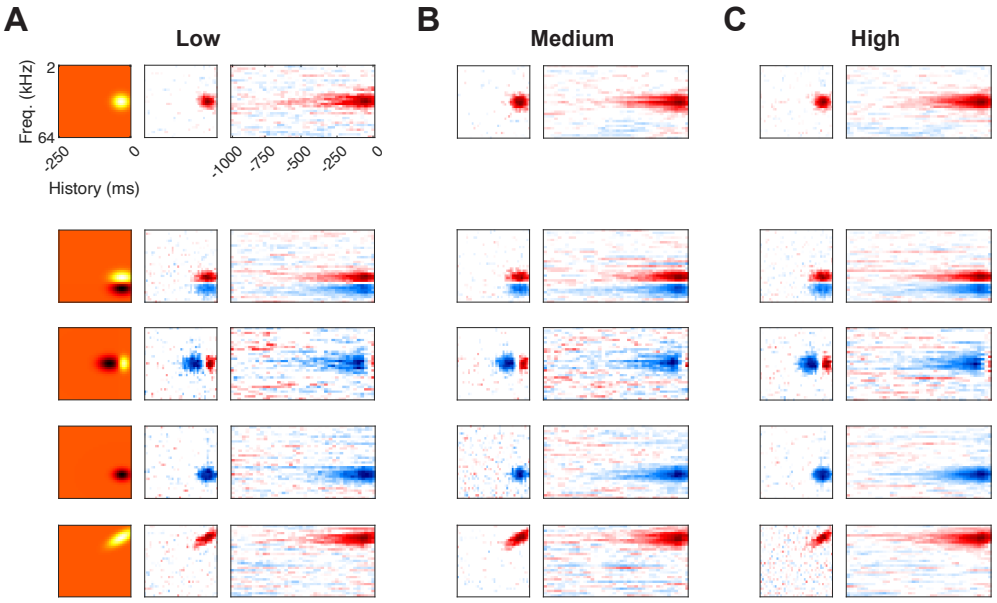


Figure 2.10: STRFs are invariant to spike rate. Responses were generated using a sigmoidal S2F transformation with noise ($\sigma = 0.5$) AP). **A.** Low, **B.** medium and **C.** high spike rate neurons were simulated. Spike rate parameters are equal to those used in Figure 2.9E.

Spiking dynamics are an additional determinant of whether neuronal activity can be captured reliably by GCaMP indicators, with potentially limited variance explained by a fluorescence trace associated with a robust, dynamic spike train. Auditory cortical neurons are driven strongly by their preferred stimuli (Wang et al., 2005) however tend to exhibit slower rates than visual cortical cells (Carandini, 2004). Additionally, experimental variables including anaesthetic preparation, behavioural context and stimulus choice can determine the types of spiking responses observed (Wang, 2007). Here, I systematically varied the spontaneous, r_0, and maximum spike rate, $r_{\max}$, of model neurons and used a sigmoidal S2F transformation. Varying spike rate in a noiseless sigmoidal S2F simulation did not affect STRFs (Fig. 2.9E). Similarly, the structure of $\mathrm{STRF_f}$ fit to low, medium and high spike rate neurons with added measurement noise ($\sigma = 0.5$ AP) did not show significant variation across all 5 RF types (Fig. 2.10A-C).

These results may have arisen from the fact that nonlinearity parameters were set to reflect GCaMP8s, which is a relatively linear indicator (Rupprecht et al., 2025). Further, variation in mean rates of the Poisson spiking process were unlikely to be representative of the different activation patterns exhibited by slow- and fast-spiking cortical neurons (Paninski et al., 2007).

2.3.6 Stimulus class

A large variety of synthetic and natural stimuli have been used to characterise spectrotemporal tuning with the choice often constrained by the requirements of the fitting procedure (C. Atencio and Schreiner, 2013). DRC stimuli (Fig. 2.1A), that have been used in simulations so far, are Gaussian-distributed and uncorrelated, making them suitable for spike-triggered techniques. Here, kernels were fit using regularised linear regression, removing the requirement to use uncorrelated stimuli. STRFs fit to two-photon calcium imaging data may exhibit dependence on the type and features of chosen stimuli. Temporal structure is likely to have a significant effect, with locked responses to precise or fast-varying features potentially lost in the fluorescence trace. To investigate this, I simulated responses to various TORC and natural sound stimuli. TORCs were made by summing several moving spectrotemporal "ripples", the auditory version of drifting gratings used in visual cortex studies. Slow TORCs (Fig. 2.1B), with maximum temporal modulation speeds of 7 Hz were used in conjunction with fast TORCs (temporal modulations up to 24 Hz, typical of stimuli used in electrophysiology; Fig. 2.1C). Natural sounds (Fig. 2.1D) used here were not optimised to be ethologically relevant, but instead were a more general set of natural recordings, including a mixture of animal and environmental sounds (e.g. footsteps).

Slow TORCs yielded STRFs with mixed success (Fig. 2.11A). $\mathrm{STRF_f}$ recovered simple excitation, frequency sideband and simple inhibition filters albeit with spurious structure at higher frequency bands. Temporal smoothing was less pronounced compared

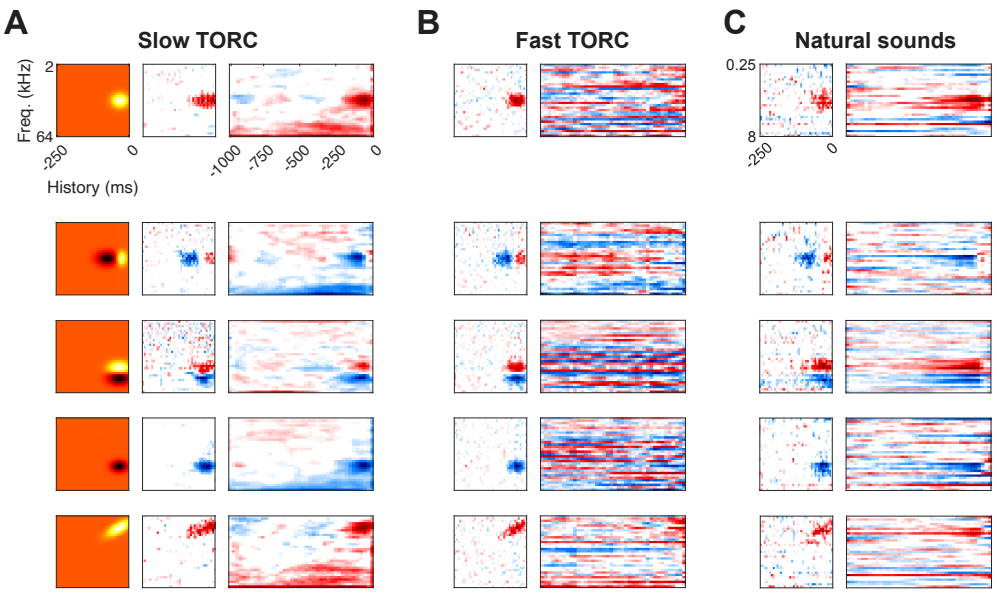


Figure 2.11: Effect of stimulus class on STRFs. STRFs obtained using responses to **A.** slow TORCs, **B.** fast TORCs and **C.** natural sounds.

to DRC estimates, potentially due to better sampling of temporal tuning properties by the periodic structure in TORCs. Fast TORCs could not recover any meaningful structure in STRF_f (Fig. 2.11B). Natural sounds provided reasonably good estimates with STRF_f recovering much of the original RF structure albeit with significantly background noise (Fig. 2.11C). This could have been due to the stimulus set or method of regularisation, among other possibilities. The adoption of L1- over L2-regularisation is likely to eliminate many of the non-zero STRF weights. STRFs provided a good match across all RF types. Natural sound and slow TORC stimuli may be well suited to recover spectrotemporal tuning from imaging data, partly because they are biased towards slower temporal modulations. Additionally, naturalistic stimuli contain features capable of driving robust responses *in vivo* not accounted for by the linear model used here. Nothwithstanding, in these simulations, DRCs yielded STRFs whose structure best matched that of the simulated RFs.

2.4 Discussion

Receptive field models provide a means for quantifying and predicting neuronal responses in sensory systems, including visual, auditory and somatosensory pathways. Auditory STRFs are typically fit using electrophysiology in awake or anaesthetised animals but are not exclusive to extracellular recordings. Other neural measures including electrocorticography (Hullett et al., 2016), local field potential (Eggermont et al., 2011; Francis et al., 2018) and whole-cell recording (Machens et al., 2004) have also been used successfully, although evidence supporting the suitability of STRFs in lower temporal resolution modalities is largely lacking. In one study, receptive fields were fit to fMRI data from humans presented with spectrotemporal modulation band-limited speech (Venezia et al., 2021). The 'STRFs' fit in this work actually described the spectral and temporal modulation preferences of voxels in higher auditory areas and were fit by a procedure which finds the average modulation spectrum weighted by the fMRI response magnitude. It remains unclear whether STRFs describing the preferences of single units in frequency and time can be fit to slow-varying signals. Although progressive improvements in the GCaMP family of calcium indicators have yielded indicators with greater linearity, signal-to-noise ratio, and kinetic properties, inferring temporally precise activity from fluorescence traces remains a challenge (Broussard et al., 2025; Pachitariu et al., 2018a). Here, I show that STRFs can be estimated from simulated fluorescence traces under optimal conditions.

$\mathrm{STRF_f}$ (fit to simulated calcium fluorescence) resembled a temporally-extended version of $\mathrm{STRF_s}$ (fit to simulated spike trains), though was not equivalent to a temporal convolution of $\mathrm{STRF_s}$ with the spike-to-fluorescence kernel, likely due to sources of variability in the model cascade following spike generation (see *Section 2.3.3*). Further work could compare the form of $\mathrm{STRF_f}$, to an 'analytic' $\mathrm{STRF_f}$, obtained by the aforementioned convolution. Elimination of all noise sources between spiking and

fluorescence (i.e. by using a finite impulse response filter with a history < 1000 ms and taking the mean STRF from multiple weight initialisations), can be used confirm that the $\mathrm{STRF_f}$ fit by regression is equivalent to the analytic $\mathrm{STRF_f}$.

Nothwithstanding, results were broadly in line with my prediction that the temporal effect of the spike-to-fluorescence transformation was to broaden features along the history axis of STRFs. This change in structure had significant implications for the interpretation of neural tuning properties, with the history axis predominantly describing indicator properties rather than neuronal tuning. Spectral properties, especially best frequency, were largely preserved in $\mathrm{STRF_f}$, thus kernels fit to calcium fluorescence data may still represent a useful model of spectral tuning to complex stimuli, extending pure tone tuning curves and frequency response areas traditionally used in imaging studies. While receptive field components placed along the spectral axis were predominantly just elongated in history, temporally apposed components were more severely distorted, with $\mathrm{STRF_f}$ exhibiting qualitatively different structure that worsened with slower indicator kinetics. This may pose a significant challenge when imaging neurons *in vivo*, since auditory cortical STRFs often contain temporal structure like the lagging inhibition examined here.

$\mathrm{STRF_f}$ failed to capture the timescale of responses because they fit to fluorescence transients that do not reflect neuronal tuning. Both latency and temporal bandwidth values obtained with fluorescence appear to be related to those obtained with spiking, by some constant (i.e. the y-intercepts in Fig 2.6D, E) whose exact determinants are unclear – though indicator kinetics are likely to be a key contributor. Despite this ambiguity, $\mathrm{STRF_f}$ can support comparisons of temporal (and spectral) tuning between units in imaged fields, with the caveat that the mapping between spiking and fluorescence of individual units is likely to differ (Éltes et al., 2019; Helmchen et al., 1996; Maravall et al., 2000). Finally, changes in $\Delta F/F$ occur in the absence of spiking

and can be linked to various intracellular cascades due to the role of calcium as a ubiquitous second messenger (Grienberger and Konnerth, 2012). Indeed, kernels are unable to predict slow variations in fluorescence occurring over timescales longer than their history axes.

The selection of a suitable spiking model is also necessary to produce a meaningful model of neuronal responses. Poisson spiking is commonly assumed due to ease of implementation as part of a linear-nonlinear cascade, however it does not account for refractoriness or other spiking nonlinearities including the types of adaptation or bursting observed in auditory cortex (Wang, 2007). Natural stimuli are known to drive robust yet highly non-linear responses, not accounted for by such a model of spiking (Theunissen et al., 2000). My model also did not take account of nonlinearities of spectrotemporal integration, such as multiplicative or divisive interactions of receptive fields which are known to interact with stimulus statistics and bias STRF estimates (Christianson et al., 2008). Such complexities have the potential to interact with the S2F transformation. One simplistic view of auditory cortex posits that it functions as a filter bank specialised for the detection of complex spectrotemporal features (Chi et al., 2005). Here, five idealistic linear receptive fields were simulated to provide a proof-of-concept demonstration. In reality, auditory cortex employs a diversity of non-linear filters with complex structure that is often covariant in frequency and time. Only a proportion of measured STRFs in auditory cortex have structure as simple as the receptive fields used here, however these were used as their tuning properties and morphology are commonly observed in real data and they are easily interpretable.

Stimulus choice also has broad implications when characterising neuronal tuning, and it is not possible for a single stimulus ensemble to effectively probe the full range of computations performed at the neuronal level. DRC stimuli are used commonly in STRF estimation because they can be varied parametrically and are a logical extension

of the pure tone paradigm. Here, they provided the best STRF estimates across conditions in terms of recapitulating simulated receptive field structure. Ripple stimuli (TORCs) have also been used widely in cortex, thalamus and midbrain studies, so these were used for comparison here. TORCs used in previous studies contained ripples with temporal modulations up to 50 Hz driving responses (Depireux et al., 2001; Kowalski et al., 1996a). Relatively linear auditory cortical units exhibit phase-locked spiking to the ripple envelope at this frequency, however it is unclear whether fluorescence (or indeed the intracellular calcium concentration it is measuring) can be entrained to these modulations (Helmchen et al., 1996). $\mathrm{STRF_f}$ fit to fast TORC responses did not contain recognisable structure, while slow TORCs were somewhat more successful. This suggests that slower-varying input signals may provide a more useful substrate for probing neuronal tuning in imaging experiments. Indeed, even in electrophysiology it is found that auditory cortical units phase-lock best to ripples below 10 Hz (Eggermont, 1994; Schreiner and Urbas, 1988).

Further work can extend this simple model by investigating different types of regularisation, which will likely have significant implications for the kernels fit. For instance, priors of temporal, but not spectral, sparsity may reduce the extent of smoothing and yield better STRF estimates compared to deconvolved spikes, which attempt to account for neural response smoothing before the model is fit. Simultaneous *in vivo* electrophysiology with two-photon imaging will also allow numerous avenues to be explored, especially because accurate spike inference from fluorescence data requires ground-truth recordings (Broussard et al., 2025). Newer fitting procedures may also be useful, such as neural network models utilising hidden layers to more effectively account for the S2F transformation (Harper et al., 2016; Keshishian et al., 2020). Ultimately, the feasibility of STRFs in two-photon imaging experiments must be assessed *in vivo*.

Chapter 3

In vivo two-photon calcium imaging of spectrotemporal receptive fields

3.1 Introduction

Neurons in auditory cortex process input to extract acoustic features of interest underlying conscious perception and behaviour, and a major goal of auditory neuroscience is to quantify these input-output functions as receptive fields. The spectrotemporal receptive field (STRF) model represents a valuable first-order approximation of the stimulus-response transformation and has uncovered a richness of processing occurring at the cellular level. While modern electrophysiology allows access to the simultaneous activity of neurons at sub-millisecond resolution (Jun et al., 2017), two-photon calcium imaging can measure the responses of hundreds to thousands of genetically and anatomically defined individual cells. Thus, two-photon calcium imaging has rapidly grown to rival electrophysiology as a gold-standard recording modality in sys-

tems neuroscience, despite relatively slow kinetics (typically hundreds of milliseconds) that struggle to resolve individual action potentials in time. Building on previous generations, the new GCaMP8 family of calcium indicators produces fast and robust responses to single action potentials offering promising avenues for imaging at more biological timescales (Zhang et al., 2023). Additionally, optical techniques are an attractive option in behavioural research where neurons can be imaged over long durations and chronically (Lee et al., 2023). By contrast, maintaining stable recordings over longer durations is particularly challenging in electrophysiological preparations (David, 2018).

Using an end-to-end parametrised model in *Chapter 2*, I simulated the effect of the spike-to-fluorescence transformation on STRFs, showing that they can be fit to fluorescence data under different spiking rates, indicator kinetics and noise conditions. While encouraging, these results require *in vivo* validation to support two-photon imaging of spectrotemporal receptive fields as an emerging methodology. Importantly, it is unknown whether the complex auditory stimuli used in previous electrophysiological studies (Cooke et al., 2018; Elhilali et al., 2004) can drive robust fluorescence changes in a two-photon imaging setting. Furthermore, the kinetics and sensitivity reported for GCaMP indicators are largely based on 1-2 action potential events under optimised imaging conditions. It is unclear whether these properties are maintained under typical imaging conditions with neurons stimulated using continuous, dynamic stimuli. This chapter describes the estimation of STRFs *in vivo* in the auditory cortex of awake mice. GCaMP8s-expressing mice were presented with dynamic random chord (DRC) and temporally orthogonal ripple combination (TORC) stimuli, two classes applied successfully in previous neurophysiological experiments, to investigate the feasibility of estimating STRFs from two-photon calcium imaging data.

3.2 Methods

3.2.1 Animals

Three (2 male, 1 female) *Igs7*$^{tm3(tetO\text{-}GCaMP8s,CAG\text{-}tTA2)Genie}$/J (JAX 037952, Jackson Laboratories, USA) × B6.Cg-Tg(*Camk2a*-cre)T29-1Stl/J (JAX 005359, Jackson Laboratories, USA) mice were used. Two of the mice were 6-12 weeks old during imaging. The other mouse was 20-26 weeks old so may have exhibited some degree of age-related high-frequency hearing loss (Bowen et al., 2020). Animals were housed in groups and maintained on a 12 h light/dark cycle at 20-24 °C with ad libitum access to food and water. All procedures were approved by the local ethical review committee of the University of Oxford and performed under license from the UK Home Office in accordance with the Animal (Scientific Procedures) Act 1986.

3.2.2 Surgical procedures

Mice were premedicated with intraperitoneal dexamethasone (Dexadreson, 4 mg), atropine (Atrocare, 1 mg) and carprofen (Rimadyl, 0.15 mg), and anaesthetised using isofluorane (1.5-2%). A heating mat with a DC temperature controller was used to maintain a body temperature of 37-38 °C. Following placement in a stereotaxic frame (Model 900LS, David Kopf Instruments, USA), a right-sided dorsolateral scalp incision was made. The temporal muscles were retracted and a ~ 3 mm diameter craniotomy was performed overlying auditory cortex. A glass coverslip was used to cover the exposed cortex and fixed to the skull with cyanoacrylate adhesive (Pattex UltraGel; Henkel, DE). A custom steel headpost was attached near the midline of the skull using dental cement (C&B Superbond, Sun Medical, Japan).

3.2.3 Widefield imaging

Widefield imaging used a 470 nm LED (M470L3, Thorlabs, USA), a digital camera (pco.edge 4.2 bi; Thorlabs, USA) and a 2× objective lens (TL2X-SAP, Thorlabs, USA) integrated on a Bergamo II microscope (Thorlabs, USA). Image acquisition (10 Hz, 96 × 128 pixels) was carried out using ThorCam software (ThorLabs, USA). Auditory stimuli were created using LabView (National Instruments), digital-to-analogue converted through a soundcard (ASUS Xonar-U7), amplified (Portable Ultrasonic Power Amplifier; Avisoft Bioacoustics, Germany) and delivered through a free-field ultrasonic transducer (Vifa; Avisoft Bioacoustics, Germany) positioned ∼ 15 cm from the head. For functional mapping, sinusoidally amplitude-modulated tones were presented (500 ms duration, 10 Hz modulation frequency, 100% modulation depth). Each animal was exposed to 15 presentations of a low-frequency tone (either 4 or 5.04 kHz) and a high-frequency tone (either 25.4 or 32 kHz), at 55 or 65 dB SPL intensity and a rate of 0.33 Hz. To generate frequency maps (Fig. 3.1A), the average response to each stimulus was calculated by subtracting the baseline activity (mean signal intensity in a 1 s window before onset) from the response following stimulation (mean signal intensity in a 1 s window after onset). The responses to low- and high-frequency tones were then subtracted from each other, colour mapped and superimposed onto images of the cortical surface for reference. Functional widefield maps were used to categorise imaging fields as primary auditory cortex (A1), anterior auditory field (AAF), secondary auditory cortex or other. Data from primary fields (A1 and AAF; 28/39 fields) were used for further analysis.

3.2.4 *In vivo* two-photon imaging

Mice were allowed to recover for at least ten days following surgical procedures. Animals were handled daily prior to imaging and habituated with shorter durations of

head fixation. Imaging was carried out using a two-photon laser-scanning microscope (B-Scope, ThorLabs, USA) with excitation light emitted by a Mai-Tai eHP laser (SpectraPhysics, USA) at a 70 fs pulse width and 80 MHz repetition rate. The beam was passed into a modulator (Conoptics, USA) followed by an 8 kHz resonant scanner (x-plane) and a galvanometric scan mirror (y-plane) to achieve a laser power of 30-40 mW at the objective. Frames were acquired at a 512 × 512 pixel resolution and 28.41 Hz frame rate with the resonant scanner running in bidirectional mode. A 16×/0.80 NA objective (Nikon, Japan) was used with a 525/50 filter followed by GaAsP photomultiplier tubes (Hamamatsu, Japan). All imaging was carried out at 150-250 µm below the cortical surface over a 500 × 500 µm field of view. Imaging hardware was controlled using ScanImage 4.1 software[1] running on a personal computer.

3.2.5 Stimuli

DRC stimuli were constructed as previously described (see *Chapter 2*). Pure tone pips of a fixed amplitude (25 ms duration, 5 ms half-cosine on/off ramps, 1/6-octave spacing 2-64 kHz) were randomly interleaved with pip times determined by a Poisson process. Three sets of DRC stimuli with low (λ = 0.1 pips band^{-1}sec^{-1}), medium (λ = 0.3 pips band^{-1}sec^{-1}) and high (λ = 3.5 pips band^{-1}sec^{-1}) pip densities were created. Each set consisted of 4 stimuli of 28 s, giving a total duration of 112 s. Slow TORC stimuli were created as previously described (see *Chapter 2*). Moving ripples had spectral modulations up to 1.4 cycles/oct (0.2 cycles/oct spacing) and temporal modulations between 2 to 7 Hz (1 Hz spacing). Moving ripple envelopes were applied to randomised phase noise carriers (100 carrier frequencies per octave, 2-64 kHz) and summed to yield 26 TORCs (8 s each, total duration of 208 s). Individual DRCs and TORCs were repeated 10 and 5 times, respectively, to give stimulus sets lasting ∼20

[1] https://scanimage.org

minutes.

Pure tones at 18 log-spaced frequencies (1-64 kHz) and 3 intensity levels (40, 55 and 70 dB SPL) were presented (200 ms duration, 10 ms half-cosine on/off ramps, 1.3 s inter-trial interval) to measure FRAs. All stimuli were presented in a pseudorandom order and repeated 10 times at each frequency-intensity combination.

3.2.6 Presentation of stimuli

Stimuli were presented using a TDT System 3 RZ6 real-time processor (Tucker-Davis Technologies, USA) controlled with a personal computer. Custom-written MATLAB scripts used RPvdsEx circuits to interface with the processor and co-ordinate stimulus presentation with frame acquisition by the two-photon microscope. The processor was used to drive an ultrasonic free-field transducer (Vifa; Avisoft Bioacoustics, Germany) positioned $\sim$ 10 cm from the left ear canal. Calibration was achieved using a 1/4-inch free field microphone (G.R.A.S., Denmark) placed at the position of the animal's head and a custom-written calibration routine run in MATLAB to create an inverse filter for the speaker. This achieved a flat ($\pm$5 dB SPL) power spectrum within a 1-64 kHz range. Mice were imaged in a sound- and light-attenuating enclosure with an ambient noise floor of $\sim$ 40 dB with most power below 1 kHz.

3.2.7 Isolating neural responses

Imaging files from a total of 39 fields were loaded into the Suite2p pipeline (Pachitariu et al., 2016[2]) which carried out automatic registration, mechanical drift correction, cell detection and neuronal/neuropil trace extraction. Suite2p also carried out spike deconvolution on fluorescence traces, based on a simple non-negative deconvolution algorithm without regularization (Friedrich et al., 2017; Pachitariu et al., 2018b).

[2]https://github.com/MouseLand/suite2p

STRFs were fit to both deconvolved spikes and fluorescence traces (see *Section 3.3.4*). All regions of interest identified by Suite2p were manually inspected to check alignment with the cell membrane and that cells were healthy – this included looking for a clear rim of fluorescence around a dark centre and inspecting the fluorescence trace for evidence of cell death (seen as 'bursting'). Fluorescence traces were then exported to MATLAB and filtered based on the output of the default Suite2p classifier which identifies whether a region of interest is likely to be a cell. Raw fluorescence traces were neuropil corrected using the formula:

$$F_{\text{cell}} = F - 0.7 \cdot F_{\text{neu}} \tag{3.1}$$

F, F_{cell} and F_{neu} denote raw, neuropil-corrected, and neuropil fluorescence traces, respectively. $\Delta F/F$ was calculated by finding the median between the 10th and 70th percentile of F_{cell} and normalising by this value.

3.2.8 Analysis of pure tone responses

The tone response was calculated as the time-averaged $\Delta F/F$ within a 6 frame ($\sim$ 200 ms) window from the tone onset. This was trial-averaged at each frequency-intensity combination to calculate a mean response. A two-way analysis of variance (ANOVA) was used to identify neurons with a significant ($\alpha = 0.05$) effect of tone frequency, intensity, or a frequency/intensity interaction on their $\Delta F/F$ response. Neurons showing a significant effect of frequency or interaction with intensity were deemed to be frequency tuned.

3.2.9 Neuron inclusion

Auditory cortical neurons respond to specific spectrotemporal features in the stimulus. Encoding models exploit this to recover the stimulus dimensions that drive responses.

To identify neurons whose fluorescence traces were modulated in a reliable and repeatable manner across stimulus presentations, I calculated the signal power for each neuron (Sahani and Linden, 2003). First, a response vector to N stimulus repetitions, $\mathbf{r} = (r_t)_{t=1...T}$ is constructed, where r_t is the value of $\Delta F/F$ at frame t. Assuming the presence of a deterministic stimulus-driven component (the signal), $\boldsymbol{\mu}$, common to each response and random zero-mean noise, $\boldsymbol{\eta}^{(n)}$, in the nth trial, the measured response is given by

$$\mathbf{r}^{(n)} = \boldsymbol{\mu} + \boldsymbol{\eta}^{(n)} \tag{3.2}$$

To estimate the power of $\boldsymbol{\mu}$, the response power is first defined as the average squared deviation from the mean:

$$P(\mathbf{r}) = \langle (r_t - \langle r_t \rangle) \rangle^2 \tag{3.3}$$

where $\langle \cdot \rangle$denotes a time average. The expected response power in the nth trial is:

$$P(\mathbf{r}^{(n)}) = P(\boldsymbol{\mu}) + \langle (\eta_t^{(n)})^2 \rangle \tag{3.4}$$

Furthermore, the power of the average response and the average power *per* response can be calculated using the following two equations, respectively, where bars indicate trial averages:

$$P(\overline{\mathbf{r}^{(n)}}) = P(\boldsymbol{\mu}) + P(\overline{\boldsymbol{\eta}^{(n)}}) \tag{3.5}$$

$$\overline{P(\mathbf{r}^{(n)})} = P(\boldsymbol{\mu}) + \overline{P(\boldsymbol{\eta}^{(n)})} \tag{3.6}$$

If $\boldsymbol{\eta}$ is assumed to be independent across trials:

$$P(\overline{\boldsymbol{\eta}^{(n)}}) = \overline{P(\boldsymbol{\eta}^{(n)})}/N \tag{3.7}$$

and solving for $P(\boldsymbol{\mu})$ yields an estimate for signal power, $\hat{P}(\boldsymbol{\mu})$

$$\hat{P}(\boldsymbol{\mu}) = (N \cdot P(\overline{\mathbf{r}^{(n)}}) - \overline{P(\mathbf{r}^{(n)})})/(N-1). \tag{3.8}$$

Subtracting this from $\overline{P(\mathbf{r}^{(n)})}$ yields an estimate for noise power, $\hat{P}(\boldsymbol{\eta})$:

$$\hat{P}(\boldsymbol{\eta}) = \overline{P(\mathbf{r}^{(n)})} - \hat{P}(\boldsymbol{\mu}). \tag{3.9}$$

Furthermore, the variance of $\hat{P}(\boldsymbol{\mu})$ can be calculated as follows:

$$\mathrm{Var}\left[\hat{P}(\boldsymbol{\mu})\right] = \frac{4}{N}\left(\frac{1}{T^2}\boldsymbol{\mu}'\Sigma\boldsymbol{\mu} - \frac{2}{T}\mu\boldsymbol{\sigma}'\boldsymbol{\mu} + \mu\sigma\mu\right) + \frac{2}{N\,(N-1)}\left(\frac{1}{T^2}\mathrm{Tr}\left[\Sigma\Sigma\right] - \frac{2}{T}\boldsymbol{\sigma}'\boldsymbol{\sigma} + \sigma^2\right) \tag{3.10}$$

where Σ is the noise covariance matrix, $\boldsymbol{\sigma}$ is formed by averaging the columns of Σ, σ is the average of all elements in Σ and μ is the time average of $\boldsymbol{\mu}$. See Sahani and Linden, 2003 for a full discussion.

Crucially, neurons were included for further analysis provided that their responses to a stimulus set contained significant signal power defined as:

$$\hat{P}(\boldsymbol{\mu}) - \sqrt{\mathrm{Var}[\hat{P}\,(\boldsymbol{\mu})]} > 0. \tag{3.11}$$

A further metric used to quantify the degree of intertrial variability in neural responses is the noise ratio:

$$\text{Noise ratio} = \frac{\hat{P}\,(\boldsymbol{\eta})}{\hat{P}\,(\boldsymbol{\mu})}. \tag{3.12}$$

The noise ratio has been used as an inclusion threshold in previous experiments (Rabinowitz et al., 2012; Rahman et al., 2019), however was not used here due to the uncertainty inherent to signal power estimates. The calculation of the noise ratio involves normalisation by the signal power (or a related quantity in the case of CC_{norm};

see *Section 3.2.13*) which can lead to spurious results in neurons where the standard error interval includes zero. While the requirement of significantly non-zero signal power proved stringent for my dataset, it increased confidence in the STRFs of included neurons.

3.2.10 Cochleagrams

Stimulus cochleagrams were generated as previously described (see *Chapter 2*) but were binned to match the imaging frame rate of 28.41 Hz.

3.2.11 STRF estimation

Linear STRFs were fit as previously described (see *Chapter 2*) within a generalised linear model assuming Gaussian noise. Onset responses were removed by discarding the first second of $\Delta F/F$ to each stimulus before trial averaging. STRF weights were estimated by linear regression, using either L1- or L2-regularisation. Regularisers were incorporated as penalty terms in the objective function:

$$E_{L1} = \sum_t (y_t - \hat{y}_t)^2 + \lambda \sum_{fh} |W_{fh}| \tag{3.13}$$

$$E_{L2} = \sum_t (y_t - \hat{y}_t)^2 + \lambda \sum_{fh} {W_{fh}}^2. \tag{3.14}$$

In the above equations, y_t and $\hat{y}_t$ represent the measured and predicted responses (respectively) to the training set in frame t, λ is the strength of the regulariser, and W_{fh} represents the weight of the kernel at frequency f and history h. The strength of regularisation was chosen based on the performance of kernels estimated at a range of λ values on held-out data. The training and test set for this process were created using the MATLAB glmnet.m function (Qian et al., 2013) from the 90% training set provided. Thus, the held-out 10% was not used in any capacity except for STRF

evaluation. Additionally, separable kernels were fit to all neurons. These drastically reduce the number of parameters to be fit by assuming separability in frequency and history and have been applied successfully in rodents (Sahani and Linden, 2003), ferrets (Simon et al., 2007), and avian forebrain (Woolley et al., 2009). Frequency and history marginals were fit by least-squares regression until convergence of the following:

$$E_S = \sum_t (y_t - \hat{y}_t)^2. \tag{3.15}$$

3.2.12 Tuning properties of STRFs

STRFs were smoothed using a 2D-Gaussian kernel (3 frequency bins-by-4 history bins) and positive half-wave rectified to facilitate comparison with FRAs. Briefly, spectral and temporal tuning profiles were obtained by averaging the smoothed STRFs along the temporal and spectral axes, respectively. The maxima of these profiles defined best frequency and peak latency, while the range of the profile greater than half the maximum gave spectral and temporal bandwidths. The excitatory-inhibitory index (EI) was calculated as described in *Section 2.2.7*.

3.2.13 Model evaluation

Predicted responses, $\hat{y}_t$, were generated by convolving the STRF with the held-out stimulus cochleagram. A normalised form of the Pearson's product-moment correlation coefficient known as CC_{norm} was used to measure the similarity between $\hat{y}_t$ and the recorded fluorescence y_t (Hsu et al., 2004; Schoppe et al., 2016). CC_{norm} accounts for the predictable portion of variance in a noisy signal by the following:

$$CC_{\text{norm}} = \frac{CC_{\text{abs}}}{CC_{\text{max}}}, \tag{3.16}$$

where

$$CC_{\max} = \sqrt{\frac{2}{1+\sqrt{\frac{1}{CC_{\text{half}}^2}}}} = \sqrt{\frac{2}{1+\frac{1}{CC_{\text{half}}}}}, \text{ for } CC_{\text{half}} > 0. \tag{3.17}$$

CC_{abs} is the value of the correlation coefficient and $CC_{\max}$ represents the theoretical maximum correlation coefficient that a perfect model could achieve. $CC_{\max}$ is the correlation between the recorded neural response and the underlying driving function and can be estimated by splitting data into halves, calculating the correlation coefficient between them (CC_{half}) and repeating over many combinations. Since calculations over the large number of combinations is computationally expensive, $CC_{\max}$ is calculated instead by exploiting its mathematical relationship with signal power, which can be estimated more readily (Schoppe et al., 2016 for details). CC_{norm} is effectively bounded between -1 and 1, but can occasionally take on values outside of this range due to severe overfitting.

3.3 Results

To investigate whether STRFs can be fit to two-photon imaging data, I performed awake experiments in 3 mice expressing GCaMP8s. Of a total of 39 fields imaged, 28 were confirmed to be primary auditory areas (12 A1, 9 AAF and 7 at the A1/AAF high-frequency border), 3 were in secondary auditory cortex, and 8 were not clearly localised with widefield imaging (Fig. 3.1A). Results presented hereon pertain to the primary auditory areas A1 or AAF. Data from outside primary or secondary auditory cortices did not typically provide reliable FRAs or STRFs, and were therefore excluded. Following pre-processing and exclusion, a total of 12,289 neurons were included in the dataset.

DRC (Fig. 3.1B) and TORC stimuli were used to characterise STRFs, as previously described (deCharms et al., 1998; Klein et al., 2000). DRCs are composed of randomly

interleaved tone pips while TORCs are created by summing moving ripples, the auditory corollary of drifting gratings (see *Section 3.2.5*). Three sets of DRC stimuli were created with varying tone pip densities, to investigate the effect of stimulus density (see *Section 3.3.6*). Here, I first describe results for DRC stimuli and STRFs obtained using TORCs are addressed separately (see *Section 3.3.8*).

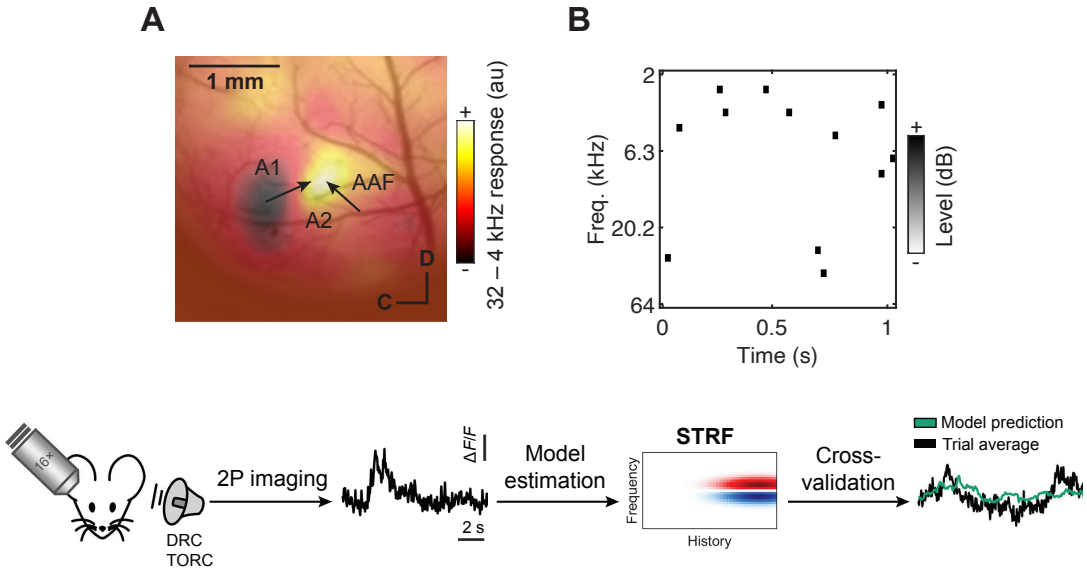


Figure 3.1: Experimental design for estimation of STRFs by *in vivo* two-photon calcium imaging. A. Widefield microscopy reveals dorsoventral tonotopic gradients of A1 and AAF. Overlaid colours indicate functional widefield map obtained by measuring responses to sinusoidally-amplitude modulated tones at 4 and 32 kHz. Mean response to 4 kHz was subtracted from the mean response to 32 kHz to obtain functional map. C = caudal, D = dorsal. **B.** DRC spectrotemporal envelope. Random tone pips (black) occurred at times generated by a Poisson process. Example shown is a medium density DRC ($\lambda = 0.3$ pips band^{-1} sec^{-1}). **C.** Spectrotemporal envelope of a slow TORC stimulus. A total of 26 TORCs were used in imaging experiments, each lasting 8 s. **D.** Experimental schematic. $n = 3$ transgenic GCaMP8s mice were imaged during presentation of DRC, TORC and pure tone stimuli. Recorded $\Delta F/F$ was used to fit linear STRFs and FRAs (not shown in figure). Prediction accuracy of STRFs was measured by comparing predicted responses (green) with held-out data (black).

During a typical imaging session, mice were presented with pure tones followed by complex stimuli to enable tuning comparisons. FRAs plotted using tone responses

also provided a form of ground truth to which STRFs could be compared. Data from each neuron was randomly partitioned into training (90%) and test (10%) sets that were used to fit and evaluate STRFs, respectively (Fig. 3.1C). Note that test data was a contiguous portion of the trial-averaged response rather than discontinuous query points or a stimulus repeat. This allowed me to test the ability of the model to generalise to new portions of stimulus. Linear STRFs were fit using L1- and L2-regularised linear regression which sought to minimise the mean squared error on the training set. Performance measures reported below are based on test set performance.

3.3.1 STRFs can be fit to fluorescence traces

Two-photon microscopy samples the activity of many neurons which vary in their level and consistency of stimulus-driven activity. Therefore, it is necessary to exercise some form of inclusion criterion to isolate neurons that respond to the sounds presented in the experiment. Assuming that auditory cortical neurons are feature selectors driven at specific points throughout the stimulus (Chi et al., 2005), responses should be relatively consistent across trials in which these stimuli are repeated. Since traces are noisy and not identical across trials, it can be assumed that the imaged response is the sum of a constant stimulus-driven component and stochastic noise. The fitting procedure aims to identify a mapping from stimulus to response, so florescence traces dominated by a stimulus-driven component provide better STRFs.

One measure of the stimulus-driven component is the signal power, which can be estimated by considering the power, or variance, within and across stimulus repeats (Sahani and Linden, 2003). The remainder of the total variance is noise, which can often be orders of magnitude larger than signal power. An example imaging field in A1 is shown (Fig. 3.2A) along with a plot of signal power against noise power for each neuron in that field (Fig. 3.2B). Across all cells in this field, mean signal power was 0.0026 $\pm$ 0.0065 $\Delta F/F^2$ bin^{-1} and noise power was 0.11 $\pm$ 0.08 $\Delta F/F^2$ bin^{-1}.

Error bars represent the standard error and were used to determine whether each cell had a significantly non-zero signal power (blue points, see *Section 3.2.9*). These neurons were selected for further analysis. In this example field, 87 of 379 (23.0%) neurons met this criterion, and they had a mean signal power of 0.024 ± 0.022 $\Delta F/F^2$ bin^{-1} and noise power of 0.24 ± 0.16 $\Delta F/F^2$ bin^{-1}. STRFs were fit using responses to individual stimulus sets and animals were often presented with more than one stimulus set per session, so while 8,970 neurons were imaged, there were 13,214 neuron-DRC set combinations of which 456 (3.5%) had significant signal power. Although this was a stringent criterion, it increased confidence in the STRFs analysed throughout this chapter as their weights were known to be determined by a stimulus-driven response.

This example imaging field in Fig. 3.2 contained 27 pure tone-tuned neurons, determined by FRAs with a statistically significant effect of frequency or a frequency/intensity interaction (defined as $p < 0.05$, two-way ANOVA). In contrast to this example, most imaging fields contained more pure-tone-tuned neurons than those included for STRF fitting. Neuronal best frequency was calculated from the pure tone responses as the frequency that elicited the greatest response in the level-averaged tuning curve. Most neurons within this field were tuned to 9 or 14.7 kHz (Fig. 3.2C). The FRA for an exemplar cell with a best frequency of 9 kHz is shown (Fig 3.2D) along with the corresponding STRF demonstrating a clear excitatory field centred approximately at this frequency (Fig. 3.2E). STRFs were fit over 42 history bins whose widths matched the acquisition rate of fluorescence data ($\sim$ 28 Hz). This STRF provided a similar account of spectral tuning to the FRA with excitation at $\sim$ 9 kHz along with various other inhibitory and excitatory features extending back into stimulus history. The temporal tuning of this kernel was broader than observed in comparable simulations in *Chapter 2*, with long bandwidth components seen throughout the STRF. This may have been due to the greater complexity of biological receptive fields compared to the simple ones used in modelling, or slower indicator kinetics *in vivo*. To aid visualisation and the

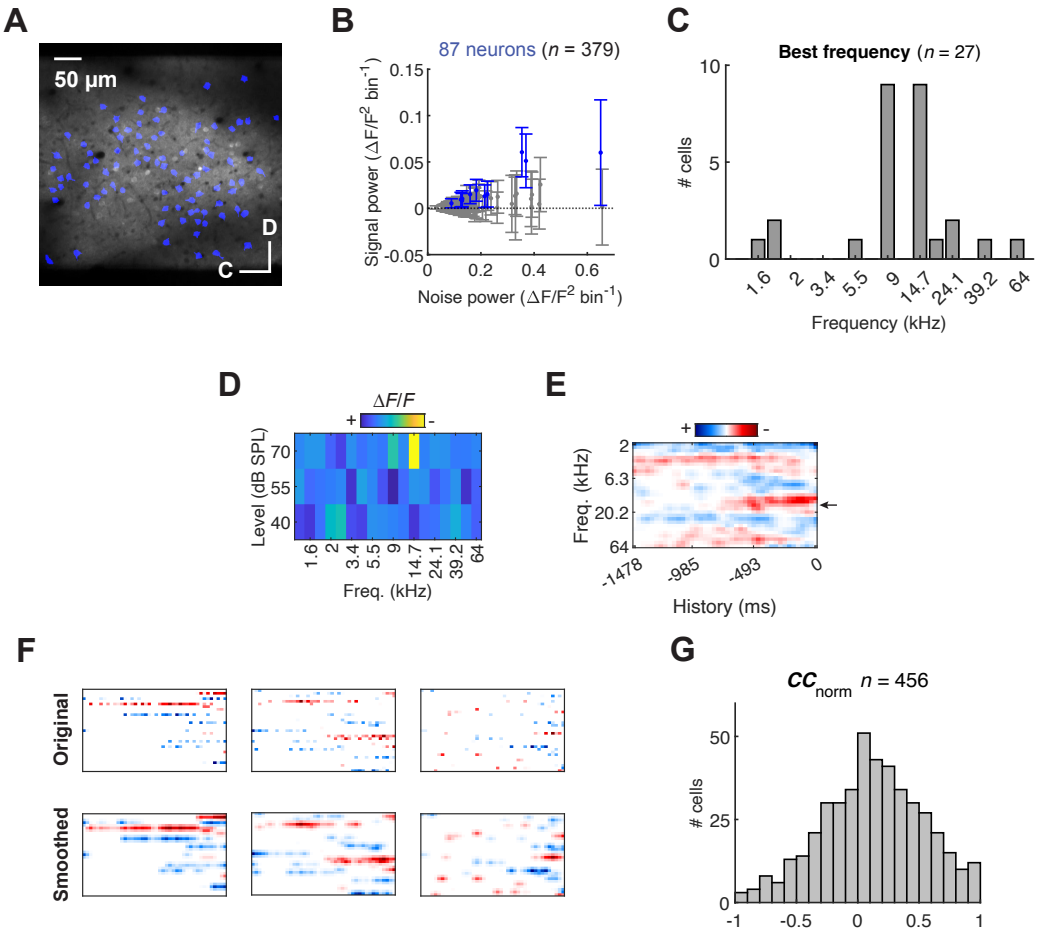


Figure 3.2: STRFs can be fit with two-photon imaging. A. Two-photon micrograph of an imaged field. Neurons with significant signal power were included for analysis (highlighted in blue). **B.** Plot of signal power against noise power for all neurons ($n = 379$) in the field of view with included neurons ($n = 87$) shown in blue. Error bars represent ± 1 standard error on the signal power. **C.** Distribution of FRA best frequencies of frequency-tuned cells (see *Section 3.2.8*). **D.** FRA and **E.** STRF of a neuron in this field. Arrow indicates the FRA best frequency for comparison with STRF tuning. FRA colour bar indicates trial-averaged $\Delta F/F$ at each frequency-level combination. STRF colour bar indicates magnitude and sign of each linear weight. Stimulus energy at positions with excitatory (red) weights is associated with an increase in $\Delta F/F$ at the current time point (0 ms history). Energy falling in inhibitory (blue) regions inhibits neurons. The linear STRF represents the optimal stimulus feature driving a neuron. Respective colour bars apply to all subsequent plots. **F.** Original STRFs (top) were smoothed (bottom) with a 3×4-point Gaussian kernel to facilitate the measurement of tuning properties. **G.** Distribution of CC_{norm} values in all cells included for analysis. CC_{norm} was calculated using original (non-smoothed) kernels. Note cells with extreme values (outside the range $-1 \leq CC_{\text{norm}} \leq 1$; see *Section 3.2.13*) are not shown ($n = 14$).

measurement of tuning properties, STRFs were convolved with a 3 frequency bin × 4 history bin 2D-Gaussian kernel (Fig. 3.2F). Though smoothing risks providing an impression of fidelity, it is practised in the STRF literature to aid visualisation and the measurement of tuning properties. Smoothed kernels are displayed throughout the remainder of this chapter.

Successful encoding models should be able to predict responses to novel stimuli. To evaluate the predictive power of STRFs, I generated predicted responses to a test set and compared them to true neural responses. For each pair of predicted and imaged responses, I calculated CC_{norm}, a form of the correlation coefficient which is normalised to a maximum achievable value for each neuron given its response variability (see *Section 3.2.13*). The neuron shown here had a modest CC_{norm} of 0.159 and was unable to predict most of the activity throughout the test set (not shown). Mean CC_{norm} across all neurons was 0.17 ± 0.44 (Fig. 3.2G). Many neurons had negative CC_{norm} values, indicating that STRF predictions were poorly or potentially anticorrelated with test data. Possible explanations, including the quality of driven responses and ability of the linear STRF model to fit to fluorescence data are discussed later (see *Section 3.3.7* and *Section 3.4.2*).

3.3.2 Effect of regularisation

Short imaging times and restricted stimulus sets necessitated regularisation of STRFs which inevitably overfit to noisy data. Here, I compared the effects of L1- and L2-regularisation incorporated as a penalty term into the error term. L1-regularisation imposes a prior of sparsity by shrinking unlikely weights to zero while L2-regularisation seeks to distribute power across coefficients to achieve smoothness. I also fit separable STRFs which can be defined as the outer product of frequency and history marginals. This method significantly reduces the number of parameters to be fit by assuming independence of the frequency and history dimensions, and has been successful in

mouse (Cooke et al., 2018; Linden et al., 2003) and ferret auditory cortex (Rabinowitz et al., 2011).

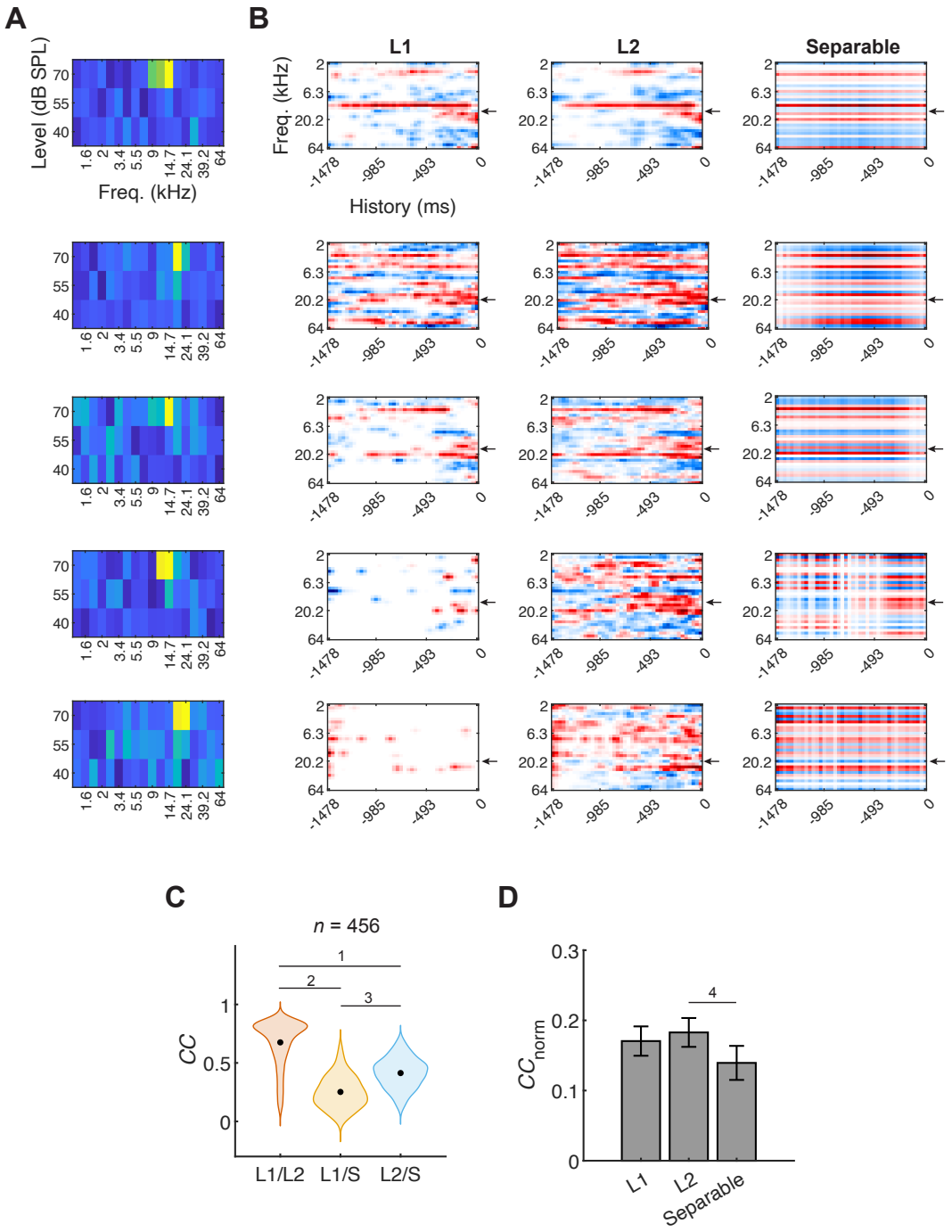


Figure 3.3: Choice of regulariser has a significant effect on STRF structure. **A.** FRAs for 5 frequency-tuned neurons. **B.** STRFs for example neurons fit with different regularisation procedures. First and second columns show inseparable kernels fit by linear regression with L1 and L2 regularisation, respectively. Third column shows separable STRFs. **C.** Structural similarity of differently-regularised STRFs measured by the Pearson's product-moment correlation coefficient of pairwise comparisons within cells (one-way ANOVA with Tukey-Kramer test, $^{1\text{-}3}p < 9.6 \times 10^{-10}$; note true p-values were below the lower limit imposed by the MATLAB function multcompare.m). **D.** Effect of regularisation on performance in held-out data (one-way ANOVA with Tukey-Kramer test, $^{4}p = 0.0028$).

The results of these analyses on 5 exemplar neurons are displayed (FRAs in Fig. 3.3A, and corresponding STRFs in Fig. 3.3B). The first neuron had an FRA best frequency of $\sim$ 14.7 kHz (Fig. 3.3A, first row) and the STRF accordingly exhibited a strong excitatory component centred at $\sim$ 15 kHz (Fig. 3.3B, first row). Both L1- and L2-regularised kernels maintained this clear frequency preference along with broad similarity across their other components. A similar result was seen for the second neuron, with largely similar STRFs under both regularisers (Fig. 3.3A, B, second row), and a strong excitatory component at the best frequency of the FRA. In contrast, L1-regularised kernels for the remainder of the cells plotted here were considerably sparser than L2-regularisation (Fig. 3.3B, third-fifth row). Excitatory bands were better defined with L1-regularisation, and less energy was contained further back in stimulus history. Overall, L1-regularisation sparsened STRFs, leading to more 'speckly' structure lacking the smooth spectrotemporal modulations observed in L2-regularisation. However, L2-regularised kernels were more likely to contain disorganised and unexpected features, such as frequency responses that were not observed in the pure tone FRA. In the case of poorly-driven neurons, L2-regularisation may have captured correlated stimulus features. Both regularisers yielded noisy structure, with few kernels containing clearly demarcated excitatory or inhibitory components. Whether this is chiefly related to response noise, poor stimulus-driven activity, or complex feature

detection not revealed by electrophysiology is unclear.

In contrast to the L1- and L2-regularised kernels, separable kernels had little to no plausible structure and contained significant 'banding'. These patterns were most likely artefacts arising from inseparable structure. However, areas containing strong excitatory or inhibitory components in the inseparable kernels were not banded in separable kernels. In the fourth neuron, excitation at $\sim$ 20 kHz and adjacent inhibition was maintained. To measure the degree of structural similarity between STRFs, I calculated the correlation coefficients for pairwise comparisons (Fig. 3.3C). L1- and L2-regularisation shared the most similarity (mean $CC = 0.68 \pm 0.20$), while separable and L1-regularised kernels were least similar ($CC = 0.25 \pm 0.14$). L2-regularised and separable kernels were slightly more similar with a mean CC of 0.41 ± 0.13. All mean CC values were significantly different from one another (one-way ANOVA with Tukey-Kramer test; see Fig. 3.3C and caption).

Next, I compared performance of STRFs with different regularisers on the same set of held-out data (Fig. 3.3D). L2-regularisation (mean $CC_{\text{norm}} = 0.18 \pm 0.43$) performed significantly better than separable kernels ($CC_{\text{norm}} = 0.14 \pm 0.51$; one-way repeated measures ANOVA with Tukey-Kramer test, $p = 0.0028$), but not better than L1-regularised kernels ($p = 0.11$). L1-regularisation (mean $CC_{\text{norm}} = 0.17 \pm 0.44$) did not outperform separable kernels ($p = 0.14$). The selection of an appropriate regulariser is a critical choice, with implications for STRF structure and predictions (David et al., 2007; Meyer et al., 2017). While L2-regularised kernels provided a slightly higher degree of structural plausibility and predictive power, I reasoned that L1-regularisation is likely to eliminate a significant portion of measurement noise and reduce the effect of stimulus correlations. The remainder of STRFs presented in this work are thus L1-regularised.

3.3.3 Tuning properties of fluorescence-based STRFs

Although STRFs are estimated using complex stimuli, they often reveal tuning in agreement with simple tuning curves derived from pure tone responses (Eggermont et al., 2011; Linden et al., 2003; Massoudi et al., 2015). To test this, I compared frequency tuning properties measured with DRC-based STRFs and pure tone FRAs (see *Section 3.2.12*), and found that several neurons showed good agreement (Fig. 3.4A). STRFs also exhibited inhibitory tuning that was absent from the FRA, such as sidebands like those shown for the second and fourth example neurons (right side, Fig. 3.4A).

Some neurons showed some correspondence between the FRA and STRF best frequency (Fig. 3.4B), and these values were weakly correlated (Spearman's rank correlation coefficient, $\rho = 0.14$, $p = 0.02$). The median best frequency measured using FRAs was 9.5 kHz (IQR 9.1-14.5), and 10.2 kHz (IQR 5.7-15.8) using STRFs. Interestingly, a best frequency of 8-16 kHz was overrepresented in FRAs and this is plausible given these frequencies form the core of the audible spectrum in mice (Panniello et al., 2018). Comparison of spectral bandwidths also revealed differences between STRFs and FRA results (Fig. 3.4C). The median bandwidth was significantly greater in FRAs (0.57 octaves, IQR 0.47-0.84) compared with STRFs (0.26 octaves, IQR 0.24-0.32; Wilcoxon sign rank test, $p = 3.8 \times 10^{-37}$). Most STRFs (241/250, 96.4%) had bandwidths narrower than half an octave. All STRFs had bandwidths narrower than 1 octave, and there was no correlation between FRA and STRF bandwidth ($\rho = 0.07$, $p = 0.24$). DRC stimuli were constructed using narrower spaced frequency bands than the pure tone frequencies, so it is possible that some FRA bandwidths were overestimates due to coarse frequency sampling. Despite this, one would expect to observe a greater variety of bandwidths in STRFs. Again, it is possible that speckling in kernels caused fragmentation of broader excitatory regions, however the Gaussian smoothing

should have at least partially counteracted this. In extracellular electrophysiology, broader bandwidths may also arise from poorly isolated units leading to broadening and smoothing of response areas. FRAs shown here can be attributed to single-neuron activity with a high level of confidence.

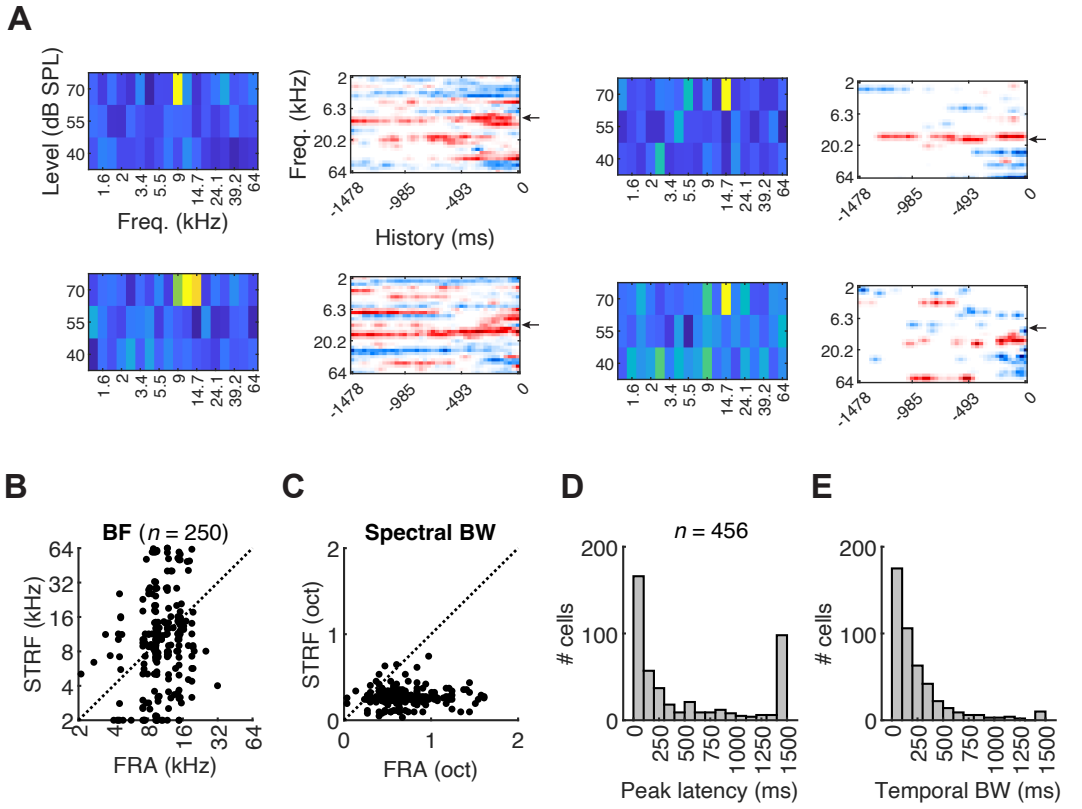


Figure 3.4: Tuning properties measured using STRFs. A. FRAs and STRFs for 4 exemplary neurons. **B.** Comparison of STRF and FRA best frequency and **C.** spectral bandwidth in neurons with frequency-tuned FRAs (n = 250). BF = best frequency, BW = bandwidth **D.** Distribution of STRF peak latency and **E.** temporal bandwidth (n = 456).

Next, I examined STRF temporal tuning and found a spread of peak latencies (Fig. 3.4D). Approximately a third of neurons (166/488, 34%) had peak latencies of less than 100 ms, which is feasible given first spike latencies in auditory cortex ($\sim$ 20ms) combined with indicator rise times ($\sim$ 10-20 ms) (Linden et al., 2003; Morrill and Hasenstaub, 2018; *Chapter 2*). Median peak latency was 217.5 ms (IQR 35-1073)

across all neurons and latencies substantially longer than 100 ms may have been due to slow GCaMP kinetics as well as noisy structure in STRFs. Three of the examples shown in Fig. 3.4A show considerable energy more than 1000 ms into stimulus history. Auditory cortex may integrate over such a duration (Norman-Haignere et al., 2022; Ulanovsky et al., 2003, 2004), however this is not typically revealed by STRFs derived from electrophysiological data (David and Shamma, 2013; Kowalski et al., 1996a). The duration of responses in the frequency-averaged temporal profile was measured as the 'temporal bandwidth'. These values also varied considerably across cells, with many neurons (175/488, 36%) having bandwidths below 100 ms (Fig. 3.4E). The median temporal bandwidth was 147.8 ms (IQR 56.3-316.7). This is surprising given the duration of temporal bandwidths expected for GCaMP8s, which should be on the order of its half-decay time of ~ 250 ms (see *Chapter 2*). Short temporal bandwidths may have arisen from the same process causing the overabundance of narrow spectral bandwidths, where truly broader STRF components are fragmented by noise or over-regularisation.

3.3.4 Spike deconvolution

Despite progressive optimisation of indicator technology, GCaMP kinetics may limit STRF estimation by smoothing the underlying spike train. For instance, a GCaMP8s transient evoked by a single action potential will rise and decay over ~ 250 ms, which is at least 25 times longer than a typical peristimulus-time histogram bin. These slow decay times may introduce STRF artifacts that appear as temporally correlated stimulus features, as the long florescence response to a sound feature captures other stimulus properties that come after it. These correlations can be cancelled out with the use of longer and more varied stimuli, but this required longer imaging times (see *Chapter 2*). Another potential way to avoid the effects of the slow florescent decay time is to attempt to convert florescent calcium traces to the more temporally precise spiking

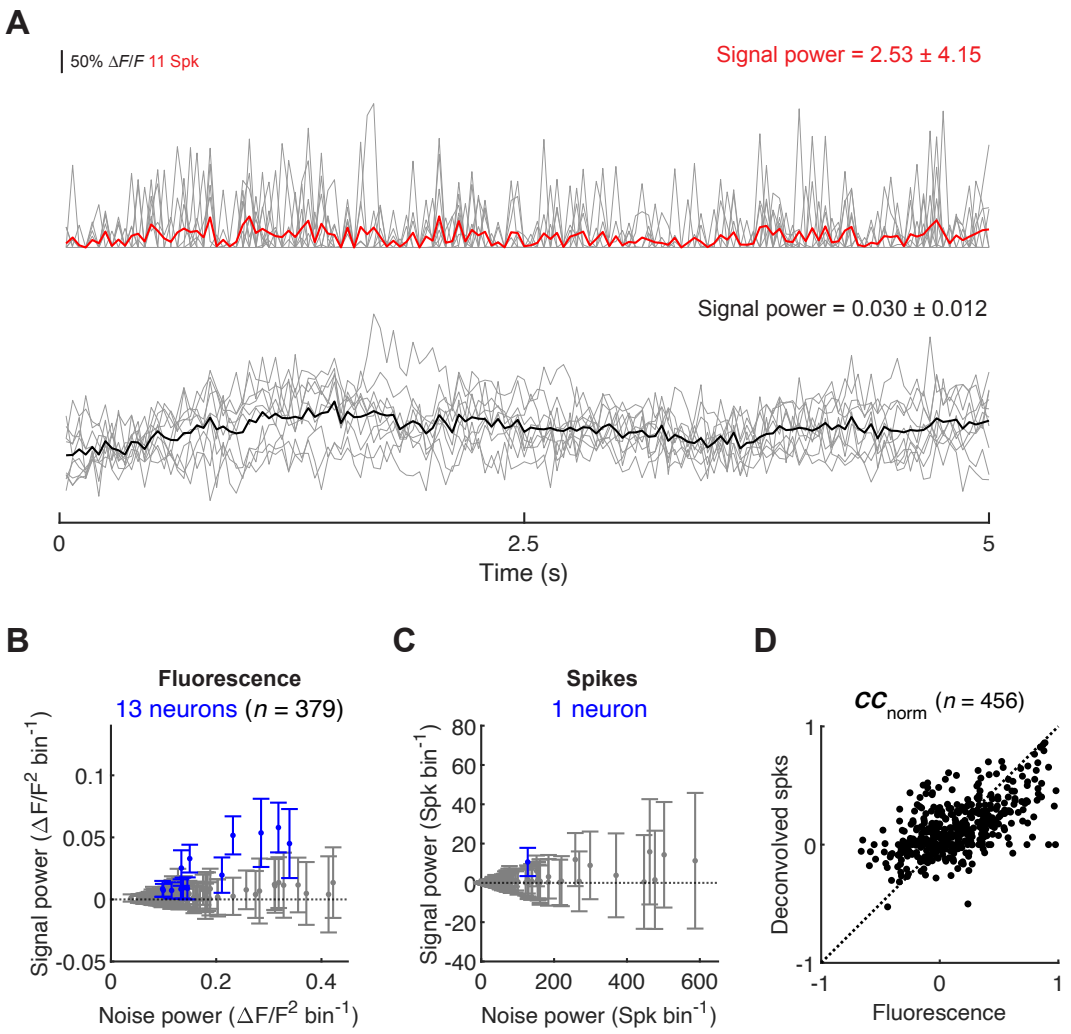


Figure 3.5: Spike deconvolution reduces the proportion of neurons included for analysis. A. Deconvolved spiking (top) and $\Delta F/F$ (bottom) for a neuron during presentation of DRC. Responses to 10 individual stimulus trials (grey) are shown with trial-average overlaid in red (deconvolved spiking) and black ($\Delta F/F$). Signal power is given $\pm$ standard error. **B.** Signal power-noise power plot calculated using deconvolved spikes and **C.** $\Delta F/F$ for one imaging field ($n = 379$). **D.** Comparison of CC_{norm} for STRFs fit to deconvolved spikes and $\Delta F/F$ in all neurons ($n = 456$). Note that cells with extreme CC_{norm} values are not shown ($n = 11$).

events that underlie them, in an estimation process known as 'spike deconvolution'. To investigate whether spike deconvolution can improve STRFs, I compared kernels fit to fluorescence data and their corresponding deconvolved spikes. I used a simple deconvolution algorithm (Pachitariu et al., 2016, 2018a), which provided a discrete spiking metric, *Spk*, at query points specified by imaged frames (Fig. 3.5A). This output resembled a coarsely binned peristimulus-time histogram and was trial-averaged before fitting STRFs.

First, I wanted to understand whether deconvolution changes the proportion of neurons included for analysis. Applying the same criteria, I found that only 0.9% of cells were included (119/13,214 with significant signal power), which was substantially less than when using fluorescence data (3.4% inclusion). Results for an example field (Fig. 3.5B, C) demonstrate this: only 1 cell was included based on deconvolved spikes, whereas 13 cells were included using fluorescence traces. This is surprising as the significantly noisy baseline of fluorescence traces was eliminated by deconvolution. *Spk* traces may have contained false positive events, especially considering that the algorithm did not enforce sparsity beyond a non-negativity constraint. The algorithm may have also performed variably, providing poorer predictions on the activity of very sparsely or densely active cells. Additionally, spiking is itself a sparse quantity since *Spk* increases to some value and often instantaneously drops to zero in adjacent bins. Response variability causing mismatch by even a single bin across trials may have disproportionately affected estimates of signal power (Pachitariu et al., 2018a). In contrast, fluorescence transients decay over longer durations and small temporal perturbations across trials are less likely to affect the signal power. Overall, deconvolution increased the number of cells qualifying for analysis in only 2 fields however this was marginal (an increase from 1 to 2 cells in one field, and from 3 to 4 in the other). As a result, cells were included for further analysis based on signal power in their fluorescence responses.

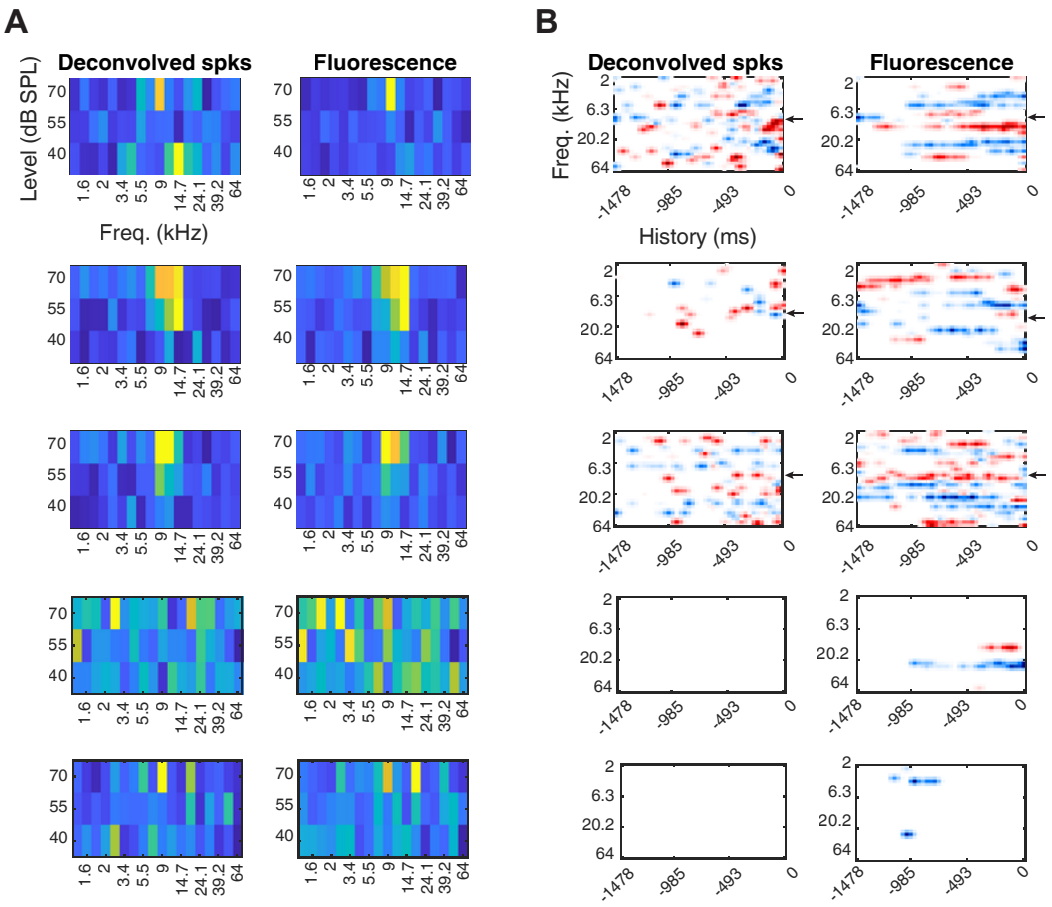


Figure 3.6: STRFs fit to deconvolved spiking. A. FRAs and **B.** STRFs for 5 cells obtained using deconvolved spiking (left) and fluorescence (right). Note that STRFs in fourth and fifth rows have no annotated FRA best frequency (black arrow), because FRA tuning did not reach statistical significance.

The results for five neurons are shown, with FRAs (Fig. 3.6A) and STRFs (Fig. 3.6B) fit to deconvolved spikes and fluorescence traces side-by-side. FRAs were largely similar whereas STRFs exhibited considerable differences. Components in deconvolved spiking STRFs did not appear temporally elongated as in fluorescence STRFs although kernels were similarly noisy. STRFs for the first neuron (Fig. 3.5A, B first row) showed agreement in frequency tuning and recapitulated the FRA. The effect of deconvolution was less clear in the other neurons with only modest improvements in STRF structure (Fig. 3.5B, second and third row). Deconvolution generally did not improve the appearance of kernels, regardless of whether (Fig. 3.5B, fourth row) or not (Fig. 3.5B, third and fifth row) fluorescence STRFs had plausible structure. STRFs fit to fluorescence and deconvolved spikes performed similarly on the same held-out test set (Fig. 3.5D). The mean CC_{norm} for deconvolved spiking STRFs was 0.15 ± 0.30 compared with 0.17 ± 0.43 for fluorescence STRFs, with no significant difference between these values (paired t-test, $p = 0.66$). Deconvolved spiking responses could therefore be fit and predicted with similar success to fluorescence by kernels lacking biologically plausible structure. This may mean that deconvolved spikes, although informative, cannot be used to recover filters which outline the tuning properties of neurons. The relationship between deconvolved spikes, fluorescence, and the underlying spike train evoked by complex stimuli is thus unclear.

3.3.5 Tuning of deconvolved spike STRFs

Median best frequency in STRFs derived from deconvolved spikes was 9.8 kHz (IQR 4.5-15.1), and 9.5 kHz (IQR 9.1-12.9) for deconvolved spike FRAs (Fig. 3.7A). The overrepresentation of mid-range frequencies in FRAs observed above in florescence data persisted here, and there was no correlation between FRA and STRF best frequency (Spearman's rank correlation coefficient, $\rho = 0.10$, $p = 0.20$). Median spectral bandwidth (Fig. 3.7B) was significantly narrower in STRFs (0.29 oct, IQR 0.24-0.36)

compared to FRAs (0.77, IQR 0.57-1.21; Wilcoxon sign rank test, $p = 5.0 \times 10^{-19}$).

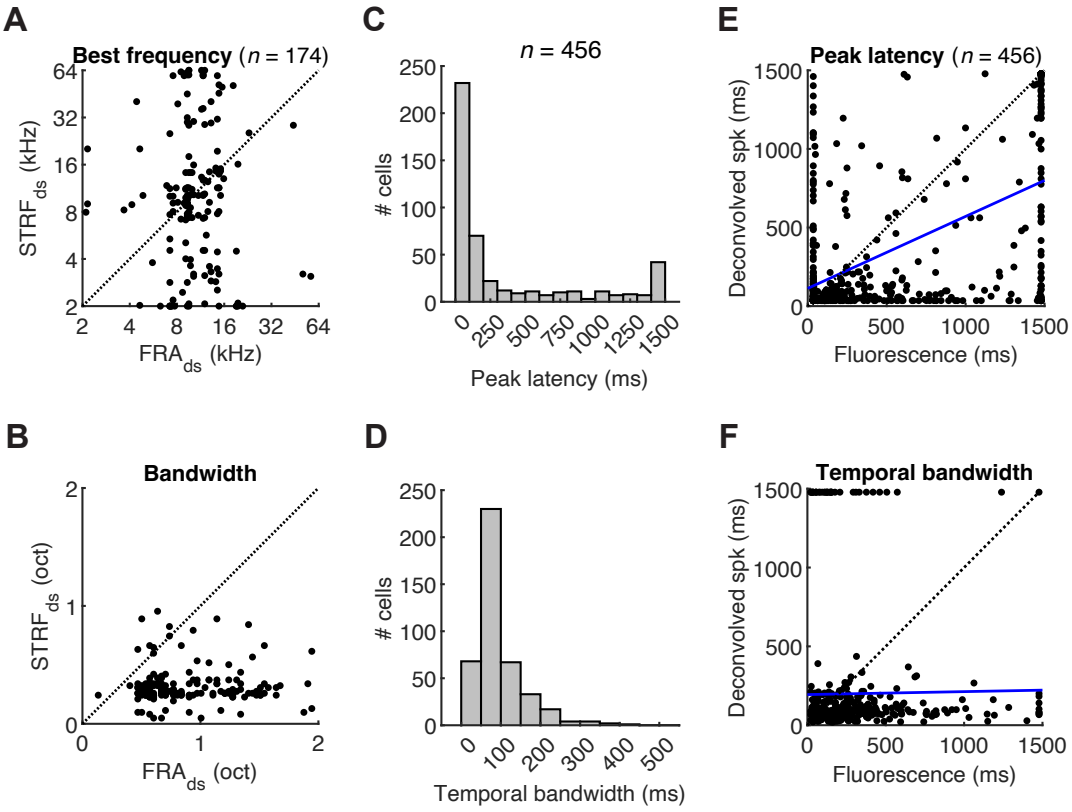


Figure 3.7: Tuning properties of deconvolved spiking STRFs (STRF_{ds}). A. Comparison of (STRF_{ds}) and deconvolved spiking FRA (FRA_{ds}) best frequency and **B.** spectral bandwidth in neurons with frequency-tuned FRA_{ds} ($n = 174$). **C.** Distribution of STRF_{ds} peak latency and **D.** temporal bandwidth ($n = 456$). **E.** Comparison of peak latency and **F.** temporal bandwidth measured in STRF_{ds} and fluorescence STRF ($n = 456$). Blue lines indicate linear best fits.

The median peak latency was 97.0 ms (IQR 35.0-513.7) and half of neurons (232/462, 50.2%) had latencies of less than 100 ms (Fig. 3.7C). The median temporal bandwidth (Fig. 3.7D) was 84.5 ms (IQR 66.9-126.7). Note that 36 neurons (7.4%) had large outlier bandwidths between 1450-1500 ms, and these are not shown in the plot in Figure 3.7D. These large outliers were also observed in a minority of fluorescence STRFs, and were likely due to noise in kernels. Peak latency was correlated between fluorescence and deconvolved spiking ($\rho = 0.39$, $p = 4.0 \times 10^{-18}$; Fig. 3.7E), as expected from

simulations (see *Chapter 2*). Temporal bandwidth was also weakly correlated across spike and florescent STRFs ($\rho = 0.18$, $p = 1.4\times10^{-4}$; Fig. 3.7F). Both peak latency and temporal bandwidth were significantly shorter in deconvolved spiking STRFs compared to fluorescence ($p = 2.5\times10^{-15}$ and $p = 2.9\times10^{-13}$, respectively). These measures are plausible and may reflect some success of the deconvolution algorithm in recovering the temporal features of the spiking response. However, since deconvolved spikes did not afford an overall improvement in the number of neurons included, STRF structure, tuning properties, or predictive power, the remaining analyses in this chapter use fluorescence traces.

3.3.6 Effect of stimulus density

Linear STRFs are not only a product of the underlying stimulus response function, but also of interactions between stimulus statistics and neuronal nonlinearities (Christianson et al., 2008; Nagel and Doupe, 2008; Theunissen et al., 2000). For instance, DRCs with fewer tone pips per second yield STRFs with simpler, mostly excitatory structure, likely due to a failure to recruit two-tone and forward suppression nonlinearities of auditory neurons (Ahrens, Linden, and Sahani, 2008; Blake and Merzenich, 2002). Additionally, sparser DRCs evoke higher maximum evoked spike rates. Two principal issues facing fluorescence traces when estimating STRFs are low signal-to-noise ratio and temporal smoothing of the underlying spike train. Therefore, I reasoned that stimulus density may be important when imaging for STRF estimation, since it can be adjusted to drive temporally sparser yet stronger responses in neurons.

To address this possibility, I designed three sets of DRC stimuli differing in tone pip density (Fig. 3.8A). During imaging sessions, mice were presented with either the low and medium, or medium and high density stimuli. The proportion of neurons included for analysis was dependent on stimulus density: 1.2% (47/3935) had significant signal power when presented the low density DRC, 4.1% (344/8421) with the medium density

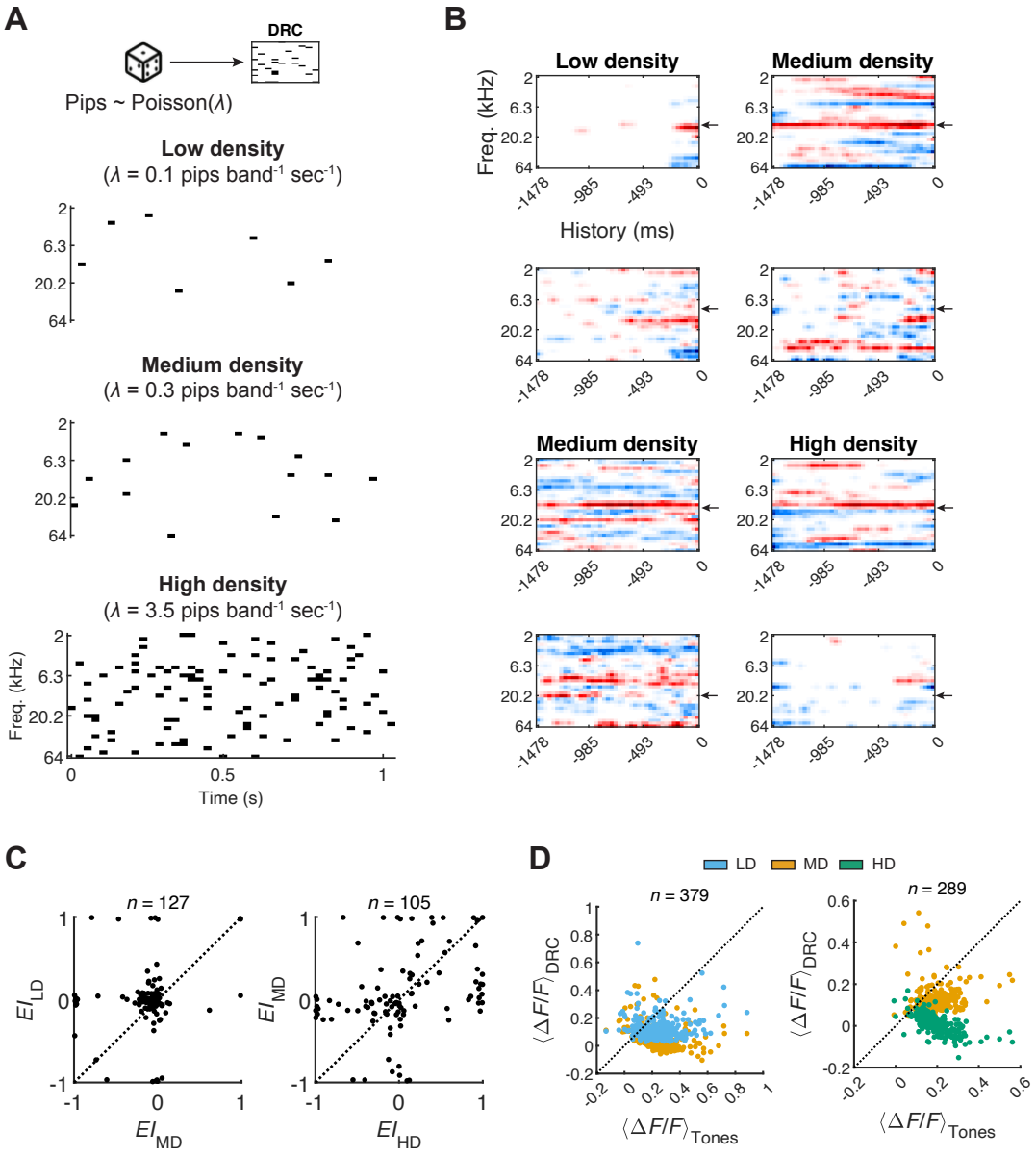


Figure 3.8: STRF structure and strength of evoked-responses are dependent on DRC tone pip density. A. Design of DRC stimuli. Tone pips were generated by a Poisson process with rate equal to tone pips band^{-1} sec^{-1} (inset). Low, medium and high density DRC spectrotemporal envelopes. **B.** STRFs fit to four neurons using different stimulus densities. **C.** Comparison of EI in STRFs fit to low and medium ($n = 127$, left) or medium and high ($n = 105$, right) density DRCs. **D.** Comparison of mean stimulus-evoked response between DRC, $\langle \Delta F/F \rangle_{\text{DRC}}$, and pure tone stimuli, $\langle \Delta F/F \rangle_{\text{Tones}}$. Each scatter shows cells from a single imaging field where animals were presented low and medium density DRCs (left, $n = 379$), and medium and high density DRCs (right). LD = low density, MD = medium density, HD = high density.

DRC, and 7.3% (63/858) with the high density DRC. STRFs are shown for four neurons (Fig. 3.8B). For the first two cells (Fig. 3.8B, first and second rows), the low density stimulus yielded sparser kernels with clearer tuning. Excitation at the $\sim$ 18 kHz best frequency was temporally smoothed in the medium density condition and most energy was located within the first 250 ms of stimulus history. Interestingly, the third and fourth (Fig. 3.8B, third and fourth rows) neurons showed sparser STRFs under the higher density condition, with some excitatory and inhibitory components maintained.

To test the hypothesis that denser stimuli probe inhibitory regions, I calculated the excitatory-inhibitory index (EI, see *Section 2.2.7*). This measure represents the balance of excitation and inhibition in STRFs, ranging from -1 (inhibition with no excitation) to +1 (excitation with no inhibition). In neurons imaged with two DRC stimuli, STRFs were fit separately to either stimulus set and EI was calculated for each (Fig. 3.8C). Neurons were included provided their responses to at least one DRC contained significant signal power. Within neuron comparison revealed a significantly greater mean EI for low density (-0.002 $\pm$ 0.39) compared with medium density stimuli (-0.14 $\pm$ 0.34; paired t-test, $p = 0.0019$; Fig. 3.8C, left). For neurons presented with the medium and high density stimuli, mean EI was 0.07 $\pm$ 0.48 and 0.04 $\pm$ 0.56, respectively (Fig. 3.8C, right). This comparison did not reveal a significant difference ($p = 0.64$). Therefore, sparser stimuli may have led to STRFs with less inhibitory structure, though this conclusion was not supported in all analyses. As mentioned with regards to tuning properties, STRFs may have been too noisy to allow an effective comparison of excitatory-inhibitory structure. Finally, performance on test data also depended on stimulus density: mean CC_{norm} was lowest in the low density (0.07 $\pm$ 0.31) and highest in the high density condition (0.21 $\pm$ 0.62). Mean CC_{norm} using the medium density DRC was 0.18 $\pm$ 0.41.

Next, I wanted to measure the ability of stimuli to evoke fluorescence changes. Count-

ing fluorescence events during complex stimuli is challenging as they do not evoke discrete transients on relevant timescales (e.g. 25 ms tone pips). As a simple measure of response amplitude, I calculated the time-averaged fluorescence, $\langle \Delta F/F \rangle$, by averaging $\Delta F/F$ over all frames during which stimuli were presented (excluding 1 s of onset response). $\langle \Delta F/F \rangle$ was dependent on stimulus density (Fig. 3.8D), and was found to be highest for pure tones (mean 0.15 ± 0.13) and lowest for the high density DRC (0.07 ± 0.15). The mean $\langle \Delta F/F \rangle$ over all neurons imaged was 0.12 ± 0.08 using the low density DRC and 0.11 ± 0.11 using the medium density DRC. Lower $\langle \Delta F/F \rangle$ with denser stimuli may be explained by frequent yet weaker responses driven by denser stimuli, as reported in Blake and Merzenich, 2002. Denser stimuli may sample features of interest more frequently with suppression and/or adaptation accounting for lower magnitude events. Given the limitations posed by imaging noise and sensitivity of GCaMP sensors, this result has implications for imaging experiments, where low amplitude and phasic neural responses are often missed (Huang et al., 2021; Wei et al., 2020). However, based on signal power and performance on held-out data, the results supported the use of high-density stimuli. Further work varying other parameters (e.g. tone pip contrast and duration) is required to optimise DRC stimuli for imaging experiments.

3.3.7 Fluorescence responses driven by complex stimuli

Significant noise in STRFs made it difficult to interpret their structure and tuning properties. This may have arisen from low signal power even in neurons qualifying for analysis. Typically, electrophysiological recordings yield a higher proportion of stimulus-driven units, at least partly due to isolation of highly active neurons (Lohse et al., 2024). For instance, application of the same criteria to a mixed dataset in mice and rats resulted in the inclusion of $\sim 45\%$ of units (Sahani and Linden, 2003). Despite bias towards well-driven units in electrophysiology, the comparative proportion

of qualifying cells in this dataset was surprisingly low.

Signal power may have been low for at least two reasons: (i) robust yet variable responses, which could be considered 'true' nonstationarity; or, (ii) poor stimulus-evoked activity which resembles response variability. The latter can be further divided into poor responses due to stimuli that suboptimally probe the neurons' complex receptive fields, and prohibitively low signal to noise ratios under the current imaging conditions. Responses across the dataset were often nonstationary and of variable strength (Fig. 3.9A). Note that the four large, approximately equidistant transients observed in the first neuron (Fig. 3.9A, first row) were likely large onset responses which persisted beyond the first second of the stimulus response and so were not removed. Similar onsets are also observed marking the beginning of the first and second stimulus in the second neuron (second row). The third neuron (third row) was an example where there was stimulus-driven activity in the response, with a reasonable degree of response variability (noise ratio = 14.17) and a structurally plausible STRF. However, poorly responsive neurons (fourth row) were also maintained in the analysis based on signal power. This may suggest that adopting an events-based criterion (e.g. minimum number of transients) may be suitable for fluorescence data.

I found that normalised signal power (signal power/total power) was positively correlated with $\langle \Delta F/F \rangle_{\mathrm{DRC}}$ (Spearman's rank correlation coefficient, $\rho = 0.41$, $p = 0$; Fig. 3.9B). Thus, neurons with well-driven fluorescence tended to have a larger signal component in their responses, suggesting that low signal power throughout my dataset was primarily related to poorly-driven fluorescence. Note, I differentiate between poorly-driven fluorescence and poorly-driven *neurons*, assuming there were sparsely yet consistently responding cells whose stimulus-evoked activity was unable to overcome measurement noise. Efficient STRF estimation relies on consistent stimulus-driven activity such that noise is cancelled out in few repeats. The noise ratio (see

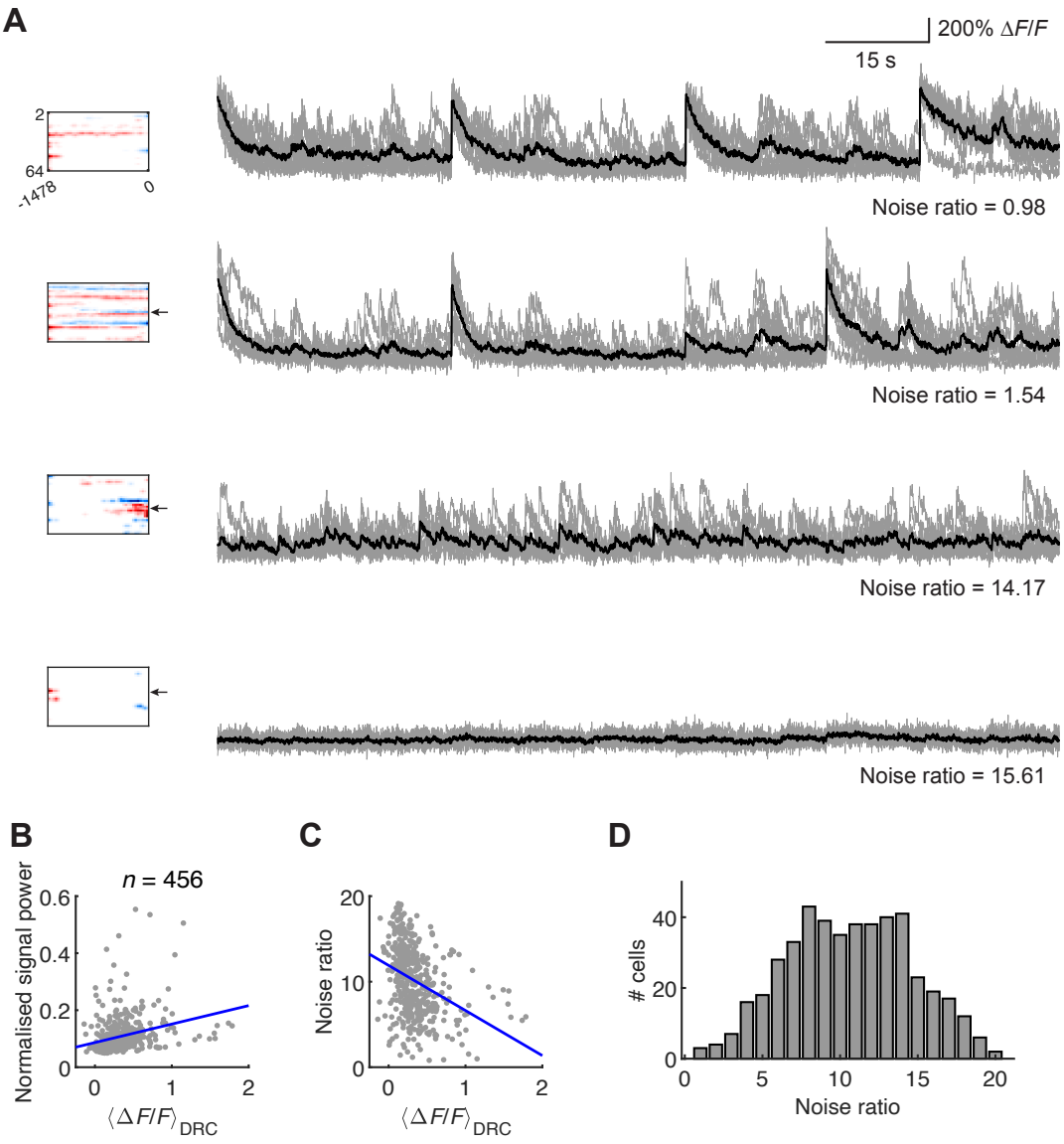


Figure 3.9: Normalised signal power and response variability are related to the strength of DRC-evoked responses, $\langle \Delta F/F \rangle_{\mathrm{DRC}}$. A. Stimulus-evoked responses to 10 stimulus trials (grey) and trial-averages (black) in 4 cells with increasing noise ratio (top to bottom). Insets show corresponding STRFs. **B.** $\langle \Delta F/F \rangle_{\mathrm{DRC}}$ is associated with normalised signal power ($n = 456$). **C.** Noise ratio (intertrial variability) is inversely correlated with $\langle \Delta F/F \rangle_{\mathrm{DRC}}$ (n as in **A**). Note that data shown in **B** and **C** are from medium density DRC. **D.** Distribution of noise ratios for neurons with significant signal power ($n = 456$).

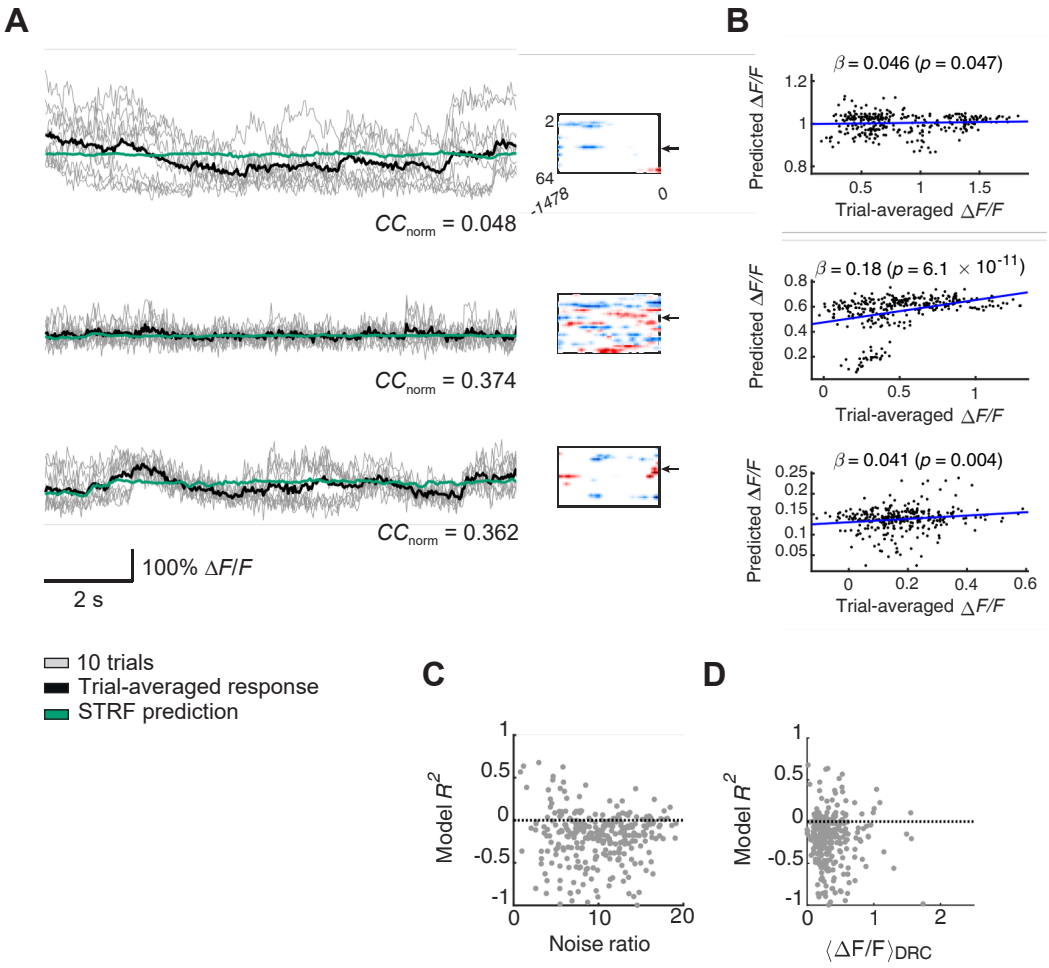


Figure 3.10: Accuracy of STRF prediction on held-out test data. A. Test data for three neurons (STRFs shown as inset, left). Responses shown to 10 stimulus trials (grey) along with the trial-averaged response used to fit the STRF (black) and the predicted response (green). **B.** Scatters showing relationship between trial-averaged response and model prediction for the three neurons in **A**. Best fits (blue) shown with regression statistics (above). **C.** Proportion of linearly-predictable variance (R^2) as a function of noise ratio ($n = 344$). R^2 was given by the total variance minus the mean squared error, all divided by the total variance. **D.** R^2 as a function of $\langle \Delta F/F \rangle_{\mathrm{DRC}}$.

Section 3.2.9), calculated as the ratio of noise power to signal power in the response, measures trial-to-trial variability. I found a negative correlation between the noise ratio and $\langle \Delta F/F \rangle_{\mathrm{DRC}}$ ($\rho = -0.41$, $p = 0$; Fig. 3.9C), indicating that neurons with robustly-driven fluorescence also exhibited lower variability. Note that a reported p-value of zero indicates that the true value was lower than the floating-point threshold in MATLAB (i.e. $p < 2 \times 10^{-308}$). The noise ratio for all imaged cells appeared log-normally distributed between 0.81 and 694,490, but was below 20 for cells with significant signal power (Fig. 3.9D). The mean noise ratio for qualifying neurons was 10.1 ± 4.0 (i.e. 9.0% of response variance was signal). This was well within limits of previous electrophysiology studies, where noise ratio thresholds such as 40 and 200 have been used for unit inclusion (Lohse et al., 2024; Rabinowitz et al., 2012; Rahman et al., 2019; Willmore et al., 2016).

Models of stimulus-response transformations typically overfit to noise, with their performance on held-out data representing a theoretical lower bound on predictive power (Sahani and Linden, 2003). Held-out data for three neurons are shown along with the STRF predictions (Fig. 3.10A) – these responses contained a large degree of variability and little predictable activity. Slow oscillatory structure most pronounced in the first and second neurons (top and middle) would not have been predictable by the STRF which covered up to ~ 1500 ms of stimulus history. For these neurons, STRFs predicted effectively no activity so the held out and predicted responses did not covary (Fig. 3.10B).

The degree of response variability in training set data did not correlate with the ability of STRFs to predict held-out data ($\rho = 0.039$, $p = 0.48$; Fig. 3.10C). Typically, a plot such as Fig. 3.10C has a line of best fit with a negative gradient and a y-intercept corresponding to the theoretical 'maximum' on the proportion of response variance that the model can predict in a noiseless neuron, given finite training data (Sahani and

Linden, 2003). While STRFs of a few neurons performed well on test data, many neurons had very small or even negative R^2 values. Inspection of held-out responses and model predictions in predictive STRFs ($R^2 > 0.5$) revealed only prediction of strong, sustained stimulus-onset responses, which were maintained despite pre-processing and formed the test set in these cells. However, for many neurons, STRFs were unable to predict stimulus-evoked activity. Although all analysed units had some degree of deterministic, stimulus-evoked activity, the quantity and diversity of fluorescence events in many neurons may not have been adequate to fit predictive STRFs. For instance, many neurons showed coherence in their strong onset responses and qualified for STRF fitting, despite the rest of the response containing little stimulus-evoked activity (e.g. Fig. 3.9A, first and second row).

Interestingly, there was no correlation between $\langle \Delta F/F \rangle_{\mathrm{DRC}}$ and R^2 ($\rho = -0.066$, $p = 0.22$), suggesting that average stimulus-evoked activity did not influence the predictive power of STRFs (Fig. 3.10D). It is possible that truly robust responses were driven in too low a proportion of neurons to establish a relationship between stimulus-evoked activity and STRF predictive power. Regardless, neurons exhibiting mostly onset (or generally few) fluorescence transients may have been subject to severe overfitting, leading to model predictions which were anticorrelated with held-out data. Further work understanding the relationship between stimulus-evoked fluorescence, response variability and the predictive power of encoding models will be required.

3.3.8 Slow TORC STRFs

'Ripple' stimuli have an extensive history in the auditory and visual systems, allowing cells' spatio- or spectrotemporal modulation preferences to be probed explicitly (Depireux et al., 2001; Kowalski et al., 1996a, 1996b; Pinter and Harris, 1981). In the auditory system, TORCs are one such stimulus, created by summing moving frequency sweeps called 'ripples' (the equivalent of drifting gratings in the visual system, see *Sec-*

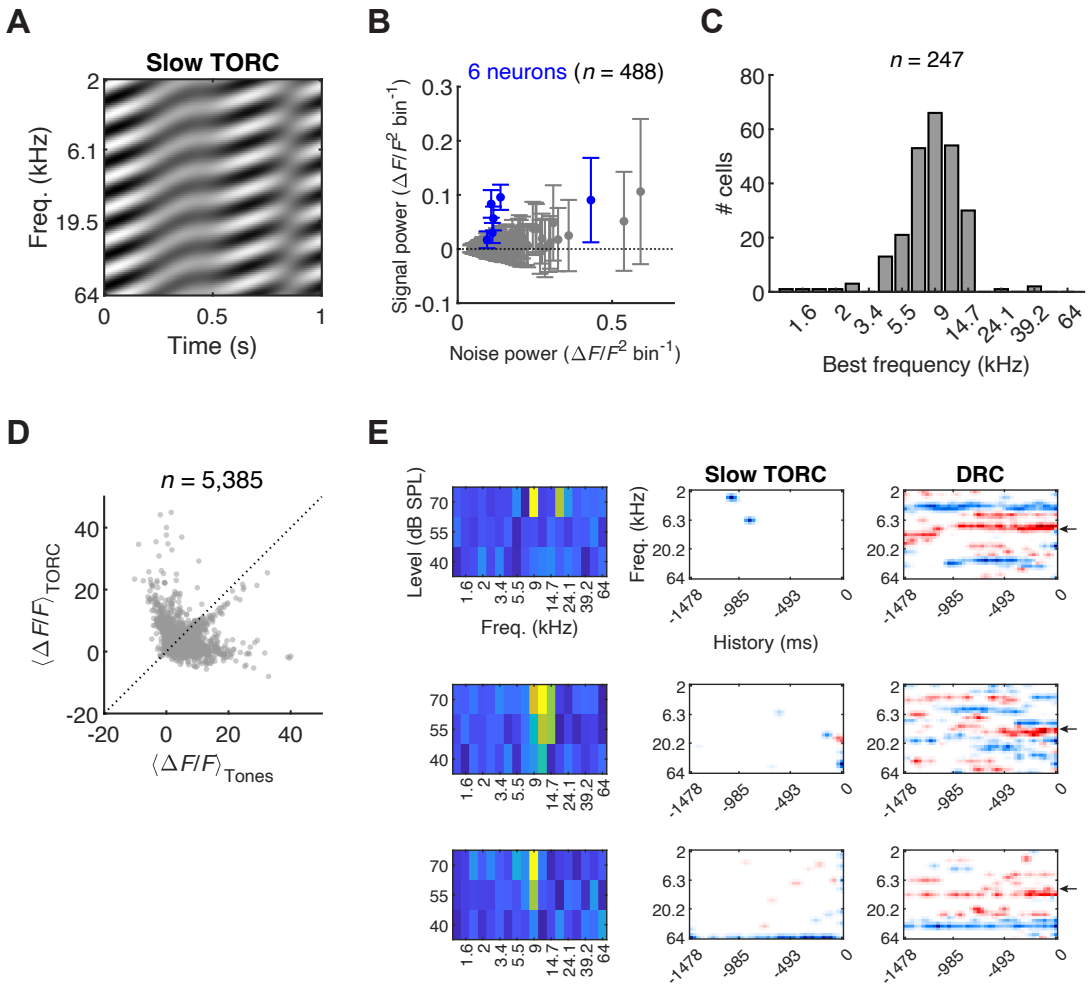


Figure 3.11: Slow TORCs STRFs do not contain discernible structure. A. Spectrotemporal envelope of a slow TORC stimulus. **B.** Signal power/noise power plot in an exemplar field imaged under TORC stimulation ($n = 488$). Cells included for analysis shown in blue. ($n = 6$). **C.** Distribution of pure tone best frequencies for cells with frequency-tuned FRAs ($n = 247$). **D.** Relationship between mean TORC-evoked response, $\langle \Delta F/F \rangle_{\text{TORC}}$, and tone-evoked response, $\langle \Delta F/F \rangle_{\text{Tones}}$ in all neurons presented TORCs ($n = 5,385$). Note all neurons were plotted here since very few would be included based on significant signal power. **E.** FRAs for three example units from this field (left) with respective STRFs obtained using slow TORC (middle) and medium density DRC stimuli (right).

tion 3.2.5). Since modelling work in *Chapter 2* revealed that TORCs band-limited to lower temporal rates are better suited to fluorescence data, the same slow TORC stimuli (Fig. 3.11A) were presented *in vivo*.

Of the 5,385 neurons presented with TORCs, only 16 (0.30%) had responses with significant signal power, which was substantially lower than for DRC stimuli (456/13,214, 3.5%). Results for one field are shown (Fig. 3.11B). This field contained numerous frequency-tuned neurons and the distribution of best frequencies was centred around 9 kHz (Fig. 3.11C). Interestingly, there was an inverse relationship between TORC and pure tone responses where cells driven strongly by one of these stimuli were driven poorly by the other (Fig. 3.11D). This suggests that neurons are driven preferentially by either TORCs or pure tones – potentially due to differences in spike thresholding, saturation or stimulus adaptation between neurons (Dean et al., 2005). Surprisingly, slow TORCs did not recover any meaningful filter structure across the dataset, including in tone-responsive neurons with good DRC-derived STRFs (Fig. 3.11E). Importantly, TORC-responsive cells showed predominantly onset-driven excitation (Fig. 3.12A) or suppression (Fig. 3.12B), and the remaining neurons were poorly-responsive to the stimuli. Although the first second of stimulus responses were removed before fitting STRFs, strong onset responses persisted in a subset of neurons. Such strong responses may be related to the broadband noise carrier of TORCs. Furthermore, the periodicity of TORCs places substantial energy within inhibitory regions of the neuronal stimulus response function, potentially leading to the poor responses here. Evoked spike rates may not have been high enough to cause measurable transients under the current imaging conditions. Additionally, fast modulations missing in slow TORCs may have been needed to drive activity in a subset of neurons, though mammalian auditory cortex is shown to phase-lock to slower envelope features (Depireux et al., 2001; Kowalski et al., 1996a; Miller et al., 2002).

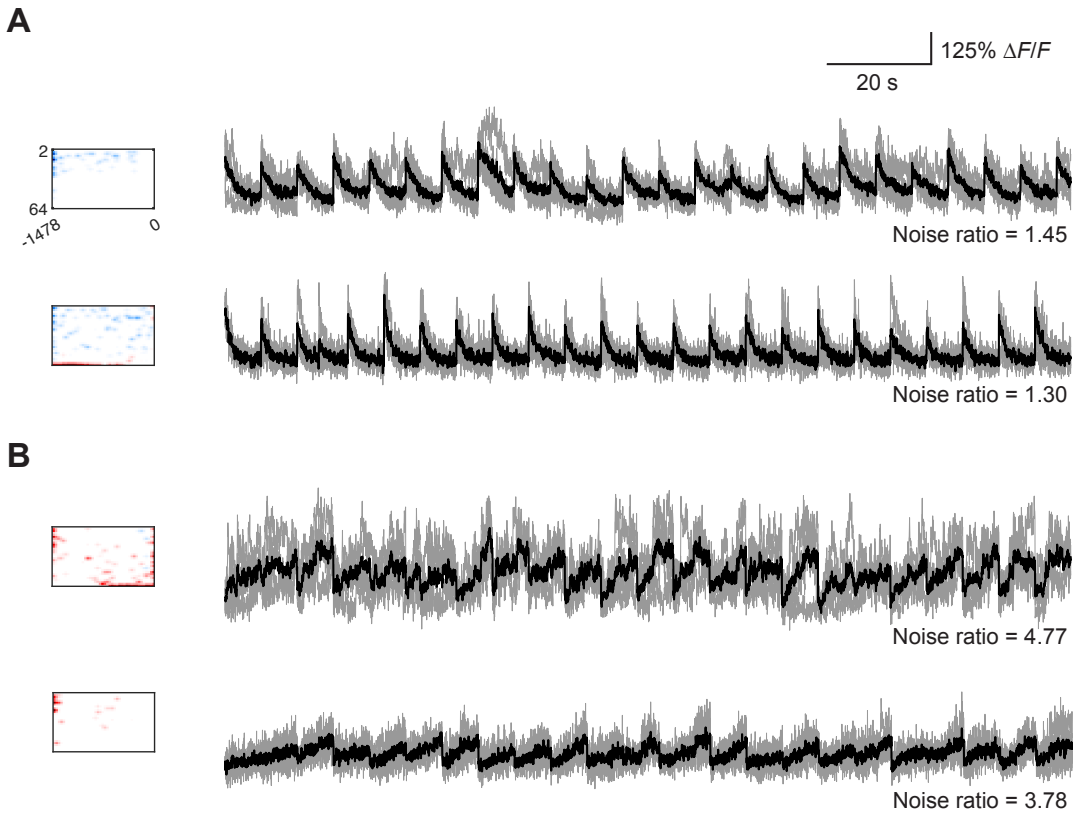


Figure 3.12: TORC-evoked fluorescence responses and respective STRFs (inset, left). A. Cells showing strong onset-driven responses and **B.** onset suppression.

3.4 Discussion

This experiment investigated the feasibility of STRFs fit to *in vivo* two-photon calcium imaging data acquired during the presentation of complex stimuli to mice. The linear STRF encapsulates the encoding properties of individual neurons and contains the stimulus features that are able to drive the strongest response. Here, I demonstrate that STRFs fit to a subset of auditory cortical cells contain structure and predictive power comparable to STRFs fit using classical electrophysiology (Machens et al., 2004; Willmore et al., 2016). STRFs contained slow temporal structure, as predicted by my computational modelling, and sometimes reflected frequency preferences of a neuron measured with pure tones.

3.4.1 Stimulus-evoked fluorescence and inclusion criteria

Although kernels could be fit with success to some neurons, responses across most of the dataset contained low signal power. Comparable electrophysiological work in mouse auditory cortex using DRC stimuli included 92/205 (44.9%) of units, based on the same signal power criterion employed here (Sahani and Linden, 2003). This is considerably higher than here (3.5%) and may reflect a strong bias in electrode recordings towards strongly-driven units. By contrast, a larger proportion of cells sampled by two-photon imaging are weakly or non-responsive to stimuli compared with electrophysiology. The method used here assumed the presence of a constant, stimulus-related component in the response accompanied by stochastic noise. Neurons were required to exhibit significant signal power in their mean stimulus response, defined by a non-zero lower bound of the standard error interval. This criterion proved to be stringent, with $> 95\%$ of cells not included in further analysis, however increased confidence in STRFs fit to the remaining neurons. This is particularly important given that it is unclear if fluorescence traces to complex stimuli contain sufficient information to even linearly

approximate the underlying stimulus response function.

Previous electrophysiological work (Rabinowitz et al., 2012; Rahman et al., 2019; Willmore and King, 2023) has used the noise ratio for inclusion, which does not place an explicit constraint on signal power. The noise ratio is calculated without considering the standard error interval of the signal power and could technically result in spuriously low noise ratios if this interval includes a value of 0. Indeed, adopting a noise ratio threshold comparable to previous studies would have resulted in the proportion of included neurons approaching that in electrophysiology. However, a higher proportion of units contain significant signal power in such datasets (Sahani and Linden, 2003). Inclusion of many cells with potentially non-zero signal power would also decrease confidence in CC_{norm} values, where normalisation of the correlation coefficient by the signal power would lead to spurious values as signal power approaches zero. It is possible that other forms of inclusion criteria may be better suited for imaging data. For instance, inclusion based on a minimum number of stimulus-evoked events may remove inactive neurons while preserving nonstationary ones, some of which have measurable tuning properties. Additionally, inclusion criteria could be combined with requirements on STRF structure (e.g. only including STRF components exceeding some threshold) or by fitting resampled STRFs and including only those which satisfy a signal-to-noise requirement (J. Fritz et al., 2003; J. B. Fritz et al., 2005).

3.4.2 Predictive power

Extrapolating the test set performance of STRFs to zero-noise conditions can give an estimate of the maximum predictable variance on a test set, given finite training data. This was not possible in my data, and many predictions were only mildly or anticorrelated with the test set (Fig. 3.2G, 3.10C), suggesting severe overfitting in cases where stimulus-evoked activity was not adequately strong or diverse. It is possible that fluorescence evoked by complex auditory stimuli generally contains lit-

tle linearly predictable activity. In comparable electrophysiology, measuring STRFs using DRCs in rodents, the theoretical maximum R^2 was 18% (Sahani and Linden, 2003), however fluorescence is subject to additional nonlinearities compared to the underlying spike train, such as those related to GCaMP (e.g. thresholding/saturation and history-dependence) and intracellular calcium dynamics (Rupprecht et al., 2025). Differences in exact model formulation (e.g. number of free parameters, choice of regulariser) and the location of sampled neurons could also be responsible. Importantly, the electrophysiology study recorded from thalamorecipient units which are known to be more linear than supragranular neurons imaged here (C. A. Atencio et al., 2009; Ramirez et al., 2014). Layer 2/3 neurons also have a greater extent of intracortical connectivity and are the targets of top-down projections from prefrontal cortex (Plakke and Romanski, 2014). Indeed, task-related plasticity is greater in these layers when animals engage in a tone detection task (Francis et al., 2018). Therefore, responses in these layers are potentially less determined entirely by sound stimuli and subject to modulatory top-down influences (Soto et al., 2006). Contextual top-down signals may have played a role in our observed nonstationarity, especially given the longer imaging durations used here ($\sim$ 20 min for STRF estimation) and experiments being conducted in awake animals.

The responses of some neurons appeared to be robust yet highly variable across trials of a stimulus (Fig. 3.9A). Nonstationarity across trials can be attributed to various causes (Curto et al., 2009; Deweese and Zador, 2004; Kisley and Gerstein, 1999). First, it may be inherent to neurons themselves. Intertrial variability is even exhibited by neurons stimulated by simple, deterministic electrical stimuli *in vitro*, representing a potentially advantageous computational mechanism for neural networks (Carandini, 2004; Heggelund and Albus, 1978). Weak stimulus-evoked activity can also explain apparent nonstationarity, with very low spike rate events (e.g. 1-2 action potentials) potentially missed by typical calcium indicators (Huang et al., 2021). My imaging

was carried out in layer 2/3 neurons, which are known to exhibit weaker stimulus-evoked responses than those in the deeper layers 4 and 5 that are typically targeted in microelectrode recordings (C. A. Atencio et al., 2009; Hromádka et al., 2008; Ramirez et al., 2014). Spike rates may also generally be lower in auditory cortex than other areas (Carandini, 2004). Thus, apparent nonstationarity may be related to missed detection of low action potential-count events in imaged neurons. Furthermore, fast half-rise times ($\sim$ 5 ms) reported for GCaMP8s are a challenge to capturing short-duration events given our slower imaging frame rate ($\sim$ 28 Hz). Fluorescence peaks triggered by short-duration events may not be sampled reliably at this frame rate, with $\Delta F/F$ dropping within the range of noise by the time the next frame is triggered.

3.4.3 Choice of estimation procedure

STRFs were estimated here by linear regression with L1-regularisation within a generalised linear model framework assuming Gaussian noise. Separable kernels were also estimated and found to contain highly corrupted structure, but nonetheless performed surprisingly well based on CC_{norm}, compared to inseparable kernels. L1-regularisation was ultimately used throughout subsequent analyses because it imposes a prior of sparsity on what is expected to be a slow and noisy STRF (see *Chapter 2*). In some instances, kernels appeared over-regularised with a loss of local smooth structure. Imposing different priors may improve structure and predictive power of STRFs estimated using limited noisy data (David et al., 2007; M. Sahani and Linden, 2002; Sharpee et al., 2004). For example, the automatic locality determination algorithm (Park and Pillow, 2011) achieves this by imposing a prior of global sparsity yet local smoothness, reflecting the biological structure of cortical receptive fields. Alternatively, reductions in model dimensionality by matrix factorisation or assuming parametric forms can improve the structure and predictive capabilities of STRFs (Thorson et al., 2015). For instance, two-photon imaging of direction-sensitive visual cells in *Drosophila* estimated

spatiotemporal receptive fields by calculating a fluorescence-weighted average (i.e. reverse correlation) and fitting a difference of Gaussians model to denoise kernels (Leong et al., 2016). Although, the drifting grating stimuli used evoked robust, periodic $\Delta F/F$ responses providing a suitable substrate for reverse correlation in the first place. Selection of an appropriate fitting procedure for imaging data requires further investigation but reducing overall model dimensionality is likely to have a positive effect.

Interestingly, spike deconvolution did not improve the appearance, tuning or predictive power of STRFs. This could be attributed to variable performance of the deconvolution algorithm in addition to poor stimulus-evoked responses. Although more elaborate algorithms exist, including those with additional biophysically-motivated parameters (Deneux et al., 2016), the simple non-negative deconvolution algorithm used here is shown to perform similarly well at a fraction of the computational power (Pachitariu et al., 2016). A novel approach to recover fast temporal structure from fluorescence data is to use the positive half-wave rectified first derivative of $\Delta F/F$. Application of this method in zebrafish tectum under visual stimulation yielded traces well-correlated with simultaneous electrophysiology and recovered linear temporal filters of cells (Ramdya et al., 2006). This was trialled as a preprocessing step (data not shown), but failed to improve STRF structure or performance during cross-validation. Fitting within the large parameter space encompassed by a 31×40 kernel may require more information than provided by this methodology, where some features of the original trace are lost during differentiation and rectification.

Finally, the linear model used here could be augmented by the addition of a static output nonlinearity (Chichilnisky, 2001), potentially accounting for indicator nonlinearities (e.g. thresholding, saturation) in addition to neuronal nonlinearities. Alternatively, a linear or nonlinear spike-to-fluorescence step could be incorporated following the linear step, with parameters defined *a priori* or fit in conjunction with the STRF.

The temporal properties of the indicator could also be incorporated directly into the STRF as a parametric temporal marginal distribution. It would be of interest to examine systematic failures in predictions to understand whether they can be explained by trivial factors such as spiking nonlinearities or indicator properties. In many cases, held-out data lacked predictable features, highlighting the need to optimise stimuli to drive fluorescence responses before STRFs can be fairly evaluated.

3.4.4 Conclusions

Ultimately, the present work shows promise for fitting valid and reliable STRFs to calcium imaging data, but work addressing the different facets of STRF estimation in imaging experiments is required. These results provide a proof-of-concept supporting additional investigations into this area. Determination of optimal imaging conditions, stimuli and fitting procedures will reveal the principles of efficient experimental design, allowing the full potential of imaging STRFs to be realised.

Chapter 4

General discussion

The results of this thesis show that linear spectrotemporal receptive fields (STRFs) fit to calcium imaging data could provide useful descriptors of neural encoding. By first simulating spike trains and GCaMP fluorescence from a model receptive field, I found that STRFs can recover simulated receptive fields in the presence of noise (measurement and neural noise), indicator nonlinearity, and low spike rates (*Chapter 2).* I then validated these results *in vivo*, showing plausible structure and reasonable predictive power in a subset of STRFs fit to the calcium traces of stimulus-driven neurons (*Chapter 3*). These results represent a key step in developing this methodology, which could uncover new aspects of complex stimulus processing by imaging auditory cortex.

4.1 Fluorescence responses to complex stimuli

Across sessions, reliable stimulus responses were evoked in a minority of imaged cells and even fewer units yielded STRFs whose predictions correlated with test data. Considering that: (a) dynamic random chord (DRC) and temporally orthogonal ripple combination (TORC) stimuli have been applied successfully to estimate STRFs across

various species and preparations (Cooke et al., 2018; David et al., 2009; deCharms et al., 1998; Klein et al., 2000), and (b) imaging durations were comparable to recording durations in neurophysiological studies (J. Fritz et al., 2003; Linden et al., 2003; Willmore et al., 2016), my results suggest that the current methodology requires modification to optimize the yield of responsive neurons.

There is mounting evidence that sparsity is a feature of both cellular responses (i.e. infrequently active and/or slow-spiking units) and global representation in auditory cortex (Liang et al., 2019). Stimuli including tones, frequency sweeps, noise bursts and natural sounds evoke responses in less than 5% of auditory cortical neurons at a given time, as measured with intracellular electrophysiology in unanaesthetised rats (Hromádka et al., 2008). Moreover, half of all neurons show no significant change in firing rate to any stimulus. Extreme sparsity of cellular responses has also been demonstrated with single-cell recording under anaesthesia, where preferred stimuli can be represented by single-spike responses, i.e. a binary code (DeWeese et al., 2003).

In my experiments, events consisting of single (or few) action potentials may have been obscured in fluorescence traces because imaging was performed over large fields, where few pixels are dedicated to each neuron. It is thought that less than 10% of single (and 30% of double) action potential events are captured under population imaging conditions using GCaMP6f (Huang et al., 2021). Therefore, one barrier to fitting STRFs may have been inadequate sensitivity to a number of stimulus-evoked spiking events. Separately, the low proportion of stimulus-driven cells measured by intracellular electrophysiology (Hromádka et al., 2008) may explain the low proportions within each imaged field here. *Chapter 2* included a simulation where transients evoked by single action-potentials were obscured by measurement noise. Despite this, STRFs recapitulating the simulated filter were able to be fit to these calcium traces. This discrepancy may be explained by the linear-nonlinear-Poisson spiking model employed,

where spiking events occurred consistently throughout the simulated response. Spiking was not sparse and may have occurred frequently enough for summated transients to overcome simulated noise. Extending my spiking model to account for phenomena such as refractoriness and adaptation while simulating yet noisier traces could yield different results that better approximate the behaviour of real neurons.

Importantly, indicators are validated under idealistic conditions where one or a few cells occupy the field of view. This increases the signal-to-noise ratio and allows acquisition at higher rates, ensuring that small amplitude transients are better captured. By contrast, observable fluorescence transients in my experiments may have occurred only when units were strongly driven. STRFs fit to these transients would be thresholded, in that stimulus features causing smaller deflections in firing rate would not contributing to the estimate. Imaging over a smaller field of view may counteract this issue, but would diminish one of the principal advantages of microscopy – that is, measuring the responses of many individual neurons simultaneously.

GCaMP8s was selected here because it provides a good balance between kinetics, linearity and sensitivity. *Chapter 2* simulations using parameters reported for GCaMP8s provided good STRF fits and marginal gains using a faster indicator. Interestingly, I found that transients *in vivo* would often decay over considerably longer timescales (> 1 s) than reported in the original GCaMP8 paper (Zhang et al., 2023, although see Rupprecht et al., 2025). This disparity may be due to sustained neuronal activity in response to complex sounds compared to the single action potential stimulation used when validating the indicator (Broussard et al., 2025; Helmchen et al., 1996). However, long decay times were also observed with pure tones, where a response considerably longer than the 200 ms stimulus would not be expected. Regardless of the cause, the kinetics of fluorescence traces appeared slower than expected. Understanding the relationship between spiking and GCaMP fluorescence evoked by complex sounds will

ultimately require imaging with simultaneous intracellular electrophysiology, which have yet to be performed for complex sounds in auditory cortex.

Importantly, STRF estimation requires sufficiently diverse stimulus-driven events to identify features of interest while cancelling out noise. Comparison of spatiotemporal receptive fields in barrel cortex, estimated using both membrane potential and spikes, suggests that at least 1000 spikes are needed to obtain reliable estimates (Ramirez et al., 2014). Though binning spikes in peristimulus time histograms reduces the number of perturbations from the mean firing rate to less than 1000, the number of transients imaged per cell here were generally far lower. Reaching this many evoked events may require prohibitively long imaging times, thus necessitating stimuli which drive numerous large-amplitude events in a short time.

Furthermore, neurons imaged here were in cortical layers 2 and 3, which respond more sparsely to sounds (including DRC stimuli) compared to thalamorecipient layers (Hromádka et al., 2008; Kang and Kanold, 2024; Liang et al., 2019; Petersen and Crochet, 2013). Lower spontaneous and evoked-rates have also been recorded in layer 2/3 barrel cortex using DRC-like whisker stimuli, and responses had lower signal-to-noise ratios and less linearly predictable components compared with deeper layers (Ramirez et al., 2014). Additionally, layer 2/3 circuits are more likely to exhibit changes based on contextual cues (Allen et al., 2003; Trachtenberg et al., 2000; Zhou et al., 2014), potentially due to denser innervation from corticocortical projections (Hackett, 2011; Plakke and Romanski, 2014). Optical approaches predominantly target supragranular layers (Birkner et al., 2017), however imaging deeper in cortex may yield a greater proportion of stimulus-driven cells with different contributions to hierarchical cortical processing (C. A. Atencio et al., 2009; Vasquez-Lopez et al., 2017). Combining imaging across cortical laminae could further uncover how STRFs are shaped by interactions within cortical microcircuits.

4.2 Stimulus and model optimisation for calcium imaging

The selection of appropriate stimuli represents a key variable in maximising the proportion of units recruited and driving robust activity. Considering the sparsity of auditory cortical responses combined with my imaging conditions, it is unclear whether the proportion of stimulus-driven cells here is an underestimate. Nonetheless, the stimuli used could be modified with aim of response optimisation for two-photon imaging. For instance, tone pip duration, amplitude and contrast could be varied in DRCs. Alternatively, different stimulus classes could be used with the obvious choice of natural sounds. Importantly, STRF estimates obtained by natural stimuli are very different to those obtained by other stimuli, reflecting the strongly nonlinear encoding mechanisms recruited by the natural features that are not found in many synthetic stimuli (Valentine and Eggermont, 2004; Woolley et al., 2006).

Response maximisation has also been pursued as an 'optimisation' problem, which can be addressed analytically (A. R. Chambers et al., 2014; Edin et al., 2004; Paninski et al., 2007). For instance, the 'evolutionary optimisation' algorithm employed in A. R. Chambers et al., 2014 carried out online manipulation of different stimulus dimensions based on single-unit feedback in awake mice, converging on stimulus dimensions capable of evoking higher spike rates than control stimuli. While such methods can identify stimulus features eliciting robust spiking, they typically sample a sparser and highly biased stimulus subspace compared with traditional stimuli used for STRF estimation. Additionally, tailoring stimuli to drive strong responses in individual neurons (as in the above) or under particular conditions (e.g. calcium imaging), does not necessarily provide information about how sounds are processed naturally in auditory cortex (Hromádka et al., 2008). Thus, stimulus optimisation is part of a broader optimisation problem, where the quality and ethological validity of evoked responses should be considered together.

The large stimulus space encompassed by all sounds an animal can hear makes it particularly challenging to select appropriate stimuli to probe the properties of individual neurons. With the relationship between the auditory input and spiking far from clear, measuring fluorescence simply introduces an additional layer of complexity which has the potential to interact with stimulus statistics and neuronal nonlinearities (Christianson et al., 2008; Theunissen et al., 2000; Woolley et al., 2006). Understanding these interdependencies may require spectral and temporal parameters to be varied separately, building a ground-up understanding of spectrotemporal processing as measured by imaging. For instance, calculation of modulation transfer functions of calcium STRFs may reveal properties (e.g. temporal low-pass filtering) imposed by the imaging procedure which limit the broader applicability of this method.

Reformulation of the standard STRF may also be useful in imaging experiments. For instance, the 31×42 kernels fit here contained over 1300 free parameters which could be reduced by factorisation of the STRF into low dimensional spectral and temporal components (Park and Pillow, 2013; Rabinowitz et al., 2012; Simon et al., 2007). It has been shown that simpler, low-rank formulations can outperform the standard STRF in predicting single-unit responses to animal vocalisations in primary auditory cortex, and that they are less susceptible to error in neurons with lower signal-to-noise ratios (Thorson et al., 2015). Here, I used separable kernels which fit spectral and temporal marginals, constraining STRF dimensionality as much as possible while maintaining a nonparametric form. It may be that the separable model did not contain suitable features to fit fluorescence data. Alternatively, assuming parametric forms (e.g. Gaussian, difference of Gaussians, difference of exponentials etc.) for spectrotemporal components may improve both structure and predictive power of calcium STRFs (Leong et al., 2016; Thorson et al., 2015). Finally, different fitting procedures and model formulations may be useful. For instance, generalised linear models including spike history appended to the stimulus matrix have been able to account for spiking phe-

nomena such as refractoriness and bursting (Paninski, 2003). Similarly, accounting for the exponential-like decay process of GCaMP indicators could improve the biophysical accuracy of the otherwise phenomenological linear model.

4.3 The future of imaging STRFs

Receptive field-based models have been a valuable tool in systems neuroscience over the last half-century. While modern electrophysiology makes it possible to record thousands of units simultaneously, optical approaches possess immense potential by allowing stable, long-term access to highly defined populations of neurons (such as those projecting to particular target regions, or particular neuron subtypes). Further, two-photon imaging can be targeted with precision to subcellular compartments, such as dendrites and dendritic spines. Dendritic imaging has already provided insights into the functional organisation of thalamocortical and intracortical projections onto dendritic trees in auditory cortex (Li et al., 2018; Richardson et al., 2009). With further optimisation, the possibility of fitting STRFs separately to dendrites, spines and soma could illuminate how receptive fields are constructed in cortex. While it is known that cortical neurons constructively combine thalamic receptive fields (Miller et al., 2001), imaging could uncover interactions between dendritic inputs and integration at the soma. Combining this approach with behaviour could shed light on the role of different inputs in task-related plasticity. It is thought that top-down inputs from frontal areas modulate response gain and feature selectivity, with stronger task-related plasticity observed in layer 2/3 where intracortical projections are prevalent (Francis et al., 2018). Imaging STRFs within task epochs and longitudinally across training could reveal interactions between top-down and bottom-up inputs. This would help understand how their coordinated activity gives rise to plasticity at different timescales underlying selective attention and memory.

Recent advances in genetically-encoded voltage indicators may also provide a complementary optical approach. Externally loaded dyes have previously been used measure activity in auditory areas in mice (Sawatari et al., 2011) and guinea pigs (Nishimura et al., 2007; Song et al., 2006). More recently, the development of genetically-encoded indicators with superior sensitivity and kinetics have enabled the measurement of subthreshold activity at sub-millisecond resolution (Abdelfattah et al., 2019). Voltage imaging has the potential to revolutionise imaging by combining high spatial resolution with a temporal resolution so far limited to electrophysiology. However, widespread adoption of voltage imaging has been hampered by technical and biological obstacles. Firstly, voltage imaging setups typically generate audible emissions due to high-speed scanning, rendering this technique unsuitable for auditory experiments. Additionally, the number of cells imaged simultaneously is typically fewer than calcium imaging due to the high acquisition rates and sensitivity required to capture fluorescence changes evoked by action potentials (Bando et al., 2019). Imaging durations are typically shorter to avoid photobleaching of voltage indicators (Abdelfattah et al., 2019), and may therefore be unsuitable for behavioural experiments or where more data is needed to fit elaborate models. Nonetheless, voltage imaging has the potential to sample activity in large neuronal populations at a temporal resolution not achieved by calcium imaging.

4.4 Conclusions

STRFs have been used widely in the auditory, visual and somatosensory domains. Although only a linear approximation of the underlying stimulus-response function, they provide an intuitive descriptor of neuronal function which can be easily extended. Measuring spectrotemporal tuning by imaging would unlock a new avenue of experimental work in auditory cortex, investigating complex stimulus representations from

the subcellular to the circuit level. Optimising the stimuli, imaging conditions and procedures used to fit STRFs will help develop this powerful new methodology.

Bibliography

Abbas, P. J., & Sachs, M. B. (1976). Two-tone suppression in auditory-nerve fibers: Extension of a stimulus-response relationship. *The Journal of the Acoustical Society of America*, *59*(1), 112–122. https://doi.org/10.1121/1.380841

Abdelfattah, A. S., Kawashima, T., Singh, A., Novak, O., Liu, H., Shuai, Y., Huang, Y.-C., Campagnola, L., Seeman, S. C., Yu, J., Zheng, J., Grimm, J. B., Patel, R., Friedrich, J., Mensh, B. D., Paninski, L., Macklin, J. J., Murphy, G. J., Podgorski, K., … Schreiter, E. R. (2019). Bright and photostable chemigenetic indicators for extended in vivo voltage imaging. *Science*, *365*(6454), 699–704. https://doi.org/10.1126/science.aav6416

Aertsen, A. M. H. J., & Johannesma, P. I. M. (1981). A comparison of the Spectro-Temporal sensitivity of auditory neurons to tonal and natural stimuli. *Biological Cybernetics*, *42*(2), 145–156. https://doi.org/10.1007/BF00336732

Aertsen, A. M. H. J., Olders, J. H. J., & Johannesma, P. I. M. (1981). Spectro-temporal receptive fields of auditory neurons in the grassfrog: III. analysis of the stimulus-event relation for natural stimuli. *Biological Cybernetics*, *39*(3), 195–209. https://doi.org/10.1007/BF00342772

Ahrens, M. B., Linden, J. F., & Sahani, M. (2008). Nonlinearities and Contextual Influences in Auditory Cortical Responses Modeled with Multilinear Spec-

trotemporal Methods. *The Journal of Neuroscience*, *28*(8), 1929–1942. https://doi.org/10.1523/JNEUROSCI.3377-07.2008

Ahrens, M. B., Paninski, L., & Sahani, M. (2008). Inferring input nonlinearities in neural encoding models. *Network: Computation in Neural Systems*, *19*(1), 35–67. https://doi.org/10.1080/09548980701813936

Allen, C. B., Celikel, T., & Feldman, D. E. (2003). Long-term depression induced by sensory deprivation during cortical map plasticity in vivo [Publisher: Springer Science and Business Media LLC]. *Nature Neuroscience*, *6*(3), 291–299. https://doi.org/10.1038/nn1012

Ashley, C. C., & Ridgway, E. B. (1968). Simultaneous Recording of Membrane Potential, Calcium Transient and Tension in Single Muscle Fibres [Publisher: Springer Science and Business Media LLC]. *Nature*, *219*(5159), 1168–1169. https://doi.org/10.1038/2191168a0

Atencio, C., & Schreiner, C. (2013, November). Stimulus choices for spike-triggered receptive field analysis.

Atencio, C. A., Sharpee, T. O., & Schreiner, C. E. (2008). Cooperative Nonlinearities in Auditory Cortical Neurons. *Neuron*, *58*(6), 956–966. https://doi.org/10.1016/j.neuron.2008.04.026

Atencio, C. A., Sharpee, T. O., & Schreiner, C. E. (2009). Hierarchical computation in the canonical auditory cortical circuit [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, *106*(51), 21894–21899. https://doi.org/10.1073/pnas.0908383106

Atencio, C. A., Sharpee, T. O., & Schreiner, C. E. (2012). Receptive field dimensionality increases from the auditory midbrain to cortex. *Journal of Neurophysiology*, *107*(10), 2594–2603. https://doi.org/10.1152/jn.01025.2011

Bando, Y., Grimm, C., Cornejo, V. H., & Yuste, R. (2019). Genetic voltage indicators. *BMC Biology*, *17*(1), 71. https://doi.org/10.1186/s12915-019-0682-0

BIBLIOGRAPHY

Bandyopadhyay, S., Shamma, S. A., & Kanold, P. O. (2010). Dichotomy of functional organization in the mouse auditory cortex. *Nature Neuroscience*, *13*(3), 361–368. https://doi.org/10.1038/nn.2490

Bathellier, B., Ushakova, L., & Rumpel, S. (2012). Discrete Neocortical Dynamics Predict Behavioral Categorization of Sounds [Publisher: Elsevier BV]. *Neuron*, *76*(2), 435–449. https://doi.org/10.1016/j.neuron.2012.07.008

Bednar, J. A., & Wilson, S. P. (2016). Cortical Maps [Publisher: SAGE Publications]. *The Neuroscientist*, *22*(6), 604–617. https://doi.org/10.1177/1073858415597645

Birkner, A., Tischbirek, C. H., & Konnerth, A. (2017). Improved deep two-photon calcium imaging in vivo [Publisher: Elsevier BV]. *Cell Calcium*, *64*, 29–35. https://doi.org/10.1016/j.ceca.2016.12.005

Bizley, J. K., & Cohen, Y. E. (2013). The what, where and how of auditory-object perception. *Nature Reviews Neuroscience*, *14*(10), 693–707. https://doi.org/10.1038/nrn3565

Blake, D. T., & Merzenich, M. M. (2002). Changes of AI Receptive Fields With Sound Density. *Journal of Neurophysiology*, *88*(6), 3409–3420. https://doi.org/10.1152/jn.00233.2002

Boone, W., & Piccinini, G. (2016). The cognitive neuroscience revolution. *Synthese*, *193*(5), 1509–1534. https://doi.org/10.1007/s11229-015-0783-4

Bowen, Z., Winkowski, D. E., & Kanold, P. O. (2020). Functional organization of mouse primary auditory cortex in adult C57BL/6 and F1 (CBAxC57) mice. *Scientific Reports*, *10*(1), 10905. https://doi.org/10.1038/s41598-020-67819-4

Brosch, M., & Schreiner, C. E. (1997). Time Course of Forward Masking Tuning Curves in Cat Primary Auditory Cortex [Publisher: American Physiological Society]. *Journal of Neurophysiology*, *77*(2), 923–943. https://doi.org/10.1152/jn.1997.77.2.923

Broussard, G. J., Diana, G., Urra Quiroz, F. J., Sermet, B. S., Rebola, N., Lynch, L. A., DiGregorio, D. A., & Wang, S. S.-H. (2025, January). Precise calcium-to-spike inference using biophysical generative models. https://doi.org/10.1101/2024.12.31.630967

Calabrese, A., Schumacher, J. W., Schneider, D. M., Paninski, L., & Woolley, S. M. N. (2011). A Generalized Linear Model for Estimating Spectrotemporal Receptive Fields from Responses to Natural Sounds (M. F. Kubke, Ed.). *PLoS ONE*, *6*(1), e16104. https://doi.org/10.1371/journal.pone.0016104

Carandini, M. (2004). Amplification of Trial-to-Trial Response Variability by Neurons in Visual Cortex (Charles Stevens, Ed.). *PLoS Biology*, *2*(9), e264. https://doi.org/10.1371/journal.pbio.0020264

Carruthers, I. M., Laplagne, D. A., Jaegle, A., Briguglio, J. J., Mwilambwe-Tshilobo, L., Natan, R. G., & Geffen, M. N. (2015). Emergence of invariant representation of vocalizations in the auditory cortex [Publisher: American Physiological Society]. *Journal of Neurophysiology*, *114*(5), 2726–2740. https://doi.org/10.1152/jn.00095.2015

Chambers, A. R., Hancock, K. E., Sen, K., & Polley, D. B. (2014). Online Stimulus Optimization Rapidly Reveals Multidimensional Selectivity in Auditory Cortical Neurons [Publisher: Society for Neuroscience]. *Journal of Neuroscience*, *34*(27), 8963–8975. https://doi.org/10.1523/jneurosci.0260-14.2014

Chambers, J. D., Elgueda, D., Fritz, J. B., Shamma, S. A., Burkitt, A. N., & Grayden, D. B. (2019). Computational Neural Modeling of Auditory Cortical Receptive Fields [Publisher: Frontiers Media SA]. *Frontiers in Computational Neuroscience*, *13*. https://doi.org/10.3389/fncom.2019.00028

Chang, M., & Kawai, H. D. (2018). A characterization of laminar architecture in mouse primary auditory cortex [Publisher: Springer Science and Business Media LLC].

Brain Structure and Function, *223*(9), 4187–4209. https://doi.org/10.1007/s00429-018-1744-8

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., & Kim, D. S. (2013). Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature*, *499*(7458), 295–300. https://doi.org/10.1038/nature12354

Chi, T., Ru, P., & Shamma, S. A. (2005). Multiresolution spectrotemporal analysis of complex sounds. *The Journal of the Acoustical Society of America*, *118*(2), 887–906. https://doi.org/10.1121/1.1945807

Chichilnisky, E. J. (2001). A simple white noise analysis of neuronal light responses. *Network: Computation in Neural Systems*, *12*(2), 199–213. https://doi.org/10.1080/713663221

Christianson, G. B., Sahani, M., & Linden, J. F. (2008). The Consequences of Response Nonlinearities for Interpretation of Spectrotemporal Receptive Fields. *The Journal of Neuroscience*, *28*(2), 446–455. https://doi.org/10.1523/JNEUROSCI.1775-07.2007

Coleman, J. R., & Clerici, W. J. (1987). Sources of projections to subdivisions of the inferior colliculus in the rat. *Journal of Comparative Neurology*, *262*(2), 215–226. https://doi.org/10.1002/cne.902620204

Cook, J. R., Li, H., Nguyen, B., Huang, H.-H., Mahdavian, P., Kirchgessner, M. A., Strassmann, P., Engelhardt, M., Callaway, E. M., & Jin, X. (2022). Secondary auditory cortex mediates a sensorimotor mechanism for action timing [Publisher: Springer Science and Business Media LLC]. *Nature Neuroscience*, *25*(3), 330–344. https://doi.org/10.1038/s41593-022-01025-5

Cooke, J. E., King, A. J., Willmore, B. D. B., & Schnupp, J. W. H. (2018). Contrast gain control in mouse auditory cortex. *Journal of Neurophysiology*, *120*(4), 1872–1884. https://doi.org/10.1152/jn.00847.2017

Curto, C., Sakata, S., Marguet, S., Itskov, V., & Harris, K. D. (2009). A Simple Model of Cortical Dynamics Explains Variability and State Dependence of Sensory Responses in Urethane-Anesthetized Auditory Cortex [Publisher: Society for Neuroscience]. *The Journal of Neuroscience*, *29*(34), 10600–10612. https://doi.org/10.1523/jneurosci.2053-09.2009

David, S. V. (2018). Incorporating behavioral and sensory context into spectro-temporal models of auditory encoding. *Hearing Research*, *360*, 107–123. https://doi.org/10.1016/j.heares.2017.12.021

David, S. V., Mesgarani, N., Fritz, J. B., & Shamma, S. A. (2009). Rapid Synaptic Depression Explains Nonlinear Modulation of Spectro-Temporal Tuning in Primary Auditory Cortex by Natural Stimuli. *The Journal of Neuroscience*, *29*(11), 3374–3386. https://doi.org/10.1523/JNEUROSCI.5249-08.2009

David, S. V., Mesgarani, N., & Shamma, S. A. (2007). Estimating sparse spectro-temporal receptive fields with natural stimuli. *Network: Computation in Neural Systems*, *18*(3), 191–212. https://doi.org/10.1080/09548980701609235

David, S. V., & Shamma, S. A. (2013). Integration over Multiple Timescales in Primary Auditory Cortex [Publisher: Society for Neuroscience]. *The Journal of Neuroscience*, *33*(49), 19154–19166. https://doi.org/10.1523/jneurosci.2270-13.2013

Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. MIT Press.

De Boer. (1968). Reverse Correlation I. - A heuristic introduction to the technique of triggered correlation with application to the analysis of compound systems. *Proc. Kon. Ned. Acad. Wetensch*, *71*, 472–486. https://cir.nii.ac.jp/crid/1570572699754725632

BIBLIOGRAPHY

De Boer, E., & Kuyper, P. (1968). Triggered Correlation. *IEEE Transactions on Biomedical Engineering*, *BME-15*(3), 169–179. https://doi.org/10.1109/TBME.1968.4502561

Dean, I., Harper, N. S., & McAlpine, D. (2005). Neural population coding of sound level adapts to stimulus statistics. *Nature Neuroscience*, *8*(12), 1684–1689. https://doi.org/10.1038/nn1541

deCharms, R. C., Blake, D. T., & Merzenich, M. M. (1998). Optimizing Sound Features for Cortical Neurons. *Science*, *280*(5368), 1439–1444. https://doi.org/10.1126/science.280.5368.1439

Deneux, T., Kaszas, A., Szalay, G., Katona, G., Lakner, T., Grinvald, A., Rózsa, B., & Vanzetta, I. (2016). Accurate spike estimation from noisy calcium signals for ultrafast three-dimensional imaging of large neuronal populations in vivo. *Nature Communications*, *7*(1), 12190. https://doi.org/10.1038/ncomms12190

Depireux, D. A., Simon, J. Z., Klein, D. J., & Shamma, S. A. (2001). Spectro-Temporal Response Field Characterization With Dynamic Ripples in Ferret Primary Auditory Cortex. *Journal of Neurophysiology*, *85*(3), 1220–1234. https://doi.org/10.1152/jn.2001.85.3.1220

DeWeese, M. R., Wehr, M., & Zador, A. M. (2003). Binary Spiking in Auditory Cortex. *The Journal of Neuroscience*, *23*(21), 7940–7949. https://doi.org/10.1523/JNEUROSCI.23-21-07940.2003

Deweese, M. R., & Zador, A. M. (2004). Shared and Private Variability in the Auditory Cortex [Publisher: American Physiological Society]. *Journal of Neurophysiology*, *92*(3), 1840–1855. https://doi.org/10.1152/jn.00197.2004

Edin, F., Machens, C. K., Schütze, H., & Herz, A. V. (2004). Searching for Optimal Sensory Signals: Iterative Stimulus Reconstruction in Closed-Loop Experiments [Publisher: Springer Science and Business Media LLC]. *Journal of Computa-*

tional Neuroscience, *17*(1), 47–56. https://doi.org/10.1023/b:jcns.0000023868.18446.a2

Eggermont, J. J. (1993). Wiener and Volterra analyses applied to the auditory system. *Hearing Research*, *66*(2), 177–201. https://doi.org/10.1016/0378-5955(93)90139-R

Eggermont, J. J. (1994). Temporal modulation transfer functions for AM and FM stimuli in cat auditory cortex. Effects of carrier type, modulating waveform and intensity [Publisher: Elsevier BV]. *Hearing Research*, *74*(1-2), 51–66. https://doi.org/10.1016/0378-5955(94)90175-9

Eggermont, J. J., Munguia, R., Pienkowski, M., & Shaw, G. (2011). Comparison of LFP-Based and Spike-Based Spectro-Temporal Receptive Fields and Cross-Correlation in Cat Primary Auditory Cortex (M. Coleman, Ed.). *PLoS ONE*, *6*(5), e20046. https://doi.org/10.1371/journal.pone.0020046

Elgueda, D., & Delano, P. H. (2020). Corticofugal modulation of audition. *Current Opinion in Physiology*, *18*, 73–78. https://doi.org/10.1016/j.cophys.2020.08.016

Elhilali, M., Fritz, J. B., Klein, D. J., Simon, J. Z., & Shamma, S. A. (2004). Dynamics of Precise Spike Timing in Primary Auditory Cortex. *The Journal of Neuroscience*, *24*(5), 1159–1172. https://doi.org/10.1523/JNEUROSCI.3825-03.2004

Éltes, T., Szoboszlay, M., Kerti-Szigeti, K., & Nusser, Z. (2019). Improved spike inference accuracy by estimating the peak amplitude of unitary [Ca^{2+}] transients in weakly GCaMP6f-expressing hippocampal pyramidal cells [Publisher: Wiley]. *The Journal of Physiology*, *597*(11), 2925–2947. https://doi.org/10.1113/jp277681

BIBLIOGRAPHY

Escabí, M. A., & Schreiner, C. E. (2002). Nonlinear Spectrotemporal Sound Analysis by Neurons in the Auditory Midbrain. *The Journal of Neuroscience*, *22*(10), 4114–4131. https://doi.org/10.1523/JNEUROSCI.22-10-04114.2002

Francis, N. A., Elgueda, D., Englitz, B., Fritz, J. B., & Shamma, S. A. (2018). Laminar profile of task-related plasticity in ferret primary auditory cortex. *Scientific Reports*, *8*(1), 16375. https://doi.org/10.1038/s41598-018-34739-3

Friedrich, J., Zhou, P., & Paninski, L. (2017). Fast online deconvolution of calcium imaging data (J. Vogelstein, Ed.). *PLOS Computational Biology*, *13*(3), e1005423. https://doi.org/10.1371/journal.pcbi.1005423

Fritz, J. B., David, S., & Shamma, S. (2013). *Attention and Dynamic, Task-Related Receptive Field Plasticity in Adult Auditory Cortex* [ISSN: 0947-2657]. Springer New York. https://doi.org/10.1007/978-1-4614-2350-8_9

Fritz, J., Elhilali, M., & Shamma, S. (2005). Active listening: Task-dependent plasticity of spectrotemporal receptive fields in primary auditory cortex. *Hearing Research*, *206*(1-2), 159–176. https://doi.org/10.1016/j.heares.2005.01.015

Fritz, J., Shamma, S., Elhilali, M., & Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nature Neuroscience*, *6*(11), 1216–1223. https://doi.org/10.1038/nn1141

Fritz, J. B., David, S. V., Radtke-Schuller, S., Yin, P., & Shamma, S. A. (2010). Adaptive, behaviorally gated, persistent encoding of task-relevant auditory information in ferret frontal cortex. *Nature Neuroscience*, *13*(8), 1011–1019. https://doi.org/10.1038/nn.2598

Fritz, J. B., Elhilali, M., & Shamma, S. A. (2005). Differential Dynamic Plasticity of A1 Receptive Fields during Multiple Spectral Tasks. *The Journal of Neuroscience*, *25*(33), 7623–7635. https://doi.org/10.1523/JNEUROSCI.1318-05.2005

Galvez, D., Diamos, G., Ciro, J., Cerón, J. F., Achorn, K., Gopi, A., Kanter, D., Lam, M., Mazumder, M., & Reddi, V. J. (2021). The People's Speech: A Large-Scale

Diverse English Speech Recognition Dataset for Commercial Usage [arXiv: 2111.09344]. *CoRR*, *abs/2111.09344*. https://arxiv.org/abs/2111.09344

Gaucher, Q., Panniello, M., Ivanov, A. Z., Dahmen, J. C., King, A. J., & Walker, K. M. (2020). Complexity of frequency receptive fields predicts tonotopic variability across species. *eLife*, *9*, e53462. https://doi.org/10.7554/eLife.53462

Gerstner, W., Kistler, W. M., Naud, R., & Paninski, L. (2014). *Neuronal dynamics: From single neurons to networks and models of cognition*. Cambridge University Press.

Gerstner, W., Kreiter, A. K., Markram, H., & Herz, A. V. M. (1997). Neural codes: Firing rates and beyond. *Proceedings of the National Academy of Sciences*, *94*(24), 12740–12741. https://doi.org/10.1073/pnas.94.24.12740

Grienberger, C., & Konnerth, A. (2012). Imaging Calcium in Neurons. *Neuron*, *73*(5), 862–885. https://doi.org/10.1016/j.neuron.2012.02.011

Griffiths, T. D., & Warren, J. D. (2004). What is an auditory object? *Nature Reviews Neuroscience*, *5*(11), 887–892. https://doi.org/10.1038/nrn1538

Guo, W., Chambers, A. R., Darrow, K. N., Hancock, K. E., Shinn-Cunningham, B. G., & Polley, D. B. (2012). Robustness of Cortical Topography across Fields, Laminae, Anesthetic States, and Neurophysiological Signal Types [Publisher: Society for Neuroscience]. *Journal of Neuroscience*, *32*(27), 9159–9172. https://doi.org/10.1523/jneurosci.0065-12.2012

Hackett, T. A. (2011). Information flow in the auditory cortical network [Publisher: Elsevier BV]. *Hearing Research*, *271*(1-2), 133–146. https://doi.org/10.1016/j.heares.2010.01.011

Harper, N. S., Schoppe, O., Willmore, B. D. B., Cui, Z., Schnupp, J. W. H., & King, A. J. (2016). Network Receptive Field Modeling Reveals Extensive Integration and Multi-feature Selectivity in Auditory Cortical Neurons (I. Nelken, Ed.)

[Publisher: Public Library of Science (PLoS)]. *PLOS Computational Biology*, *12*(11), e1005113. https://doi.org/10.1371/journal.pcbi.1005113

Heffner, R., & Masterton, R. (1990). Comparative perception [Publisher: Wiley].

Heggelund, P., & Albus, K. (1978). Response variability and orientation discrimination of single cells in striate cortex of cat [Publisher: Springer Science and Business Media LLC]. *Experimental Brain Research*, *32*(2). https://doi.org/10.1007/bf00239727

Helmchen, F., Imoto, K., & Sakmann, B. (1996). Ca2+ buffering and action potential-evoked Ca2+ signaling in dendrites of pyramidal neurons [Publisher: Elsevier BV]. *Biophysical Journal*, *70*(2), 1069–1081. https://doi.org/10.1016/s0006-3495(96)79653-4

Helmchen, F., & Denk, W. (2005). Deep tissue two-photon microscopy [Publisher: Springer Science and Business Media LLC]. *Nature Methods*, *2*(12), 932–940. https://doi.org/10.1038/nmeth818

Holdgraf, C. R., De Heer, W., Pasley, B., Rieger, J., Crone, N., Lin, J. J., Knight, R. T., & Theunissen, F. E. (2016). Rapid tuning shifts in human auditory cortex enhance speech intelligibility. *Nature Communications*, *7*(1), 13654. https://doi.org/10.1038/ncomms13654

Homma, N. Y., Atencio, C. A., & Schreiner, C. E. (2021). Plasticity of Multidimensional Receptive Fields in Core Rat Auditory Cortex Directed by Sound Statistics. *Neuroscience*, *467*, 150–170. https://doi.org/10.1016/j.neuroscience.2021.04.028

Homma, N. Y., See, J. Z., Atencio, C. A., Hu, C., Downer, J. D., Beitel, R. E., Cheung, S. W., Najafabadi, M. S., Olsen, T., Bigelow, J., Hasenstaub, A. R., Malone, B. J., & Schreiner, C. E. (2024). Receptive-field nonlinearities in primary auditory cortex: A comparative perspective. *Cerebral Cortex*, *34*(9), bhae364. https://doi.org/10.1093/cercor/bhae364

Hromádka, T., DeWeese, M. R., & Zador, A. M. (2008). Sparse Representation of Sounds in the Unanesthetized Auditory Cortex (D. Plenz, Ed.). *PLoS Biology*, *6*(1), e16. https://doi.org/10.1371/journal.pbio.0060016

Hsu, A., Borst, A., & Theunissen, F. E. (2004). Quantifying variability in neural responses and its application for the validation of model predictions. *Network: Computation in Neural Systems*, *15*(2), 91. https://doi.org/10.1088/0954-898X/15/2/002

Hu, C., Hasenstaub, A. R., & Schreiner, C. E. (2024). Basic Properties of Coordinated Neuronal Ensembles in the Auditory Thalamus. *The Journal of Neuroscience*, *44*(19), e1729232024. https://doi.org/10.1523/JNEUROSCI.1729-23.2024

Huang, L., Ledochowitsch, P., Knoblich, U., Lecoq, J., Murphy, G. J., Reid, R. C., De Vries, S. E., Koch, C., Zeng, H., Buice, M. A., Waters, J., & Li, L. (2021). Relationship between simultaneously recorded spiking activity and fluorescence signal in GCaMP6 transgenic mice. *eLife*, *10*, e51675. https://doi.org/10.7554/eLife.51675

Hullett, P. W., Hamilton, L. S., Mesgarani, N., Schreiner, C. E., & Chang, E. F. (2016). Human Superior Temporal Gyrus Organization of Spectrotemporal Modulation Tuning Derived from Speech Stimuli. *The Journal of Neuroscience*, *36*(6), 2014–2026. https://doi.org/10.1523/JNEUROSCI.1779-15.2016

Illing, R.-B., Kraus, K. S., & Michler, S. A. (2000). Plasticity of the superior olivary complex. *Microscopy Research and Technique*, *51*(4), 364–381. https://doi.org/10.1002/1097-0029(20001115)51:4⟨364::AID-JEMT6⟩3.0.CO;2-E

Issa, J. B., Haeffele, B. D., Agarwal, A., Bergles, D. E., Young, E. D., & Yue, D. T. (2014). Multiscale Optical Ca2+ Imaging of Tonal Organization in Mouse Auditory Cortex. *Neuron*, *83*(4), 944–959. https://doi.org/10.1016/j.neuron.2014.07.009

BIBLIOGRAPHY

Ito, T., Bishop, D. C., & Oliver, D. L. (2016). Functional organization of the local circuit in the inferior colliculus. *Anatomical Science International*, *91*(1), 22–34. https://doi.org/10.1007/s12565-015-0308-8

Joachimsthaler, B., Uhlmann, M., Miller, F., Ehret, G., & Kurt, S. (2014). Quantitative analysis of neuronal response properties in primary and higher-order auditory cortical fields of awake house mice [Publisher: Wiley]. *European Journal of Neuroscience*, *39*(6), 904–918. https://doi.org/10.1111/ejn.12478

Johnson, D. H. (1980). The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. *The Journal of the Acoustical Society of America*, *68*(4), 1115–1122. https://doi.org/10.1121/1.384982

Jones, J. P., & Palmer, L. A. (1987). The two-dimensional spatial structure of simple receptive fields in cat striate cortex [Publisher: American Physiological Society]. *Journal of Neurophysiology*, *58*(6), 1187–1211. https://doi.org/10.1152/jn.1987.58.6.1187

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, Ç., Barbic, M., Blanche, T. J., Bonin, V., Couto, J., Dutta, B., Gratiy, S. L., Gutnisky, D. A., Häusser, M., Karsh, B., … Harris, T. D. (2017). Fully integrated silicon probes for high-density recording of neural activity. *Nature*, *551*(7679), 232–236. https://doi.org/10.1038/nature24636

Kang, H., & Kanold, P. O. (2024). Sparse representation of neurons for encoding complex sounds in the auditory cortex [Publisher: Elsevier BV]. *Progress in Neurobiology*, *241*, 102661. https://doi.org/10.1016/j.pneurobio.2024.102661

Kelly, J. B., Judge, P. W., & Phillips, D. P. (1986). Representation of the cochlea in primary auditory cortex of the ferret (Mustela putorius) [Publisher: Elsevier BV]. *Hearing Research*, *24*(2), 111–115. https://doi.org/10.1016/0378-5955(86)90054-7

Keshishian, M., Akbari, H., Khalighinejad, B., Herrero, J. L., Mehta, A. D., & Mesgarani, N. (2020). Estimating and interpreting nonlinear receptive field of sensory neural responses with deep neural network models [Publisher: eLife Sciences Publications, Ltd]. *eLife*, *9*. https://doi.org/10.7554/elife.53445

Kisley, M. A., & Gerstein, G. L. (1999). Trial-to-Trial Variability and State-Dependent Modulation of Auditory-Evoked Responses in Cortex [Publisher: Society for Neuroscience]. *The Journal of Neuroscience*, *19*(23), 10451–10460. https://doi.org/10.1523/jneurosci.19-23-10451.1999

Klein, D., Depireux, D., Simon, J., & Shamma, S. (2000). Robust Spectrotemporal Reverse Correlation for the Auditory System: Optimizing Stimulus Design. *Journal of Computational Neuroscience*, *9*(1), 85–111. https://doi.org/10.1023/A:1008990412183

Kline, A. M., Aponte, D. A., & Kato, H. K. (2023). Distinct nonlinear spectrotemporal integration in primary and secondary auditory cortices. *Scientific Reports*, *13*(1), 7658. https://doi.org/10.1038/s41598-023-34731-6

Kowalski, N., Depireux, D. A., & Shamma, S. A. (1996a). Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra. *Journal of Neurophysiology*, *76*(5), 3503–3523. https://doi.org/10.1152/jn.1996.76.5.3503

Kowalski, N., Depireux, D. A., & Shamma, S. A. (1996b). Analysis of dynamic spectra in ferret primary auditory cortex. II. Prediction of unit responses to arbitrary dynamic spectra. *Journal of Neurophysiology*, *76*(5), 3524–3534. https://doi.org/10.1152/jn.1996.76.5.3524

Kuchibhotla, K., & Bathellier, B. (2018). Neural encoding of sensory and behavioral complexity in the auditory cortex. *Current Opinion in Neurobiology*, *52*, 65–71. https://doi.org/10.1016/j.conb.2018.04.002

BIBLIOGRAPHY

Kulesza, R. J. (2007). Cytoarchitecture of the human superior olivary complex: Medial and lateral superior olive [Publisher: Elsevier BV]. *Hearing Research*, *225*(1-2), 80–90. https://doi.org/10.1016/j.heares.2006.12.006

Laudanski, J., Edeline, J.-M., & Huetz, C. (2012). Differences between Spectro-Temporal Receptive Fields Derived from Artificial and Natural Stimuli in the Auditory Cortex (M. S. Malmierca, Ed.). *PLoS ONE*, *7*(11), e50539. https://doi.org/10.1371/journal.pone.0050539

Lee, T.-Y., Weissenberger, Y., King, A. J., & Dahmen, J. C. (2023, August). Midbrain encodes sound detection behavior without auditory cortex. https://doi.org/10.7554/elife.89950.1

Leong, J. C. S., Esch, J. J., Poole, B., Ganguli, S., & Clandinin, T. R. (2016). Direction Selectivity in Drosophila Emerges from Preferred-Direction Enhancement and Null-Direction Suppression. *Journal of Neuroscience*, *36*(31), 8078–8092. https://doi.org/10.1523/JNEUROSCI.1272-16.2016

Li, R., Wang, M., Yao, J., Liang, S., Liao, X., Yang, M., Zhang, J., Yan, J., Jia, H., Chen, X., & Li, X. (2018). Two-Photon Functional Imaging of the Auditory Cortex in Behaving Mice: From Neural Networks to Single Spines [Publisher: Frontiers Media SA]. *Frontiers in Neural Circuits*, *12*. https://doi.org/10.3389/fncir.2018.00033

Liang, F., Li, H., Chou, X.-l., Zhou, M., Zhang, N. K., Xiao, Z., Zhang, K. K., Tao, H. W., & Zhang, L. I. (2019). Sparse Representation in Awake Auditory Cortex: Cell-type Dependence, Synaptic Mechanisms, Developmental Emergence, and Modulation [Publisher: Oxford University Press (OUP)]. *Cerebral Cortex*, *29*(9), 3796–3812. https://doi.org/10.1093/cercor/bhy260

Linden, J. F., Liu, R. C., Sahani, M., Schreiner, C. E., & 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. https://doi.org/10.1152/jn.00751.2002

Lohse, M., King, A. J., & Willmore, B. D. (2024). Subcortical origin of nonlinear sound encoding in auditory cortex [Publisher: Elsevier BV]. *Current Biology*, *34*(15), 3405–3415.e5. https://doi.org/10.1016/j.cub.2024.06.057

Machens, C. K., Wehr, M. S., & Zador, A. M. (2004). Linearity of Cortical Receptive Fields Measured with Natural Sounds. *The Journal of Neuroscience*, *24*(5), 1089–1100. https://doi.org/10.1523/JNEUROSCI.4445-03.2004

Malik, W. Q., Schummers, J., Sur, M., & Brown, E. N. (2011). Denoising Two-Photon Calcium Imaging Data (M. Perc, Ed.) [Publisher: Public Library of Science (PLoS)]. *PLoS ONE*, *6*(6), e20490. https://doi.org/10.1371/journal.pone.0020490

Maravall, M., Mainen, Z., Sabatini, B., & Svoboda, K. (2000). Estimating Intracellular Calcium Concentrations and Buffering without Wavelength Ratioing [Publisher: Elsevier BV]. *Biophysical Journal*, *78*(5), 2655–2667. https://doi.org/10.1016/s0006-3495(00)76809-3

Marmarelis, P. Z., & Marmarelis, V. Z. (1978). Applicability of the White-Noise Method and the Use of Quasiwhite Test Signals. In *Analysis of Physiological Systems* (pp. 181–221). Springer US. https://doi.org/10.1007/978-1-4613-3970-0_5

Marmarelis, V. Z. (2004, August). *Nonlinear Dynamic Modeling of Physiological Systems* (1st ed.). Wiley. https://doi.org/10.1002/9780471679370

Massoudi, R., Van Wanrooij, M. M., Versnel, H., & Van Opstal, A. J. (2015). Spectrotemporal Response Properties of Core Auditory Cortex Neurons in Awake Monkey (M. S. Malmierca, Ed.) [Publisher: Public Library of Science (PLoS)]. *PLOS ONE*, *10*(2), e0116118. https://doi.org/10.1371/journal.pone.0116118

BIBLIOGRAPHY

Merchán, M. A., Juiz, J. M., Godfrey, D. A., Mugnaini, E., Smith, P. H., Joris, P. X., Banks, M. I., & Yin, T. C. T. (1993). Responses of cochlear nucleus cells and projections of their axons. In M. A. Merchán, J. M. Juiz, D. A. Godfrey, & E. Mugnaini (Eds.), *The Mammalian Cochlear Nuclei: Organization and Function* (pp. 349–360). Plenum.

Merzenich, M. M., Knight, P. L., & Roth, G. L. (1975). Representation of cochlea within primary auditory cortex in the cat [Publisher: American Physiological Society]. *Journal of Neurophysiology*, *38*(2), 231–249. https://doi.org/10.1152/jn.1975.38.2.231

Merzenich, M. M., & Brugge, J. F. (1973). Representation of the cochlear partition on the superior temporal plane of the macaque monkey [Publisher: Elsevier BV]. *Brain Research*, *50*(2), 275–296. https://doi.org/10.1016/0006-8993(73)90731-2

Meyer, A. F., Williamson, R. S., Linden, J. F., & Sahani, M. (2017). Models of Neuronal Stimulus-Response Functions: Elaboration, Estimation, and Evaluation. *Frontiers in Systems Neuroscience*, *10*. https://doi.org/10.3389/fnsys.2016.00109

Miller, L. M., Escabí, M. A., Read, H. L., & Schreiner, C. E. (2001). Functional Convergence of Response Properties in the Auditory Thalamocortical System [Publisher: Elsevier BV]. *Neuron*, *32*(1), 151–160. https://doi.org/10.1016/s0896-6273(01)00445-7

Miller, L. M., Escabí, M. A., Read, H. L., & Schreiner, C. E. (2002). Spectrotemporal Receptive Fields in the Lemniscal Auditory Thalamus and Cortex. *Journal of Neurophysiology*, *87*(1), 516–527. https://doi.org/10.1152/jn.00395.2001

Miyawaki, A., Llopis, J., Heim, R., McCaffery, J. M., Adams, J. A., Ikura, M., & Tsien, R. Y. (1997). Fluorescent indicators for Ca2+based on green fluorescent proteins and calmodulin [Publisher: Springer Science and Business Media LLC]. *Nature*, *388*(6645), 882–887. https://doi.org/10.1038/42264

Morandell, K., Yin, A., Del Rio, R. T., & Schneider, D. M. (2023, July). Movement-related modulation in mouse auditory cortex is widespread yet locally diverse. https://doi.org/10.1101/2023.07.03.547560

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

Mrsic-Flogel, T. D., King, A. J., & Schnupp, J. W. H. (2005). Encoding of Virtual Acoustic Space Stimuli by Neurons in Ferret Primary Auditory Cortex. *Journal of Neurophysiology*, *93*(6), 3489–3503. https://doi.org/10.1152/jn.00748.2004

Nagel, K. I., & Doupe, A. J. (2008). Organizing Principles of Spectro-Temporal Encoding in the Avian Primary Auditory Area Field L. *Neuron*, *58*(6), 938–955. https://doi.org/10.1016/j.neuron.2008.04.028

Nishimura, M., Shirasawa, H., Kaizo, H., & Song, W.-J. (2007). New Field With Tonotopic Organization in Guinea Pig Auditory Cortex [Publisher: American Physiological Society]. *Journal of Neurophysiology*, *97*(1), 927–932. https://doi.org/10.1152/jn.00689.2006

Norman-Haignere, S. V., Long, L. K., Devinsky, O., Doyle, W., Irobunda, I., Merricks, E. M., Feldstein, N. A., McKhann, G. M., Schevon, C. A., Flinker, A., & Mesgarani, N. (2022). Multiscale temporal integration organizes hierarchical computation in human auditory cortex [Publisher: Springer Science and Business Media LLC]. *Nature Human Behaviour*, *6*(3), 455–469. https://doi.org/10.1038/s41562-021-01261-y

Pachitariu, M., Stringer, C., Dipoppa, M., Schröder, S., Rossi, L. F., Dalgleish, H., Carandini, M., & Harris, K. D. (2016, June). Suite2p: Beyond 10,000 neurons with standard two-photon microscopy. https://doi.org/10.1101/061507

Pachitariu, M., Stringer, C., & Harris, K. D. (2018a). Robustness of Spike Deconvolution for Neuronal Calcium Imaging. *The Journal of Neuroscience*, *38*(37), 7976–7985. https://doi.org/10.1523/JNEUROSCI.3339-17.2018

Pachitariu, M., Stringer, C., & Harris, K. D. (2018b). Robustness of Spike Deconvolution for Neuronal Calcium Imaging. *The Journal of Neuroscience*, *38*(37), 7976–7985. https://doi.org/10.1523/JNEUROSCI.3339-17.2018

Paninski, L. (2003). Convergence properties of three spike-triggered analysis techniques. *Network: Computation in Neural Systems*, *14*(3), 437–464. https://doi.org/10.1088/0954-898X_14_3_304

Paninski, L., Pillow, J., & Lewi, J. (2007). Statistical models for neural encoding, decoding, and optimal stimulus design (P. Cisek, T. Drew, & J. F. Kalaska, Eds.). *Progress in Brain Research*, *165*, 493–507. https://doi.org/10.1016/S0079-6123(06)65031-0

Panniello, M., King, A. J., Dahmen, J. C., & Walker, K. M. M. (2018). Local and Global Spatial Organization of Interaural Level Difference and Frequency Preferences in Auditory Cortex. *Cerebral Cortex*, *28*(1), 350–369. https://doi.org/10.1093/cercor/bhx295

Paredes, R. M., Etzler, J. C., Watts, L. T., Zheng, W., & Lechleiter, J. D. (2008). Chemical calcium indicators [Publisher: Elsevier BV]. *Methods*, *46*(3), 143–151. https://doi.org/10.1016/j.ymeth.2008.09.025

Park, M., & Pillow, J. W. (2011). Receptive Field Inference with Localized Priors (O. Sporns, Ed.) [Publisher: Public Library of Science (PLoS)]. *PLoS Computational Biology*, *7*(10), e1002219. https://doi.org/10.1371/journal.pcbi.1002219

Park, M., & Pillow, J. W. (2013). Bayesian inference for low rank spatiotemporal neural receptive fields. In C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Weinberger (Eds.), *Advances in neural information processing systems*

(Vol. 26). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2013/file/68a83eeb494a308fe5295da69428a507-Paper.pdf

Peterka, D. S., Takahashi, H., & Yuste, R. (2011). Imaging Voltage in Neurons [Publisher: Elsevier BV]. *Neuron*, *69*(1), 9–21. https://doi.org/10.1016/j.neuron.2010.12.010

Petersen, C. C., & Crochet, S. (2013). Synaptic Computation and Sensory Processing in Neocortical Layer 2/3. *Neuron*, *78*(1), 28–48. https://doi.org/10.1016/j.neuron.2013.03.020

Pickles, J. O. (2015). Auditory pathways. In *Handbook of Clinical Neurology* (pp. 3–25, Vol. 129). Elsevier. https://doi.org/10.1016/B978-0-444-62630-1.00001-9

Pinter, R. B., & Harris, L. R. (1981). Temporal and spatial response characteristics of the cat superior colliculus [Publisher: Elsevier BV]. *Brain Research*, *207*(1), 73–94. https://doi.org/10.1016/0006-8993(81)90680-6

Plakke, B., & Romanski, L. M. (2014). Auditory connections and functions of prefrontal cortex [Publisher: Frontiers Media SA]. *Frontiers in Neuroscience*, *8*. https://doi.org/10.3389/fnins.2014.00199

Pospischil, M., Piwkowska, Z., Bal, T., & Destexhe, A. (2011). Comparison of different neuron models to conductance-based post-stimulus time histograms obtained in cortical pyramidal cells using dynamic-clamp in vitro. *Biological Cybernetics*, *105*(2), 167–180. https://doi.org/10.1007/s00422-011-0458-2

Purves, D. (Ed.). (2001). *Neuroscience* (2nd ed). Sinauer Associates.
Includes bibliographical references and index.

Qian, J., Hastie, T., Friedman, J., Tibshirani, R., & Simon, N. (2013). Glmnet for Matlab. http://hastie.su.domains/glmnet_matlab/

Rabinowitz, N. C., Willmore, B. D. B., Schnupp, J. W. H., & King, A. J. (2012). Spectrotemporal Contrast Kernels for Neurons in Primary Auditory Cortex.

The Journal of Neuroscience, *32*(33), 11271–11284. https://doi.org/10.1523/JNEUROSCI.1715-12.2012

Rabinowitz, N. C., Willmore, B. D., Schnupp, J. W., & King, A. J. (2011). Contrast Gain Control in Auditory Cortex. *Neuron*, *70*(6), 1178–1191. https://doi.org/10.1016/j.neuron.2011.04.030

Rahman, M., Willmore, B. D. B., King, A. J., & Harper, N. S. (2019). A dynamic network model of temporal receptive fields in primary auditory cortex (F. E. Theunissen, Ed.). *PLOS Computational Biology*, *15*(5), e1006618. https://doi.org/10.1371/journal.pcbi.1006618

Rahman, M., Willmore, B. D. B., King, A. J., & Harper, N. S. (2020). Simple transformations capture auditory input to cortex. *Proceedings of the National Academy of Sciences*, *117*(45), 28442–28451. https://doi.org/10.1073/pnas.1922033117

Ramdya, P., Reiter, B., & Engert, F. (2006). Reverse correlation of rapid calcium signals in the zebrafish optic tectum in vivo. *Journal of Neuroscience Methods*, *157*(2), 230–237. https://doi.org/10.1016/j.jneumeth.2006.04.021

Ramirez, A., Pnevmatikakis, E. A., Merel, J., Paninski, L., Miller, K. D., & Bruno, R. M. (2014). Spatiotemporal receptive fields of barrel cortex revealed by reverse correlation of synaptic input. *Nature Neuroscience*, *17*(6), 866–875. https://doi.org/10.1038/nn.3720

Richardson, R. J., Blundon, J. A., Bayazitov, I. T., & Zakharenko, S. S. (2009). Connectivity Patterns Revealed by Mapping of Active Inputs on Dendrites of Thalamorecipient Neurons in the Auditory Cortex [Publisher: Society for Neuroscience]. *The Journal of Neuroscience*, *29*(20), 6406–6417. https://doi.org/10.1523/jneurosci.0258-09.2009

Rieke, F., Warland, D., Ruyter van Steveninck, R. R. d., & Bialek, W. S. (1999). *Spikes: Exploring the neural code* (First MIT Press paperback edition). The

MIT Press.

Includes index. - Bibliography. - Originally published: 1996.

Rothschild, G., Nelken, I., & Mizrahi, A. (2010). Functional organization and population dynamics in the mouse primary auditory cortex. *Nature Neuroscience*, *13*(3), 353–360. https://doi.org/10.1038/nn.2484

Rotman, Y., Bar-Yosef, O., & Nelken, I. (2001). Relating cluster and population responses to natural sounds and tonal stimuli in cat primary auditory cortex. *Hearing Research*, *152*(1-2), 110–127. https://doi.org/10.1016/S0378-5955(00)00243-4

Rudnicki, M., Schoppe, O., Isik, M., Völk, F., & Hemmert, W. (2015). Modeling auditory coding: From sound to spikes. *Cell and Tissue Research*, *361*(1), 159–175. https://doi.org/10.1007/s00441-015-2202-z

Runyan, C. A., Piasini, E., Panzeri, S., & Harvey, C. D. (2017). Distinct timescales of population coding across cortex. *Nature*, *548*(7665), 92–96. https://doi.org/10.1038/nature23020

Rupprecht, P., Rózsa, M., Fang, X., Svoboda, K., & Helmchen, F. (2025, March). Spike inference from calcium imaging data acquired with GCaMP8 indicators. https://doi.org/10.1101/2025.03.03.641129

Rusakov, D. A., Savtchenko, L. P., & Latham, P. E. (2020). Noisy Synaptic Conductance: Bug or a Feature? *Trends in Neurosciences*, *43*(6), 363–372. https://doi.org/10.1016/j.tins.2020.03.009

Rust, N. C., Schwartz, O., Movshon, J. A., & Simoncelli, E. P. (2005). Spatiotemporal Elements of Macaque V1 Receptive Fields. *Neuron*, *46*(6), 945–956. https://doi.org/10.1016/j.neuron.2005.05.021

Sabatini, B. L., Oertner, T. G., & Svoboda, K. (2002). The Life Cycle of Ca2+ Ions in Dendritic Spines [Publisher: Elsevier BV]. *Neuron*, *33*(3), 439–452. https://doi.org/10.1016/s0896-6273(02)00573-1

Sahani & Linden. (2003). How Linear are Auditory Cortical Responses? *Advances in neural information processing systems*, (15), 109–116.

Sahani, M., & Linden, J. (2002). Evidence optimization techniques for estimating stimulus-response functions. *Advances in neural information processing systems*, *15*.

Sally, S. L., & Kelly, J. B. (1988). Organization of auditory cortex in the albino rat: Sound frequency [Publisher: American Physiological Society]. *Journal of Neurophysiology*, *59*(5), 1627–1638. https://doi.org/10.1152/jn.1988.59.5.1627

Sawatari, H., Tanaka, Y., Takemoto, M., Nishimura, M., Hasegawa, K., Saitoh, K., & Song, W.-J. (2011). Identification and characterization of an insular auditory field in mice [Publisher: Wiley]. *European Journal of Neuroscience*, *34*(12), 1944–1952. https://doi.org/10.1111/j.1460-9568.2011.07926.x

Schnupp, J., Nelken, I., & King, A. J. (2010, November). *Auditory Neuroscience: Making Sense of Sound*. The MIT Press. https://doi.org/10.7551/mitpress/7942.001.0001

Schoppe, O., Harper, N. S., Willmore, B. D. B., King, A. J., & Schnupp, J. W. H. (2016). Measuring the Performance of Neural Models. *Frontiers in Computational Neuroscience*, *10*. https://doi.org/10.3389/fncom.2016.00010

Schreiner, C. E., & Urbas, J. V. (1988). Representation of amplitude modulation in the auditory cortex of the cat. II. Comparison between cortical fields [Publisher: Elsevier BV]. *Hearing Research*, *32*(1), 49–63. https://doi.org/10.1016/0378-5955(88)90146-3

See, J. Z., Atencio, C. A., Sohal, V. S., & Schreiner, C. E. (2018). Coordinated neuronal ensembles in primary auditory cortical columns. *eLife*, *7*, e35587. https://doi.org/10.7554/eLife.35587

See, J. Z., Homma, N. Y., Atencio, C. A., Sohal, V. S., & Schreiner, C. E. (2021). Information diversity in individual auditory cortical neurons is associated with

functionally distinct coordinated neuronal ensembles. *Scientific Reports*, *11*(1), 4064. https://doi.org/10.1038/s41598-021-83565-7

Sharpee, T., Rust, N. C., & Bialek, W. (2004). Analyzing Neural Responses to Natural Signals: Maximally Informative Dimensions. *Neural Computation*, *16*(2), 223–250. https://doi.org/10.1162/089976604322742010

Shimomura, O., Johnson, F. H., & Saiga, Y. (1962). Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan, *Aequorea* [Publisher: Wiley]. *Journal of Cellular and Comparative Physiology*, *59*(3), 223–239. https://doi.org/10.1002/jcp.1030590302

Simon, J. Z., Depireux, D. A., Klein, D. J., Fritz, J. B., & Shamma, S. A. (2007). Temporal Symmetry in Primary Auditory Cortex: Implications for Cortical Connectivity [Publisher: MIT Press - Journals]. *Neural Computation*, *19*(3), 583–638. https://doi.org/10.1162/neco.2007.19.3.583

Song, W.-J., Kawaguchi, H., Totoki, S., Inoue, Y., Katura, T., Maeda, S., Inagaki, S., Shirasawa, H., & Nishimura, M. (2006). Cortical Intrinsic Circuits Can Support Activity Propagation through an Isofrequency Strip of the Guinea Pig Primary Auditory Cortex [Publisher: Oxford University Press (OUP)]. *Cerebral Cortex*, *16*(5), 718–729. https://doi.org/10.1093/cercor/bhj018

Soto, G., Kopell, N., & Sen, K. (2006). Network Architecture, Receptive Fields, and Neuromodulation: Computational and Functional Implications of Cholinergic Modulation in Primary Auditory Cortex. *Journal of Neurophysiology*, *96*(6), 2972–2983. https://doi.org/10.1152/jn.00459.2006

Stein, R. B., Gossen, E. R., & Jones, K. E. (2005). Neuronal variability: Noise or part of the signal? *Nature Reviews Neuroscience*, *6*(5), 389–397. https://doi.org/10.1038/nrn1668

Sutter, M. L., & Schreiner, C. E. (1991). Physiology and topography of neurons with multipeaked tuning curves in cat primary auditory cortex [Publisher: American

Physiological Society]. *Journal of Neurophysiology*, *65*(5), 1207–1226. https://doi.org/10.1152/jn.1991.65.5.1207

Terreros, G., & Delano, P. H. (2015). Corticofugal modulation of peripheral auditory responses. *Frontiers in Systems Neuroscience*, *9*. https://doi.org/10.3389/fnsys.2015.00134

Theunissen, F. E., David, S. V., Singh, N. C., Hsu, A., Vinje, W. E., & Gallant, J. L. (2001). Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. *Network (Bristol, England)*, *12*(3), 289–316.

Theunissen, F. E., Sen, K., & Doupe, A. J. (2000). Spectral-Temporal Receptive Fields of Nonlinear Auditory Neurons Obtained Using Natural Sounds. *The Journal of Neuroscience*, *20*(6), 2315–2331. https://doi.org/10.1523/JNEUROSCI.20-06-02315.2000

Thorson, I. L., Liénard, J., & David, S. V. (2015). The Essential Complexity of Auditory Receptive Fields (F. E. Theunissen, Ed.). *PLOS Computational Biology*, *11*(12), e1004628. https://doi.org/10.1371/journal.pcbi.1004628

Trachtenberg, J. T., Trepel, C., & Stryker, M. P. (2000). Rapid Extragranular Plasticity in the Absence of Thalamocortical Plasticity in the Developing Primary Visual Cortex [Publisher: American Association for the Advancement of Science (AAAS)]. *Science*, *287*(5460), 2029–2032. https://doi.org/10.1126/science.287.5460.2029

Tremblay, K. L., & Burkhard, R. F. (2012). *Normal Aspects of Hearing*. Plural Publishing, Inc.

The book starts out with a chapter on acoustics, and the rest of the book focuses on the anatomy and physiology of the peripheral and central auditory systems in a rather traditional manner: from caudal through rostral levels, ending with the descending auditory system. Note that these chapters, for the most part,

review topic areas that are best considered basic research and are not translational in nature. However, the final section attempts to tie perception to the underlying physiologic responses, and chapters are parsed into stimulus factors (such as intensity, frequency, binaural stimula The first book in the three-book series, Translational Perspectives in Auditory Neuroscience.

Triplett, M. A., & Goodhill, G. J. (2019). Probabilistic Encoding Models for Multivariate Neural Data. *Frontiers in Neural Circuits*, *13*, 1. https://doi.org/10.3389/fncir.2019.00001

Tsien, R. Y. (1980). New calcium indicators and buffers with high selectivity against magnesium and protons: Design, synthesis, and properties of prototype structures [Publisher: American Chemical Society (ACS)]. *Biochemistry*, *19*(11), 2396–2404. https://doi.org/10.1021/bi00552a018

Ulanovsky, N., Las, L., Farkas, D., & Nelken, I. (2004). Multiple Time Scales of Adaptation in Auditory Cortex Neurons. *The Journal of Neuroscience*, *24*(46), 10440–10453. https://doi.org/10.1523/JNEUROSCI.1905-04.2004

Ulanovsky, N., Las, L., & Nelken, I. (2003). Processing of low-probability sounds by cortical neurons. *Nature Neuroscience*, *6*(4), 391–398. https://doi.org/10.1038/nn1032

Valentine, P. A., & Eggermont, J. J. (2004). Stimulus dependence of spectro-temporal receptive fields in cat primary auditory cortex [Publisher: Elsevier BV]. *Hearing Research*, *196*(1-2), 119–133. https://doi.org/10.1016/j.heares.2004.05.011

Vasquez-Lopez, S. A., Weissenberger, Y., Lohse, M., Keating, P., King, A. J., & Dahmen, J. C. (2017). Thalamic input to auditory cortex is locally heterogeneous but globally tonotopic [Publisher: eLife Sciences Publications, Ltd]. *eLife*, *6*. https://doi.org/10.7554/elife.25141

BIBLIOGRAPHY

Venezia, J. H., Richards, V. M., & Hickok, G. (2021). Speech-Driven Spectrotemporal Receptive Fields Beyond the Auditory Cortex. *Hearing Research*, *408*, 108307. https://doi.org/10.1016/j.heares.2021.108307

Verschooten, E., Shamma, S., Oxenham, A. J., Moore, B. C., Joris, P. X., Heinz, M. G., & Plack, C. J. (2019). The upper frequency limit for the use of phase locking to code temporal fine structure in humans: A compilation of viewpoints [Publisher: Elsevier BV]. *Hearing Research*, *377*, 109–121. https://doi.org/10.1016/j.heares.2019.03.011

Walzl, E. M., & Woolsey, C. N. (1946). Effects of cochlear lesions on click responses in the auditory cortex of the cat. *Bulletin of the Johns Hopkins Hospital*, *79*(4), 309–319.

Wang, X. (2007). Neural coding strategies in auditory cortex. *Hearing Research*, *229*(1-2), 81–93. https://doi.org/10.1016/j.heares.2007.01.019

Wang, X., Lu, T., Snider, R. K., & Liang, L. (2005). Sustained firing in auditory cortex evoked by preferred stimuli. *Nature*, *435*(7040), 341–346. https://doi.org/10.1038/nature03565

Webster, D. B., Popper, A. N., & Fay, R. R. (1992). *The mammalian auditory pathway: Neuroanatomy*. Springer.

Wei, Z., Lin, B.-J., Chen, T.-W., Daie, K., Svoboda, K., & Druckmann, S. (2020). A comparison of neuronal population dynamics measured with calcium imaging and electrophysiology (B. S. Gutkin, Ed.). *PLOS Computational Biology*, *16*(9), e1008198. https://doi.org/10.1371/journal.pcbi.1008198

White, L. E., & Fitzpatrick, D. (2007). Vision and Cortical Map Development [Publisher: Elsevier BV]. *Neuron*, *56*(2), 327–338. https://doi.org/10.1016/j.neuron.2007.10.011

Wiesenfeld, K., & Moss, F. (1995). Stochastic resonance and the benefits of noise: From ice ages to crayfish and SQUIDs [Publisher: Springer Science and Business Media LLC]. *Nature*, *373*(6509), 33–36. https://doi.org/10.1038/373033a0

Williamson, R. S., Ahrens, M. B., Linden, J. F., & Sahani, M. (2016). Input-Specific Gain Modulation by Local Sensory Context Shapes Cortical and Thalamic Responses to Complex Sounds. *Neuron*, *91*(2), 467–481. https://doi.org/10.1016/j.neuron.2016.05.041

Willmore, B. D. B., & King, A. J. (2023). Adaptation in auditory processing. *Physiological Reviews*, *103*(2), 1025–1058. https://doi.org/10.1152/physrev.00011.2022

Willmore, B. D., Schoppe, O., King, A. J., Schnupp, J. W., & Harper, N. S. (2016). Incorporating Midbrain Adaptation to Mean Sound Level Improves Models of Auditory Cortical Processing. *The Journal of Neuroscience*, *36*(2), 280–289. https://doi.org/10.1523/JNEUROSCI.2441-15.2016

Winkler, I., Denham, S. L., & Nelken, I. (2009). Modeling the auditory scene: Predictive regularity representations and perceptual objects. *Trends in Cognitive Sciences*, *13*(12), 532–540. https://doi.org/10.1016/j.tics.2009.09.003

Winkowski, D. E., Bandyopadhyay, S., Shamma, S. A., & Kanold, P. O. (2013). Frontal Cortex Activation Causes Rapid Plasticity of Auditory Cortical Processing. *The Journal of Neuroscience*, *33*(46), 18134–18148. https://doi.org/10.1523/JNEUROSCI.0180-13.2013

Winkowski, D. E., & Kanold, P. O. (2013). Laminar Transformation of Frequency Organization in Auditory Cortex. *The Journal of Neuroscience*, *33*(4), 1498–1508. https://doi.org/10.1523/JNEUROSCI.3101-12.2013

Winkowski, D. E., Nagode, D. A., Donaldson, K. J., Yin, P., Shamma, S. A., Fritz, J. B., & Kanold, P. O. (2018). Orbitofrontal Cortex Neurons Respond to Sound and Activate Primary Auditory Cortex Neurons. *Cerebral Cortex*, *28*(3), 868–879. https://doi.org/10.1093/cercor/bhw409

BIBLIOGRAPHY

Woolley, S. M. N., Fremouw, T. E., Hsu, A., & Theunissen, F. E. (2005). Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. *Nature Neuroscience*, *8*(10), 1371–1379. https://doi.org/10.1038/nn1536

Woolley, S. M. N., Gill, P. R., Fremouw, T., & Theunissen, F. E. (2009). Functional Groups in the Avian Auditory System [Publisher: Society for Neuroscience]. *The Journal of Neuroscience*, *29*(9), 2780–2793. https://doi.org/10.1523/jneurosci.2042-08.2009

Woolley, S. M. N., Gill, P. R., & Theunissen, F. E. (2006). Stimulus-Dependent Auditory Tuning Results in Synchronous Population Coding of Vocalizations in the Songbird Midbrain [Publisher: Society for Neuroscience]. *The Journal of Neuroscience*, *26*(9), 2499–2512. https://doi.org/10.1523/jneurosci.3731-05.2006

Wundt, W. M. (1887). *Grundzüge der physiologischen Psychologie.*

Zhang, Y., & Looger, L. L. (2024). Fast and sensitive GCaMP calcium indicators for neuronal imaging. *The Journal of Physiology*, *602*(8), 1595–1604. https://doi.org/10.1113/JP283832

Zhang, Y., Rózsa, M., Liang, Y., Bushey, D., Wei, Z., Zheng, J., Reep, D., Broussard, G. J., Tsang, A., Tsegaye, G., Narayan, S., Obara, C. J., Lim, J.-X., Patel, R., Zhang, R., Ahrens, M. B., Turner, G. C., Wang, S. S.-H., Korff, W. L., ... Looger, L. L. (2023). Fast and sensitive GCaMP calcium indicators for imaging neural populations. *Nature*, *615*(7954), 884–891. https://doi.org/10.1038/s41586-023-05828-9

Zhou, M., Liang, F., Xiong, X. R., Li, L., Li, H., Xiao, Z., Tao, H. W., & Zhang, L. I. (2014). Scaling down of balanced excitation and inhibition by active behavioral states in auditory cortex [Publisher: Springer Science and Business Media LLC]. *Nature Neuroscience*, *17*(6), 841–850. https://doi.org/10.1038/nn.3701