THE UNIVERSITY *of* EDINBURGH

Analysing the Information Contributions and Anatomical Arrangement of Neurons in Population Codes

Stuart Yarrow

Doctor of Philosophy
Institute for Adaptive and Neural Computation
School of Informatics
University of Edinburgh

2014

Abstract

Population coding—the transmission of information by the combined activity of many neurons—is a feature of many neural systems. Identifying the role played by individual neurons within a population code is vital for the understanding of neural codes. In this thesis I examine which stimuli are best encoded by a given neuron within a population and how this depends on the informational measure used, on commonly-measured neuronal properties, and on the population size and the spacing between stimuli. I also show how correlative measures of topography can be used to test for significant topography in the anatomical arrangement of arbitrary neuronal properties.

The neurons involved in a population code are generally clustered together in one region of the brain, and moreover their response selectivity is often reflected in their anatomical arrangement within that region. Although such topographic maps are an often-encountered feature in the brains of many species, there are no standard, objective procedures for quantifying topography. Topography in neural maps is typically identified and described subjectively, but in cases where the scale of the map is close to the resolution limit of the measurement technique, identifying the presence of a topographic map can be a challenging subjective task. In such cases, an objective statistical test for detecting topography would be advantageous. To address these issues, I assess seven measures by quantifying topography in simulated neural maps, and show that all but one of these are effective at detecting statistically significant topography even in weakly topographic maps.

The precision of the neural code is commonly investigated using two different families of statistical measures: (i) Shannon mutual information and derived quantities when investigating very small populations of neurons and (ii) Fisher information when studying large populations. The Fisher information always predicts that neurons convey most information about stimuli coinciding with the steepest regions of the tuning curve, but it is known that information theoretic measures can give very different predictions. Using a Monte Carlo approach to compute a stimulus-specific decomposition of the mutual information (the stimulus-specific information, or SSI) for populations up to hundreds of neurons in size, I address the following questions: (i) Under what conditions can Fisher information accurately predict the information transmitted by a neuron within a population code? (ii) What are the effects of level of trial-to-trial variability (noise), correlations in the noise, and population size on the best-encoded stimulus? (iii) How does the type of task in a behavioural experiment (i.e. fine and coarse discrimination, classification) affect the best-encoded stimulus? I show that, for both unimodal and monotonic tuning curves, the shape of the SSI is dependent upon trial-to-trial variability, population size and stimulus spacing, in addition to the shape of the tuning curve. It is therefore important to take these factors into account when assessing which stimuli a neuron is informative about; just knowing the tuning curve may not be sufficient.

Acknowledgements

Firstly I would like to thank my supervisor, Peggy Seriès, for counteracting my natural tendency to go round in circles and keeping me headed in the right direction, and my collaborators Khaleel Razak and Aaron Seitz for the opportunity to work in their labs, many great discussions and their hospitality during my visits to UC Riverside. I would also like to thank Dan Butts, Mark Goldman, Matthias Bethge, David Willshaw and Jim Bednar for several useful discussions about information and topographic maps. Finally, I would like to thank my family and friends, and Michelle in particular, for their unflagging support and encouragement over the last five years.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

I confirm that the work submitted is my own, except where work which has formed part of jointly-authored publications has been included. My contribution and those of the other authors to this work have been explicitly indicated below. I confirm that appropriate credit has been given within this thesis where reference has been made to the work of others.

The work presented in Chapter 3, together with parts of other chapters, was previously published in *PLOS One* as 'Detecting and Quantifying Topography in Neural Maps' by Stuart Yarrow, Khaleel Razak, Aaron Seitz and Peggy Seriès. This study was conceived by all of the authors. I carried out the analysis based on data gathered by Khaleel Razak, prepared the figures, and drafted the manuscript. All of the authors were involved in reviewing and final editing of the published manuscript.

The work presented in Chapter 4, together with parts of other chapters, was previously published in *Neural Computation* as 'Fisher and Shannon Information in Finite Neural Populations' by Stuart Yarrow, Edward Challis and Peggy Seriès. This study was conceived by Peggy Seriès and I, building upon unpublished preliminary work by Edward Challis (Challis, 2007) and myself (Yarrow, 2008). I ran the simulations, analysed the results, created the figures and drafted the manuscript. All of the authors were involved in reviewing and final editing of the published manuscript.

The work presented in Chapter 5, together with parts of other chapters, has been submitted for publication in *Frontiers in Computational Neuroscience* as 'The Influence of Population Size, Noise and Behavioural Task on Best-Encoded Stimulus for Neurons with Unimodal or Monotonic Tuning Curves' by Stuart Yarrow and Peggy Seriès. This study was conceived by both of the authors. I ran the simulations, analysed the results, created the figures and drafted the manuscript. Both authors were involved in reviewing and final editing of the manuscript that was submitted for publication.

(*Stuart Yarrow*)

Table of Contents

List of Figures

List of Symbols

β	Gain modulation scaling factor
ΔI	Interaural intensity difference
ΔI_{50}	Characteristic (50% response) IID
Δs	Stimulus spacing
δs	Normalised stimulus spacing
ω	Tuning curve width parameter
ω_{mod}	Width scaling parameter for gain modulation
ϕ	Characteristic stimulus (circular stimulus space)
ϕ_{mod}	Centre of gain modulation profile (adapting stimulus)
Π	Statistical power
ψ	Auditory source azimuth
ψ_{50}	Characteristic (50% response) azimuth
ρ	Range coefficient for localised noise correlations
σ	Characteristic stimulus
τ	Spike count integration time
Θ	Stimulus ensemble (circular stimulus space)
θ	Stimulus value (circular stimulus space)
$\boldsymbol{m} = \{x, y\}$	Map space coordinates
$\boldsymbol{r}$	Population response
C	Noise correlation matrix
c	Scaling factor for noise correlations
C_Z	Zrehen measure
C_{PC}	Pearson distance correlation
C_{PL}	Path length measure
C_{SC}	Spearman distance correlation
C_{TC}	Topological correlation
C_{WL}	Wiring length measure
D_C	Chernoff distance
F	Fano factor
f	Tuning curve function
f_ς	Sigmoidal azimuth tuning function model
f_G	Gaussian azimuth tuning function model

f_M	Monotonic tuning curve model
f_U	Unimodal tuning curve model (linear stimulus space)
f_{bg}	Background firing rate
f_{mod}	Stimulus-dependent firing rate modulation depth
H	Shannon entropy
h	Differential Shannon entropy
I_{Fisher}	Information-theoretic quantity derived from Fisher information
I_{mut}	Mutual information
I_{SI}	Specific information
I_{SSI}	Stimulus-specific information
I_{sur}	Specific surprise
J	Fisher information
N	Number of neurons
N_{80}	Number of data required for 80% statistical power
N_{best}	N_{80} of most powerful map measure
P	Probability mass
p	Probability density
P_T	Topographic product
Q	Noise covariance matrix
R	Response ensemble
r	Response spike count
S	Stimulus ensemble (linear stimulus space)
s	Stimulus value (linear stimulus space)
Z	Hypothetical observer's knowledge of stimulus
z	Feature space coordinate

Chapter 1

Introduction

1.1 Population codes

To survive, animals must interact with their environment, assessing their surroundings to help determine their next action. For the brain to perform the information processing needed to support behaviour, it must first form an internal representation of its environment. This is the function of the sensory nervous systems: to translate light, sound, motion, pressure, heat and chemical stimuli into patterns of action potentials. The field of sensory neural coding aims to understand these patterns of spikes, the relationships between environmental stimuli and neural responses, what is encoded, and how precisely it is represented. There is no one neural code; different codes are used by different regions of brains, and information about different stimulus features can be multiplexed together by being encoded in different aspects of neural activity. Many aspects of neuronal activity have been observed to be information bearing, for instance: spike counts or firing rates (Adrian and Zotterman, 1926a), timing of spikes (for example, auditory phase locking to acoustic waveforms Galambos and Davis, 1943), and timing of individual spikes relative to oscillations in the activity of a wider neural ensemble (see e.g. O'Keefe and Recce, 1993; Montemurro et al., 2007).

Population coding is the representation of information by the collective activity of a group of neurons, and is a common mode of information transmission in neural systems. In this thesis, we address a model of population coding in which information is transmitted by firing rates alone. Although the assumption of rate coding is a simplification in that the times of individual spikes are ignored, firing rates have been observed to carry information in many instances, so the rate population coding model remains important despite its limitations. For the purposes of this thesis, I focus on population codes that represent a quantitative stimulus feature, specifically a real-valued or angular quantity. Well

known examples of this type of population code include the codes for direction of visual motion in the medial temporal area (MT; Dubner and Zeki, 1971), movement direction in motor cortex (Georgopoulos et al., 1986), place cells in the hippocampus (O'Keefe and Dostrovsky, 1971), and the concurrent representation of many local visual stimulus features—such as edge position and orientation—in the primary visual cortex (V1; Hubel and Wiesel 1959 and e.g. Berens et al. 2012).

1.1.1 Modelling population codes

Population codes have been the subject of intensive study by theoretical neuroscientists due to their widespread occurrence, relative simplicity, and amenability to mathematical analysis. This has resulted in a well-established model framework for the study of population coding. The terminology and concepts of this framework are heavily influenced by classical electrophysiology methods, so it is useful to introduce the population coding model in this context. In a typical experiment a simple, tightly controlled stimulus is presented while the spiking activity of a neuron is recorded. For concreteness, let us assume that the stimulus is a moving pattern of dots on a computer screen, the stimulus feature that we are interested in is the direction of movement, and we are recording from a motion-sensitive neuron in area MT. Over the course of the experiment each direction of motion is presented many times to establish the spread of responses associated with that direction, and the entire process is repeated for many different directions to map out how the response of the neuron varies depending on the stimulus. In each trial, the number of spikes emitted by the neuron in a defined time window following the stimulus presentation are counted. In this way, sufficient data can be gathered to estimate the mean response as a function of the stimulus: the tuning curve (although the term 'tuning curve' has a more specific meaning in auditory neuroscience, here it is used to denote a function relating a mean neuronal response to an arbitrary stimulus feature, rather than to sound frequency in particular). In addition to the tuning curve, the data from our hypothetical experiment can be used to estimate the distribution of responses for a given stimulus, often by fitting a parametric probability distribution. In this way, by repeating the experiment for many neurons within the population, the experimenter can construct a probabilistic model of the population code (Figure 1.1). This model takes the form of a conditional distribution $p(R|\Theta)$ of the neuronal responses R conditioned upon the stimulus Θ.

Experimentally measured tuning curves come in a wide variety of shapes, but two basic types are often observed. The first of these types, the unimodal tuning curve, is the classic peaked, bell-shaped curve as shown in Figure 1.1. This type of tuning curve has been observed in many species, brain areas and for many stimulus features, including motion direction in MT (as discussed above), edge orientation

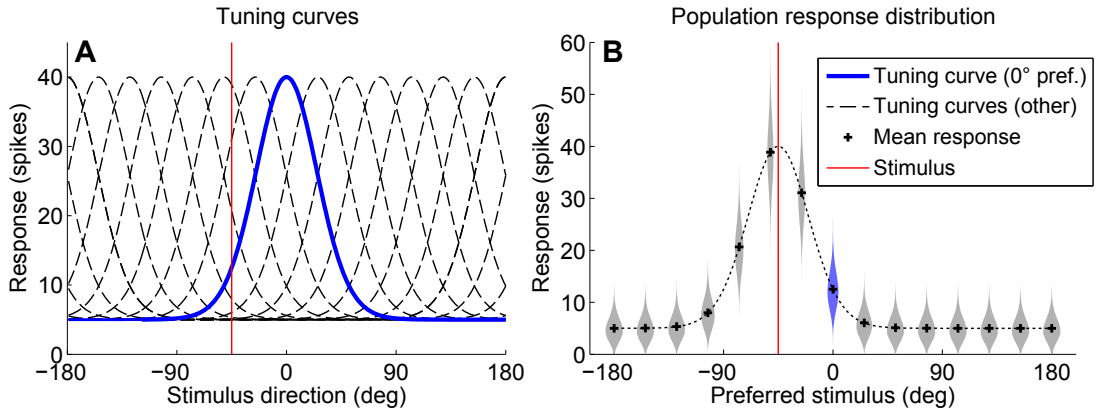

Figure 1.1: A basic population code model. A typical population coding model consists of two components: tuning curves (**A**) describing the mean response as a function of the stimulus, and a parametric probability distribution (Poisson in this case) describing the spread of responses. (**B**) illustrates the distribution of responses for a stimulus angle of -45° (indicated on both plots by a vertical red line). The mean responses and response distributions (widths of shaded regions are proportional to spike count probability) are plotted at the characteristic stimulus of the neuron—the tuning curve peak in this case. In this simplified model all tuning curves are identical, shifted copies, which means that the shape of the population response profile (dotted line) is the same as the tuning curves. The neuron with a characteristic stimulus of 0° is shown in blue to highlight the correspondence between panels.

in V1 simple cells (Hubel and Wiesel, 1962), sound frequency in many parts of the auditory system, and wind direction in the cricket cercal system (Theunissen and Miller, 1991). The second type, the monotonic or sigmoidal tuning curve, is often found when measuring the response of neurons to stimuli of different strengths, for instance relative luminance in the visual system (e.g. Sakmann and Creutzfeldt, 1969), sound intensity in the auditory system (e.g. Sachs and Abbas, 1974), and pressure of touch in the somatosensory system (Adrian and Zotterman, 1926b).

One of the principal posited advantages of population coding is that it allows the detrimental effect of variability on the precision of the code (i.e. the fidelity of the code; the precision with which the stimulus is represented) to be ameliorated by combining information from many active neurons. Trial-to-trial variability in neuronal responses (see Faisal et al., 2008, for a review) has historically been considered to be random, uninformative noise, although there is a growing body of evidence that this is not necessarily a safe assumption (see e.g. Stein et al., 2005; Goris et al., 2014). Variability in neural systems is often Poisson-like, in that the trial-to-trial variance of the spike count is close to or, more generally, proportional to the mean spike count. While measuring the variance of single neurons is quite straightforward, higher order statistics can also affect the amount of information transmitted by the code. Classical single electrode techniques only allow the recording of one, or very few, cells simultaneously, which means that

it is not possible to measure inter-neuronal dependencies in the trial-to-trial variability—so-called 'noise correlations'. In this situation it is usual to assume that the activities of each cell are conditionally independent given the stimulus, i.e. that the trial-to-trial variability is independent. However, this can lead to both under and over estimation of the precision of the code (see Averbeck et al., 2006, for a review). In order to accurately characterise correlations in variability, it is necessary to simultaneously record from multiple cells. The situation is further complicated by the number of neurons involved; in all but the simplest systems—such as the four-neuron code in the cricket cercal system (Theunissen and Miller, 1991)—it is impossible to identify and all cells involved, let alone record their activity. This means that the measured activity can only be a small sample of the activity of the population, although this is now changing due to the increasing use of multi-electrode arrays and high resolution methods for imaging activity.

1.1.2 Measuring precision

Both experimentalists and theorists have shown great interest in developing tools to assess the precision of population codes by quantifying the amount of information that they transmit. Such methods can be used to help understand the relationship between neural representations and behaviour, as well as between neural activity and environmental stimuli. Information measures are also useful for assessing the functional consequences of changes in neural response properties, such as those linked to sensory adaptation and attention. Measuring the information transmitted by a population essentially involves quantifying the degree to which the neuronal activities reflect the stimulus feature of interest; this question can be addressed using a number of statistical methods.

Two families of statistical measure are commonly used to assess the precision of neural codes. The first of these, the Fisher information (Fisher, 1925), has mainly been used to analyse large populations, while measures based on Shannon's information theory (Shannon, 1948) have been used mostly in the analysis of single neurons and small populations (see Chapter 2 for further details). As well as indicating the overall precision of the code, both families of measure can be used to quantify the information conveyed by a neuron about specific stimuli. This is useful as it allows the informational contributions of individual neurons, and hence their role within the population, to be studied. The Fisher information is itself a function of the stimulus, but the fundamental quantity of information theory, the mutual information, is a scalar and quantifies only the overall precision across all stimuli. There are a number of ways to decompose the mutual information into stimulus-specific components; here we focus on one of these: the stimulus-specific information (SSI; Butts, 2003). While both the Fisher information and information theoretic measures are useful, they are fundamentally different and

have different units, which makes it difficult to directly compare the results of analyses that use different measures. Because of this, it is not always clear exactly what they imply or how they relate to each other.

In this thesis we examine how the measures introduced above can be used to quantify the amount of information transmitted by the activity of a population as a whole. Depending on the nature of the code, it may be difficult to design a decoder to efficiently read out the population activity (i.e. while utilising all of the available information) in order to make a decision or reconstruct the stimulus. This is particularly true if the decoder architecture is constrained to be a biologically plausible neural network. The topic of population code readout has received considerable attention in the literature (see e.g. Seung and Sompolinsky, 1993; Deneve et al., 1999), and codes that allow for efficient decoding by simple networks are often considered to be advantageous (see Froudarakis et al., 2014, for a recent example). It is not clear, however, that such decoding actually occurs in the brain; it is entirely possible that information is processed in the brain primarily through transformations operating on distributed codes. Under such a scheme, there is no need for signals to be concentrated into reliable activity in single neurons. Thus, while decoding population activity is certainly a powerful tool for the study of neural coding and in neuroengineering, the relevance of neural implementations of decoders is unclear. For this reason, we leave aside the issue of neural decoding and focus on quantifying information in distributed representations.

1.1.3 Maps and topography

A recurring feature of sensory—and other—areas of the brain is that they tend to be organised such that neurons with similar functions (such as those involved in a population code) are located close together. Different areas of the brain specialise in different sensory modalities and, within modalities, in the processing of specific stimulus features. Moreover, the neurons within these areas are often arranged within their brain region in such a way that they form a topographic map of one or more stimulus features. For example, the visual cortex is largely concerned with processing visual information and consists of areas such as V1, which is thought to perform low-level feature extraction, and which is organised as a topographic map of the visual field, overlaid with maps of other features such as contour orientation. In addition to the visual cortex (Hubel and Wiesel, 1962), topographic maps are found in the auditory (Woolsey and Walzl, 1942), somatosensory (Penfield and Boldrey, 1937) and motor (Ferrier, 1874) areas, in many subcortical structures, and even in areas of cortex associated with higher functions (Silver and Kastner, 2009). The widespread occurrence of topographic maps, and the fact that they are in many cases conserved through multiple stages of neural processing, strongly suggest that they are associated with some significant evolutionary advantage.

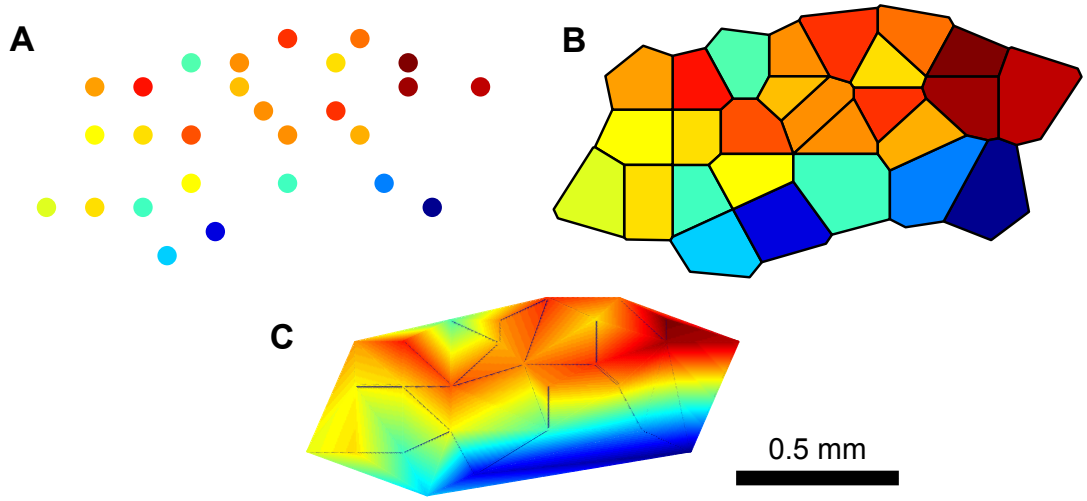

Figure 1.2: Three visualisations of the same neural map. The perceived orderliness of a topographic map can vary depending on how the data is presented. This figure shows identical mapping data (auditory source azimuth tuning in A1 of the pallid bat; Razak 2011) plotted in three ways: (**A**) no interpolation, (**B**) Voronoi tessellation (nearest neighbour interpolation), and (**C**) linear interpolation. Colour indicates the value of the mapped tuning parameter and colour scaling is continuous and consistent across panels. This dataset is explored further in Chapter 3 and Figure 3.7.

Many experimental techniques have been used to observe neural maps, including single-electrode (Hubel and Wiesel, 1962) and multi-electrode (Krüger and Bach, 1981) electrophysiology, optical intrinsic signal imaging (Grinvald et al., 1986), fMRI (Engel et al., 1994), calcium imaging (Hübener and Bonhoeffer, 2005) and microstimulation (Penfield and Boldrey, 1937). Topography of intrinsic neuronal properties has also been observed using intracellular recording techniques (see O'Donnell and Nolan, 2011, for a review). These methods vary widely in spatial resolution, and in the number of points (neurons or ensembles of neurons) at which response properties can be measured simultaneously. Because it is normally identified subjectively by visual inspection, topography on a scale close to the resolution limit of the observation technique may be difficult to identify, as the spatial density of measurements required to characterise the map becomes difficult to achieve. Also, map measurement techniques that rely on serial measurements (e.g. single-electrode electrophysiology) limit the number of points that can be measured in any one experimental subject. In cases such as these, detecting the presence of a topographic map is a non-trivial task. It is also difficult to reliably estimate the degree of topographic organisation by visual inspection (Alvarez et al., 1998), and the method used to graphically represent the map can affect the perceived degree of topography (see Figure 1.2).

1.2 Overview of thesis

Despite the simplicity of the population coding model, and although population codes have been observed by experimentalists for several decades and are one of the most intensively studied areas in theoretical neuroscience, some fundamental aspects of population coding are not yet fully understood. This section outlines the open questions addressed in this thesis and summarises the work undertaken to answer them.

1.2.1 Topography

Neurons that make up a population code are generally located close together anatomically and it is not unusual for them to be organised as a topographic map of one or more stimulus features. Intuitively, the defining property of a topographic map is that, within the map, anatomically proximate locations are occupied by neurons with similar functional properties. However, distilling this intuitive understanding into a more rigorous definition of topography is not straightforward (Bauer et al., 1999). Any formal definition of topography rests upon how similarity of both functional properties and anatomical location are quantified; different methods of measuring difference or distance lead to differing definitions of topography. Concordant with the prevalence of a loosely defined notion of topography, neural topographic maps are normally identified subjectively and described qualitatively, and surprisingly few attempts have been made to quantify the degree of topography in experimentally observed maps. Whilst the subjective and qualitative treatment of neural maps has well-established utility, some situations demand a more rigorous approach, for instance where the topographic organisation is weak or where the map can only be sampled at a limited number of points. In situations like these, an objective method of detecting significant topography in a map would be advantageous, as it would eliminate the need to rely upon subjective judgement. Topographic maps are known to be sensitive to both biological and environmental factors and a well understood quantitative measure of topography would be of broad utility for making objective comparisons between maps.

To address this, we analyse seven statistical measures to assess their suitability for quantifying and detecting topography in neural maps. The use of the measures to quantify topography and test for significance is then demonstrated in an analysis of source azimuth and interaural intensity difference maps in A1 of the pallid bat. The results of this analysis show that there is statistically significant topographic organisation of spatial tuning properties in pallid bat A1.

The form and degree of topographic organisation in neural maps depends not only on the response properties of the neurons and their physical positions within

the brain, but also on how we choose to label them, how we determine the characteristic stimuli of the neurons that define the map itself. In mapping studies, neurons are commonly labelled with the stimulus that produces the maximum response (i.e. the peak of the tung curve), but this is problematic if the tuning curve is monotonic or has multiple peaks. Another possible approach when analysing what a neuron represents and assigning labels is to consider information rather than activity: which stimuli does a neuron provide the most information about?

1.2.2 Best-encoded stimuli

A fundamental question in the analysis of a population code is: what is the significance of individual neurons within the population? What information does a neuron contribute to the code? Tuning curves are widely used to describe neuronal activity, but their implications in terms of the information contributed to the code by a neuron are still not fully understood and they remain open to misinterpretation. In particular, there is a tendency for neurons to be associated by default with the stimuli that trigger their strongest responses. This is sometimes stated explicitly, but is also implicit in the language used to describe response properties, for example in the term 'preferred stimulus' or when a neuron is described as being selective for a particular stimulus. While this is a convenient way to refer to tuning curves, it is not a reliable indication of what information a neuron contributes to a population code—what its *informational* tuning curve is, and hence which stimulus (or stimuli) it conveys the most information about: its best-encoded stimulus. Both Fisher information and information theoretic measures can be used to quantify the amount of information transmitted by a neuron about specific stimuli, and hence to identify the best-encoded stimulus. However, the two types of measure do not always indicate the same best-encoded stimuli. Specifically, the Fisher information always predicts that a neurons provides the most information about stimuli where the gradient of the tuning curve is high, for instance the sloping flanks of a peaked tuning curve. The best-encoded stimulus predicted by the SSI, however, depends on the level of trial-to-trial variability in the code (Butts and Goldman, 2006) and, while it agrees with the Fisher information when the variability is low, it yields very different results for high variability. In the latter case, the SSI predicts that the best-encoded stimulus falls at the peak of the tuning curve.

Until now, the SSI has only been calculated for very small populations with unimodal tuning curves and uncorrelated variability. Here we examine the effect of variability level, correlations in the variability, population size and stimulus spacing on the SSI for both unimodal and monotonic tuning curves. How do these factors influence the best-encoded stimulus predicted by the SSI? When do the predictions of the Fisher information and SSI agree?

1.2.3 Fisher information and mutual information

Both Fisher and Shannon information can also be used to quantify the overall precision of a population code. Although they have been proven to be equivalent for populations with an infinite number of neurons (Brunel and Nadal, 1998), their predictions can differ for smaller populations. We investigate the relationship between Fisher information and mutual information in population codes with a finite number of neurons, examining the effect of variability level, correlations in the variability, and population size. How large does a population have to be before the Fisher information can provide a useful indication of the precision of the code?

1.2.4 Outline

The remainder of this thesis is organised as follows: Chapter 2 reviews previous research on the quantification of information in population codes and topography in neural maps, describing the work that forms the background and foundation of the subsequent chapters. Chapter 3 addresses the quantification of topography in population codes by analysing map measures to assess their suitability for quantifying and detecting topography in neural maps. The following two chapters deal with quantifying information transmitted by population codes and by neurons that form part of a population code. Using population coding models with unimodal tuning curves, Chapter 4 examines the effect of population size and trial-to-trial variability, including correlations, on the relationship between the overall precision indicated by the Fisher information and the mutual information. The relationship between the best-encoded stimulus predictions of the Fisher information and the SSI is also investigated. This analysis is then extended to population codes based on monotonic tuning curves in Chapter 5, which then goes on to show how the best-encoded stimulus of population coding neurons is influenced by the spacing between stimuli. Chapter 6 discusses the implications and limitations of the work, and potential directions for future research.

Chapter 2

Population Codes: Information and Maps

This chapter introduces statistical methods for quantifying information in population codes and topography of neural maps, and reviews their application in the literature. The chapter first introduces the principle measures used to assess coding precision and best-encoded stimuli. It then goes on to review research in which these measures have been applied in the analysis of population codes. Finally, it gives a detailed review of previous research involving the quantitative analysis of neural maps.

2.1 Quantifying information

2.1.1 Information theory

Information theory is a mathematical framework developed in the 1940s by engineer and mathematician Claude Shannon (Shannon, 1948). While originally intended as a tool for analysing telecommunications systems, information theory is more generally applicable and has been widely utilised in other fields (Cover and Thomas, 2006). In contrast to many other statistical techniques, information theory does not rely on any assumptions about the form of distributions or the properties of underlying processes. It provides a powerful and general framework for studying neural coding as it takes into account all forms of statistical dependency (not just correlations) and does not rely on assumptions about the form of probability distributions.

The basic quantity of information theory is information entropy, a measure of the uncertainty or randomness of a variable. Entropy can be intuitively, but

very loosely, thought of as a generalisation of variance; while variance has a special relevance to the Gaussian distribution, entropy is equally applicable to any probability distribution. More correctly, entropy is the amount of information required, on average, to represent the value of a variable and is measured in bits (base 2 logarithms) or nats (base e logarithms); bits are used throughout this thesis. The entropy $H(\Theta)$ of a stimulus ensemble Θ is given by:

$$H(\Theta) = -\sum_{\theta \in \Theta} p(\theta) \log_2 p(\theta) \tag{2.1}$$

Shannon or mutual information I_{mut}, is a measure of the informativeness of one variable about another e.g. of a neural response R about a stimulus Θ. It is the portion of the entropy of one variable that can be explained by the other variable. Specifically, it is the total entropy minus the conditional entropy:

$$\begin{aligned} I_{mut}(\Theta, R) &= H(R) - H(R|\Theta) = H(\Theta) - H(\Theta|R) \\ &= \sum_{\theta \in \Theta} p(\theta) \sum_{r \in R} p(r|\theta) \log \frac{p(r|\theta)}{p(r)} \end{aligned} \tag{2.2}$$

Uppercase characters Θ and R represent the stimulus and response ensembles, while lowercase characters (θ, r) represent a single value within the ensemble.

Mutual information can be used to quantify the information provided by an entire response ensemble about an entire stimulus ensemble, but it cannot inform us about the precision with which specific stimuli within the ensemble are encoded. To address this, several decompositions of the mutual information have been proposed (see Butts, 2003, for a review), in particular the stimulus-specific surprise, specific information and stimulus-specific information.

Stimulus-specific surprise (SSS) is the most widely used MI decomposition. Like all of the stimulus-specific measures described here, the average of the specific surprise over the stimulus ensemble is equal to the mutual information. Equation 2.3 illustrates an intuitive interpretation: the specific surprise is the reduction in surprise (log reciprocal probability) of a given stimulus, averaged over the response ensemble. The SSS was one of the first stimulus-specific measures to be applied to population coding (Theunissen and Miller, 1991), there referred to as local transinformation. Confusingly, SSS is also referred to in some articles as stimulus-specific information.

$$I_{sur}(\theta) = \sum_{r \in R} p(r|\theta) \log \frac{p(r|\theta)}{p(r)} = \sum_{r \in R} p(r|\theta) \left[\log \frac{1}{p(\theta)} - \log \frac{1}{p(\theta|r)} \right] \tag{2.3}$$

The specific information (SI) is a mutual information decomposition that quantifies the decrease in uncertainty about the stimulus due to the observation of a given response:

$$I_{SI}(r) = \sum_{\theta \in \Theta} p(\theta|r) \log p(\theta|r) - p(\theta) \log p(\theta) \tag{2.4}$$

The specific information has a unique and advantageous property in that it is additive (DeWeese and Meister, 1999): the sum over the specific information associated with a number of individual observations is equal to the specific information of the whole set considered jointly.

The stimulus specific information (SSI) is a stimulus-specific development of the specific information (Butts, 2003). The SSI is the average specific information associated with a given stimulus, therefore it quantifies the average reduction in entropy resulting from the presentation of a given stimulus.

$$\begin{aligned} I_{SSI}(\theta) &= \sum_{r \in R} p(r|\theta) I_{SI}(r) \\ &= \sum_{r \in R} p(r|\theta) \left[\sum_{\theta' \in \Theta} p(\theta'|r) \log p(\theta'|r) - p(\theta') \log p(\theta') \right] \end{aligned} \tag{2.5}$$

In this thesis we discuss both the population SSI (the SSI of the population as a whole) and the singleton SSI, which is the SSI of a single neuron considered in isolation. A closely related quantity, the marginal SSI (mSSI) for a particular neuron within a population, is defined as the difference between the population SSI and the SSI for the population of remaining neurons with the neuron of interest removed.

SSS and SSI are both stimulus-specific decompositions of the mutual information, so how do they differ? The two measures quantify different aspects of how an observers knowledge of the stimulus changes following the presentation of a specific stimulus θ: the SSI quantifies the change in entropy of the stimulus distribution, while the SSS measures the change in probability of θ (see Figure 2.1 and Butts 2003). The SSI thus tells us the average reduction in uncertainty—about all possible values of stimulus—that results from the presentation of a given stimulus. The SSS is the average amount by which the surprise of a given stimulus reduces, and hence its probability increases, following the presentation of that stimulus. In many cases the two measures yield qualitatively similar indications of the information associated with each stimulus, because the presentation of a well-encoded stimulus results in a concentration of probability mass around that stimulus value, leading to a reduction in entropy and an increase in the probability of the presented stimulus. Although it is possible to design codes for which the SSS and SSI predict very different best-encoded stimuli (Butts, 2003), the two measures are similar for population codes with smooth tuning curves (see Chapter 4). The fact that the SSI is based on Shannon entropy makes it particularly well suited to analysing information transmission, and for this reason it is primarily the SSI that is used in this thesis to determine best-encoded stimuli.

The SSI is a relatively recent development and has not yet been explored or applied as widely as the specific surprise. Until recently, the SSI was considered to be intractable for all but small populations; (Butts and Goldman, 2006) calculated the SSI for a maximum of four neurons and it has rarely been used in the analysis of experimental data, and only to analyse single neurons (Sawtell and

Williams, 2008; Wimmer et al., 2008; Remedios et al., 2009; Montgomery and Wehr, 2010). Challis (2007) demonstrated that, at least for independent Poisson trial-to-trial variability, this can be overcome through the use of Monte Carlo integration to compute the average over the high-dimensional response ensemble. This technique was later extended to Gaussian variability, both independent and correlated (Yarrow, 2008).

All information-theoretic measures have one major disadvantage in an experimental neuroscience context. In order to calculate any of these measures directly, it is necessary to establish the full joint distribution $p(\Theta, R)$. In a model this is relatively simple, but in an experimental context it is, at best, very difficult to record the number of trials necessary to establish an accurate joint distribution. One method that has been proposed to avoid the problem of constructing the joint distribution $p(\Theta, R)$ involves calculating a lower bound on the transmitted information using spike train metrics (Victor and Purpura, 1997; Victor, 2005). Since this method relies on stimulus-dependent clustering, it is inherently suited to assessing classification of discrete stimuli. Another approach, which is suited to assessing the discrimination of continuous-valued stimuli, is to estimate the mutual information by calculating the Fisher information, as described in section 2.1.3. One of the goals of this thesis is to assess the validity of this approximation.

2.1.2 Fisher information

Fisher information (Fisher, 1925) is a measure of the precision with which a parameter (typically a stimulus in studies of neural coding) of a parametric probability distribution (e.g. a distribution of neuronal responses) can be estimated, based on a sample from that distribution (e.g. a set of neuronal responses). It is commonly used in both theoretical (see e.g. Paradiso, 1988; Seung and Sompolinsky, 1993; Abbott and Dayan, 1999; Wilke and Eurich, 2002; Berens et al., 2011) and experimental (e.g. Jenison and Reale, 2003; Harper and McAlpine, 2004; Durant et al., 2007; Gutnisky and Dragoi, 2008) studies of population coding. The Fisher information J is defined as:

$$J(\theta) = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log p(r|\theta)\right)^2 \middle| \theta\right] \tag{2.6}$$

In a population code with Gaussian variability, mean response vector (tuning curves) $\boldsymbol{f}(\theta)$ and covariance matrix $Q(\theta)$, the Fisher information about θ is given by:

$$J(\theta) = \boldsymbol{f}'(\theta)^T Q(\theta)^{-1} \boldsymbol{f}'(\theta) + \frac{1}{2}\mathrm{Tr}\left[Q(\theta)^{-1} Q'(\theta) Q(\theta)^{-1} Q'(\theta)\right] \tag{2.7}$$

Despite its name, it is not a measure of information in the information theoretic sense; its units are those of the reciprocal of variance (e.g. deg^{-2} for an angular stimulus). Fisher information is perhaps more intuitive than the information

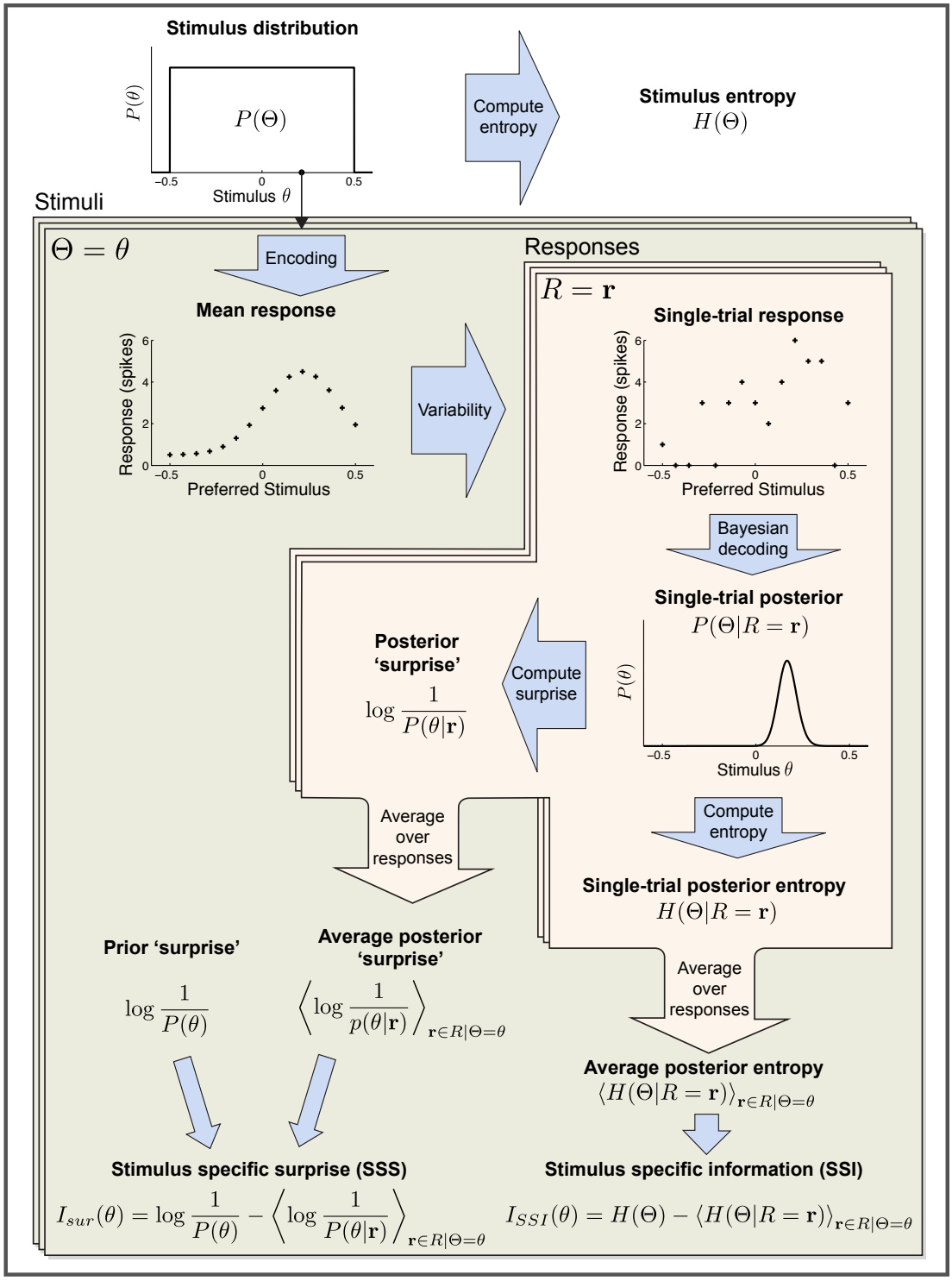

Figure 2.1: Comparison of the SSI and SSS.
An illustration of the differences between the stimulus-specific information (SSI) and stimulus-specific surprise (SSS). Where the SSI quantifies the change in stimulus entropy following the presentation of given stimulus, the SSS instead quantifies the change in the 'surprise' (log reciprocal probability) of the stimulus.

theoretic measures: its reciprocal defines a lower limit (the Cramér-Rao bound) on the variance of an unbiased estimator (a function of r that yields an estimate of θ), and hence the smallest achievable standard error. However, this level of precision is not necessarily achievable; the performance of an optimal estimator only approaches the Cramér-Rao bound asymptotically as the population size tends towards infinity; we describe this as the asymptotic regime. When the true precision of the code does not saturate that bound the Fisher information can be misleading (Bethge et al., 2002; Xie, 2002; Yarrow et al., 2012); we refer to this as the pre-asymptotic regime. Predicting what population size is required for effective saturation of the bound is non-trivial, and this question has rarely been addressed in the literature (Bethge et al., 2002; Xie, 2002). Fisher information should therefore be treated with some caution, as it is not always clear whether it indicates the true coding precision of a population.

A significant limitation of the Fisher information is that it is determined entirely by local properties of the tuning curve; as such, its relevance is restricted to tasks involving closely spaced stimuli, for example fine discrimination, and it can only be applied to continuous stimulus parameters. Because it is a function of the stimulus and, for typical Poisson-like trial-to-trial variability, it is approximately proportional to the first derivative of the tuning curve, the Fisher information predicts best-encoded stimuli at points of maximum tuning curve gradient (e.g. the flanks of peaked tuning curves). Correspondingly, the Fisher information is zero where the tuning curve gradient is zero—this includes tuning curve peaks. One major advantage of the Fisher information is that closed form solutions are available for many tuning curve and noise models. Therefore it is often easy to compute the Fisher information from parametric tuning curves and variability models, although constructing accurate parametric models from experimental data can be difficult in itself. For example, under a fitted Gaussian variability model the inverse covariance matrix is required in order to obtain the Fisher information. As this inverse matrix, and hence the Fisher information itself, is sensitive to inaccuracies in the fitted covariance matrix, it is important to first obtain an accurate estimate of the covariance matrix, which may require a large amount of data. Consequently, the Fisher information has only rarely been used to quantify the precision of experimentally-characterised population codes.

2.1.3 Linking Fisher and Shannon

Brunel and Nadal (1998) linked Fisher and Shannon information by proposing I_{Fisher}, a new information theoretic measure derived from Fisher information. They considered an optimal estimator $\hat{\Theta}(R)$, computed from R, with a Gaussian conditional distribution $p(\hat{\Theta}(R)|\Theta)$ and variance that saturates the Cramér-Rao bound. This is equivalent to assuming that the population size is infinite and

therefore that each estimate $\hat{\theta}(\boldsymbol{r})$ is based on an infinite number of independent observations. Given these assumptions, we can determine the conditional entropy of the estimator $h(\hat{\Theta}(R)|\Theta = \theta)$ from the variance, and hence from the Fisher information, using the following relation:

$$h(\hat{\Theta}(R)|\Theta = \theta) = \frac{1}{2}\log_2(2\pi e\sigma^2) = \frac{1}{2}\log_2\left(\frac{2\pi e}{J(\theta)}\right) \tag{2.8}$$

Since we are dealing with a continuous variable this is a differential entropy, denoted by a lowercase h and defined as $h(X) = -\int_X p(x)\log p(x)dx$. Equation 2.8 gives the conditional entropy for a specific stimulus value θ; to obtain $h(\hat{\Theta}(R)|\Theta)$ it is necessary to take the average over the stimulus ensemble:

$$h(\hat{\Theta}(R)|\Theta) = \mathbb{E}_\Theta[h(\hat{\Theta}(R)|\Theta = \theta)] = \int_\Theta p(\theta)\frac{1}{2}\log_2\left(\frac{2\pi e}{J(\theta)}\right)d\theta \tag{2.9}$$

The mutual information of stimulus and estimator is therefore given by:

$$I_{mut}(\Theta, \hat{\Theta}(R)) = h(\hat{\Theta}(R)) - \int_\Theta p(\theta)\frac{1}{2}\log_2\left(\frac{2\pi e}{J(\theta)}\right)d\theta \tag{2.10}$$

Using the data processing inequality to relate $I_{mut}(\Theta, \hat{\Theta}(R))$ and $I_{mut}(\Theta, R)$:

$$I_{mut}(\Theta, R) \geq h(\hat{\Theta}(R)) - \int_\Theta p(\theta)\frac{1}{2}\log_2\left(\frac{2\pi e}{J(\theta)}\right)d\theta \tag{2.11}$$

and showing that i) this inequality becomes an equality in the limit of large N and under certain regularity conditions, and ii) that $h(\hat{\Theta}(R)) \to h(\Theta)$ in the limit where the estimator is sharply peaked around its mean value (i.e. $J(\theta) \gg 1$), Brunel and Nadal show that $I_{mut}(\Theta, R)$ can be approximated by a measure that they call I_{Fisher}, since it is defined in terms of Fisher information:

$$I_{Fisher} = h(\Theta) - \int_\Theta p(\theta)\frac{1}{2}\log_2\left(\frac{2\pi e}{J(\theta)}\right)d\theta \tag{2.12}$$

As the conditional entropy defined in Equation 2.9 is determined by the Fisher information, it may be either greater than or less than the stimulus entropy and consequently I_{Fisher} can either under- or over-estimate the mutual information (see Chapter 4).

This method of estimating the mutual information from the Fisher information was used by Gordon et al. (2008) in the analysis of binaural tuning properties in the inferior colliculus of the guinea pig. That study focussed on the comparison of mutual information, Fisher information and receiver operating characteristic (ROC) analysis in the analysis of experimental data from single neurons. A notable finding was that I_{Fisher} and mutual information for single neurons were approximately linearly related, but not equal.

Yarrow (2008) carried out a preliminary investigation into the relationship between MI and I_{Fisher} in finite populations, the results of which suggested a rapid convergence of the two measures and only small differences between them in

populations of the order of ten neurons. However, that study only modelled a very limited range of variability, using a maximum variability that was equivalent to a 100 ms integration time at a Fano factor of 1. Higher levels of variability corresponding particularly to short integration times, but also to supra-Poisson response distributions, are likely to be highly relevant to biological population codes. The same study also examined the effect of noise correlations with an exponentially decaying localised structure. However, problems with the simulations lead to erroneous results suggesting that MI and I_{Fisher} diverge in large populations with both correlated and independent variability. Chapter 4 includes an expanded and revised version of this analysis.

In summary, information theory provides us with measures that are very powerful, but which can be difficult to apply in practice. Other statistical measures, such as the Fisher information, are often easier to measure or calculate, but it is not always clear exactly what they tell us, or what the precise limits of their applicability are. I_{Fisher} goes some way towards bridging the gap between mutual information and Fisher information by allowing their absolute values to be compared, in the special case of an infinite population.

2.1.4 Chernoff distance

The Chernoff distance (Chernoff, 1952) is a measure of the difference between two probability distributions. The Chernoff distance between the response distributions $P(\boldsymbol{r}|\theta_1)$ and $P(\boldsymbol{r}|\theta_2)$ associated with two different stimuli θ_1 and θ_2 is defined as:

$$D_C(\theta_1,\theta_2) = \max_{\alpha\in(0,1)} \left[-\log \sum_{\boldsymbol{r}\in R} P^{\alpha}(\boldsymbol{r}|\theta_1) P^{1-\alpha}(\boldsymbol{r}|\theta_2) \right] \tag{2.13}$$

In the study of neural codes, it can be used to quantify the amount of overlap between the response distributions associated with two stimuli, and hence the ease with which the two stimuli can be discriminated. The Chernoff distance is linked to the mutual information between stimulus and response and also to the error rate in a two-alternative discrimination task (Kang and Sompolinsky, 2001). Specifically, if we consider a set of stimuli and their associated response distributions, the Chernoff distance between the two most similar response distributions predicts the rate with which the mutual information grows with the size of the population. The Chernoff distance is also related to the Fisher information, as it has been shown that they are approximately proportional for fine discrimination tasks (Cover and Thomas, 2006; Kang et al., 2004). Although the Chernoff distance has been used to quantify the precision of population codes as a function of the distance between stimuli in a discrimination task (Kang et al., 2004), it has not previously been used to predict best-encoded stimuli; in Chapter 5 we explore the latter application of the Chernoff distance. Computing the Chernoff distance for many parametric distributions is faster than computing the SSI, making it a potentially useful

method of determining best-encoded stimulus in two-alternative tasks, but it does involve iterative optimisation on α, which accounts for much of the computational effort. However, a recently described information geometric approach (Nielsen, 2013), which we make use of in Chapter 5, can greatly reduce the computational complexity of the optimisation.

2.1.5 Applications in neuroscience

Information measures can tell us about the precision of neural representations and, through careful selection of what is being measured, can also be used to address other questions about neural codes. The nature of the neural code—specifically, which aspects of cell and population activity are information bearing—is generally unknown. By comparing the coding precision of various response properties (e.g. firing rate, spike times or inter-spike intervals), information measures can be used to address this question. An example of this type of analysis is the work of Panzeri et al. (2001) on the representation of whisker stimuli in the barrel cortex of the rat. In this study, information theory was used to examine whether spike times conveyed information about spatial aspects of the stimulus by computing the time course of information accumulation following the stimulus presentation for both spike count and spike times. In this case, spike timing was found to contribute a significant amount of information beyond that carried by the spike count alone. More generally, the inherent temporal precision of a code can be found by perturbing the spike times by introducing progressively larger amounts of temporal noise, and noting how the precision of the code degrades as a function of the amount of jitter (Quian Quiroga and Panzeri, 2009).

Informational measures can also be used to examine which aspects of the stimulus are best encoded—most precisely represented—by a cell or population. In this case, the type of code (e.g. spike count versus spike timing) is fixed, and the amount of information transmitted about various stimulus properties is compared. Machens et al. (2005) used this approach to determine the optimal stimulus ensemble—the distribution of stimuli that maximised the information transmitted by a neuron—for grasshopper peripheral auditory neurons. The optimal stimulus ensemble was found to coincide with grasshopper communication sounds and not with natural sounds in general, indicating that the communication calls and auditory system were well matched. Panzeri et al. (2001) also used the stimulus-specific surprise to identify which whiskers are most precisely represented within a given barrel.

The relationship between neural precision and behavioural performance is a key area of neural coding research. In order to examine this relationship, it is necessary to ensure that both measures, neural and behavioural, are addressing equivalent questions. Fisher information is rather inflexible in this respect, as it only tells

us about the precision of fine discrimination or stimulus reconstruction, and not about coarser discrimination, classification or detection tasks. Information theoretic measures are more flexible as they can be tailored to suit a particular task by changing the stimulus ensemble (see Chapter 5). An alternative approach is to explicitly model a decoder that mimics the decision required by the task; in this case the performance of the decoder can be directly compared to behavioural performance. See Oram et al. (1998) and Quian Quiroga and Panzeri (2009) for reviews that cover this approach.

To illustrate the range of possible applications, we briefly review selected examples of the use of information theory and Fisher information to analyse rate population codes. For more detailed information on the use of information measures to analyse neural codes see reviews by Borst and Theunissen (1999), Sanger (2003), Averbeck et al. (2006), Nelken and Chechik (2007) and Quian Quiroga and Panzeri (2009).

2.1.5.1 Tuning curves

Tuning curves are one of the most intensely studied aspects of rate population coding. As well as having been used to analyse measured tuning curves (e.g. Theunissen and Miller, 1991), information measures have been used in theoretical studies examining the effect of tuning curve properties on the precision of the code. Much of this literature has concentrated on unimodal tuning curves and in particular the effects of tuning curve width (Paradiso, 1988; Seung and Sompolinsky, 1993; Zhang and Sejnowski, 1999; Pouget et al., 1999; Eurich and Wilke, 2000; Wilke and Eurich, 2002; Kang et al., 2004; Montemurro and Panzeri, 2006; Maler, 2009; Yaeli and Meir, 2010; Berens et al., 2011), baseline firing rate (Nakahara et al., 2001; Nakahara and Amari, 2002; Wilke and Eurich, 2002) and diversity of tuning (Wilke and Eurich, 2002; Ecker et al., 2011) on the precision of the code. Codes based on monotonic tuning curves have also received some attention and optimal sigmoidal tuning curves in terms of both Fisher information (Johnson and Ray, 2004) and mutual information (McDonnell and Stocks, 2008; Nikitin et al., 2009) have been studied. Many of these theoretical articles aim to identify optimal tuning curve parameters for idealised, abstract codes, while other studies (e.g. Theunissen and Miller, 1991) used experimental data to more tightly constrain their models in order to assess the optimality of specific population codes. Although several articles have addressed Fisher-optimal tuning curves (e.g. Zhang and Sejnowski, 1999; Eurich and Wilke, 2000), it is now becoming clear that the use of Fisher information for assessing optimality can be problematic (Yaeli and Meir, 2010; Berens et al., 2011).

As well as quantifying the effects of changes to single tuning curves or populations of stereotypical tuning curves, information measures can also be used to analyse the distribution of tuning curves, or variation in tuning curve properties, across

the space of possible stimuli (e.g. Eurich and Wilke, 2000). The optimal tiling of tuning curves or receptive fields in order to cover a given stimulus space is closely linked to the stimulus distribution. Harper and McAlpine (2004) conducted a theoretical study to determine the optimal (in terms of Fisher information i.e. specific to fine discrimination) frequency tuning for populations of auditory neurons selective for interaural time difference (ITD). The study predicted that cells that responded to frequencies below a certain species-specific threshold were more likely to respond maximally to ITDs that were outside the range that occurs in nature. This arrangement leads to the flanks of the tuning curves—the regions of maximum Fisher information—coinciding with the physiological range of ITDs, and was in agreement with experimental findings in small mammals. A similar study (Bonnasse-Gahot and Nadal, 2008), this time addressing the classification of stimuli drawn from overlapping categories, found that optimising the population in terms of overall precision (mutual information estimated from Fisher information, similar to I_{Fisher}) resulted in narrower tuning curves clustered at the boundaries between categories, with broader widely spaced tuning curves within categories. This meant that Fisher information, and hence discrimination power, was concentrated around category boundaries where small differences in the stimulus have the greatest importance in terms of determining its category.

2.1.5.2 Variability and noise correlations

The effect of variability and inter-neuronal correlations in trial-to-trial variability—noise correlations—on the precision of population codes has long been of interest to neuroscientists. Research using ROC analysis (e.g. Britten et al., 1992) of data recorded during a coarse visual discrimination task indicated that noise correlations have the potential to limit the information capacity of the code by defeating the ability of large populations to mitigate the effects of variability through redundancy. However, these studies were based on the assumption of averaging the responses of neurons with identical selectivities and did not address the information content of entire diversely-tuned populations. The effect of variability and noise correlations have subsequently been extensively studied using both Fisher (e.g. Abbott and Dayan, 1999; Wilke and Eurich, 2002) and mutual information (e.g. Panzeri et al., 1999; Sompolinsky et al., 2001), and it has been shown that they do not necessarily decrease the amount of information transmitted by a population code, and can even increase it. Small pairwise correlations can have a large effect at the population level (Zohary et al., 1994) and the structure of the correlations is a critical determinant of their effect on precision. Pairwise correlations that are uniform across the population actually increase the precision of the code, as they reduce the entropy of the trial-to-trial variability (Abbott and Dayan, 1999; Jenison, 2000). Noise correlations are detrimental to precision when they have a localised structure, i.e. when the variability of similarly tuned pairs

of neurons is more strongly correlated than that of disparately tuned pairs. These correlation structures result in stimulus-dependent covariances, but stimulus-dependent correlations have also been studied. Stimulus dependent shaping of correlations can, again, either increase or decrease information, and presence of localised stimulus-independent correlations is again critical (Josić et al., 2009).

2.1.5.3 Adaptation, attention and learning

Information measures can be used to accurately assess how precision changes when properties of the neural response change through learning or shorter-term modulatory processes such as attention and adaptation. Contrast adaptation in the visual system was studied by Durant et al. (2007) and Hietanen et al. (2007), using Fisher information to show that contrast discrimination thresholds were reduced at and above to the contrast of the adapting stimulus, at the expense of discriminability at lower contrasts. Seriès et al. (2009) used Fisher information together with simulated decoding to analyse the reconstruction precision and bias associated with a fixed decoding strategy in the presence of adaptation-like modulations in tuning properties. An analysis of sound intensity coding in the mammalian midbrain, using Fisher information, showed that monotonic intensity tuning curves adapted to changes in the stimulus statistics, allowing precision to be maintained across a wide stimulus dynamic range (Dean et al., 2005). Fisher information has also been used to measure how adaptive changes in noise correlations affected the precision of orientation representation by cells in macaque V1 (Gutnisky and Dragoi, 2008).

2.2 Quantifying topography

In general, neural maps are analysed by visual inspection and topographic organisation of maps is identified subjectively. However, this is not always the case; here we review research involving quantitative analysis of neural topographic maps. We then go on to introduce an example of a map with marginally topographic organisation, which we will analyse further in a subsequent chapter.

2.2.1 Map measures in neuroscience

Measures that quantify the local consistency of neuronal selectivity have been used to analyse experimental data; for example the Local Homogeneity Index (Nauhaus et al., 2008), and the Local Coherence Index (Li et al., 2008). These methods assign high scores where neighbouring cells are similarly tuned, but they do not quantify topography in the sense of a broader preservation of neighbourhood relations.

Kaschube and collaborators have published several quantitative analyses of maps in primary visual cortex (V1), using quantities such as ocular dominance column spacing (Kaschube et al., 2003) and orientation pinwheel density (Kaschube et al., 2010), but again these measures do not quantify topography as such. Polley et al. (2006) and Bandyopadhyay et al. (2010) illustrated tonotopic maps by producing scatter plots of characteristic frequency against position on the rostrocaudal axis of the primary auditory cortex (A1), and this approach was extended by Zheng (2012), who quantified maps by computing correlation coefficients (also between frequency and position on the rostrocaudal axis of A1). Zheng also computed average pairwise distances between nth-nearest neighbour cells with matching characteristic frequencies and used this measure in a bootstrap analysis to demonstrate statistical significance. Guo et al. (2012) used a vector averaging approach to make spatially resolved estimates of tonotopic map precision in several regions within the mouse auditory cortex, and used non-parametric statistical tests to compare tonotopy across regions and a number of different experimental conditions. Alvarez et al. (1998) defined measures of topographic organisation and lateral asymmetry for retinotopic maps, but these were based on differences from a predefined reference pattern; this is only a viable approach if such an ideal map can be defined. In cases where the dimensionality of the map matches the dimensionality of the space that it represents (such as in a retinotopic map) it is trivial to define an ideal mapping, but this is not the case where there is a difference in dimensionality between map and feature spaces: there is no unique ideal map. Willshaw (2006) measured the emergence of topography in a model of retinocollicular map development by quantifying receptive field size and overlap, making use of the fact that in a mature, ordered map, receptive fields tend to be local and less-overlapping. Willshaw et al. (2014) also went on to quantify topography in one-to-one retinocollicular maps by computing the size of the largest map subdomain within which neighbourhood relations were perfectly preserved. However, similarly to the approach taken by Alvarez et al. (1998), this method depends on the existence of a well-defined ideal mapping.

The literature on iterative map generation methods, such as self-organising feature maps, contains a wealth of information on quantifying the fidelity of topographic mappings (see reviews by Goodhill and Sejnowski, 1997; Bauer et al., 1999; Vidaurre and Muruzábal, 2007) and much of this is applicable to biological maps. These measures are explored further in Chapter 5.

2.2.2 An example of marginal topography: binaural and spatial maps in pallid bat A1

2.2.2.1 Maps in the mammalian auditory system

In contrast to the visual system, where only a single thalamic nucleus is interposed between the retina and primary sensory cortex, the auditory system relies upon multiple stages of sub-cortical processing. While frequency is encoded at the cochlea by an ordered array of transducers analogous to the retina, other auditory features must be extracted by further neural information processing. The tonotopic arrangement of the sensory epithelium is preserved through many subsequent stages of processing, so tonotopic maps are found in, e.g. the superior olivary complex (SOC), inferior (IC) and superior (SC) colliculi, medial geniculate nucleus (MGN), and auditory cortex.

Binaural interaction occurs initially in the SOC, and it is here that neuronal selectivity for IID and source azimuth emerges. Although azimuth maps have been observed in the external nucleus of the IC in some species (Binns et al., 1992), the most pronounced are found in the SC, where some form of azimuth map has been found in every species studied to date (see Cohen and Knudsen, 1999, for a review of spatial representations in the auditory system). The underlying population code in the SC is based on unimodal tuning curves, and the maps of space found in the SC are place maps i.e. the position of a sound source is reflected in the locus of the neural activity that it elicits within the SC. Perhaps surprisingly, given the arrangement that exists in the SC and the fact that auditory cortex is known to be important for spatial hearing (reviewed by King et al., 2007; Cohen and Knudsen, 1999), equivalent maps of space have not been reported in the auditory cortex, other than in the pallid bat.

2.2.2.2 Pallid bat auditory cortex

The pallid bat echolocates for general orientation and obstacle avoidance and listens to prey-generated noise to localise and hunt terrestrial insects (Bell, 1982). A1 of the pallid bat consists of two subregions, one that is specialised for the processing of frequency modulated echolocation calls and a second that responds to broadband, noisy stimuli. This second region is likely to be important for passive detection and localisation of prey. The passive hearing subregion of A1 in the pallid bat is further divided into at least two clusters of neurons based in IID selectivity: the 'peaked' cluster and the binaural inhibition (EI) cluster (following the nomenclature of Razak, 2011). The peaked cluster is made up of neurons that respond to sounds arriving with similar amplitude at both ears, and have bell-shaped azimuth tuning functions, while the EI cluster consists of neurons that

are excited by input from the contralateral ear and inhibited by the ipsilateral ear, which leads to sigmoidal azimuth tuning functions.

It has been suggested that topographic arrangements of IID and azimuth selectivity exist within the EI cluster in the pallid bat (Razak and Fuzessery, 2000, 2002; Razak, 2011). However, these maps were identified only subjectively, and the strength or clarity of the maps varies considerably between animals. Razak showed that the overall degree of activation of the EI cluster varied systematically with source azimuth (Figure 7 of Razak, 2011), but did not give any quantitative evidence for a systematic relationship between tuning properties and locations of neurons. Interestingly, the systematic maps in the EI cluster are not place maps like those identified in the SC; such a map is unlikely as many EI neurons do not have well defined receptive field maxima. Instead, it is the azimuth and IID at which the neurons become active—the sloping region of the tuning function (see Figure 3.3)—that is systematically arranged. This representation is consistent with those proposed in subcortical auditory lemniscal pathways (reviewed in McAlpine and Grothe, 2003). In Chapter 3, an analysis of data from the auditory cortex of the pallid bat was used as an example of how statistical measures can be used to quantify neural maps and detect significant topographic order. This analysis provides substantial additional evidence for a systematically ordered arrangement of spatial tuning in both the peaked and EI clusters.

Chapter 3

Anatomical Order in Population Codes: Quantifying Topography in Neural Maps

The work presented in this chapter was previously published as 'Detecting and Quantifying Topography in Neural Maps' in *PLOS One* (Yarrow et al., 2014).

The Matlab implementations of measures and statistical tests used in this chapter are available in the form of a toolbox from https://github.com/StuYarrow/MapTools.

This chapter examines seven map measures drawn from the map development modelling literature (Pearson distance correlation, Spearman distance correlation, Zrehen's measure, topographic product, topological correlation, wiring length and path length), with the aim of establishing an objective, quantitative method for comparing experimentally characterised maps and detecting statistically significant topography. We first assess the statistical power of each measure when applied to the detection of different types of map: linear gradient; convoluted, similar to maps of orientation in V1; and maps composed of randomly arranged homogeneous clusters. The results of these simulations showed that six of the measures were well suited to detecting topography and only one (wiring length) was less useful, due to low statistical power.

The chapter then goes on to illustrate the use of map measures to detect topography in experimental data. The recently identified systematic arrangement of azimuth selectivity (Razak, 2011) and corresponding binaural selectivity (interaural intensity difference; Razak and Fuzessery, 2002) in A1 of the pallid bat (*Antrozous pallidus*) is an example of a very small map that has been identified using single unit recordings. Because the map covers only a small area of cortex

($< 3\,\mathrm{mm}^2$) and because of the limited time available for making serial single-neuron recordings in each bat, Razak (2011) identified the map based on relatively few recordings (between 14 and 36 cells per bat). Here we quantify the topography in characteristic frequency, source azimuth and interaural intensity difference (IID) selectivity in A1 of the pallid bat. In addition to the well-known tonotopy, the results show that the arrangements of source azimuth and IID selectivity have significant topography at the scale of single binaural clusters in all eight bats studied. This analysis demonstrates the feasibility of objective quantification of topography and detection of statistically significant topography in experimentally characterised neural maps.

3.1 Methods

Before discussing measures of topography, it is useful to establish a formal definition of a map and define some terminology. Here the word 'map' refers only to the arrangement of neuronal properties in physical, anatomical space; no topography is implied. In order to observe a map, it is necessary to identify a number of anatomical elements, the nature of which depends on the experimental technique used. These can be individual neurons in the case of single-unit electrophysiology or multi-photon calcium imaging, local neuronal populations in the case of multi-unit recordings, or local haemodynamic response in the case of fMRI or intrinsic imaging. The units can be arranged in a regular grid (e.g. fMRI voxels) or scattered (e.g. single neurons). The positions of the units in map space (i.e. within the brain) are measured. Often these positions are 2-dimensional, as in cortical maps, but 3-D positions could also be used. Each unit is assigned a label based on its functional properties; typical examples are the characteristic frequency of auditory neurons or the preferred orientation of visual neurons. We refer to the space that the labels are defined within as the feature space. Both characteristic frequency and orientation feature spaces are 1-D, but 2-D (e.g. visual neuron receptive field centres) or higher dimensional spaces are possible. For concreteness, all the examples in this chapter involve a 2-D map space and a 1-D feature space (see Figure 3.1).

3.1.1 Map measures

A variety of map measures have been used to assess iterative models of topographic map development (see reviews by Goodhill and Sejnowski, 1996, 1997; Bauer et al., 1999; Vidaurre and Muruzábal, 2007). Many of these can be directly applied to experimentally measured maps, but some have inherent limitations that prevent this. Some map measures rely on the existence of a known training data set

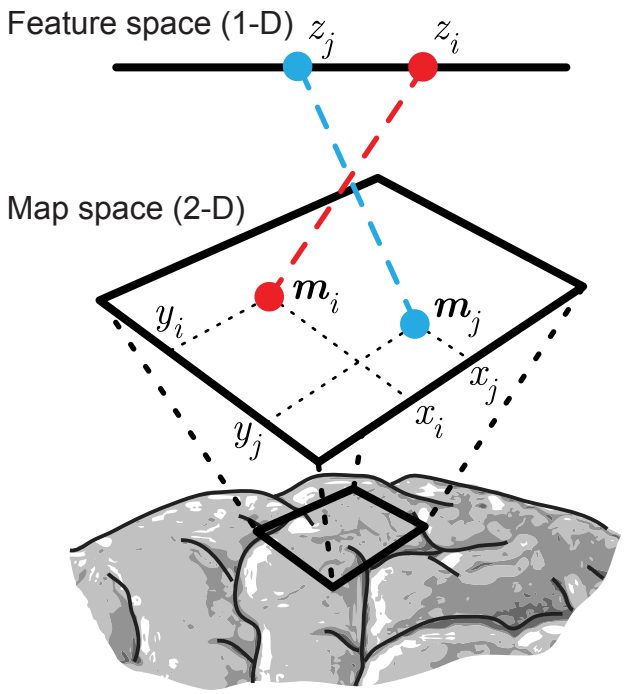

Figure 3.1: Elements of a topographic map. Fundamental elements of a map from a 1-dimensional feature space to a 2-dimensional map space. Dashed lines represent the link between the positions of neurons in feature space with their positions in map space. Here two neurons i and j are shown together with their map space (anatomical) and feature space (characteristic stimulus) coordinates.

from which the map is derived or learned (e.g. the measure proposed by Kaski and Lagus, 1996, and the topographic function of Villmann et al., 1997), some assume that neurons lie on a regular grid, and some are only applicable where the feature space has the same dimensionality as the map (the directional product measure of Vidaurre and Muruzábal, 2007 relies upon the latter two assumptions). For the purposes of this study, seven measures were selected, all of which can be calculated based on receptive field data alone, and are flexible with regard to the dimensionality of the feature space and the map space. The measures are also flexible with respect to the topologies of the map and feature spaces, but it is necessary to ensure that the distance measures are correctly adapted to the topology of the space. For example, applying a linear distance measure to angular variables will give nonsensical results. In this chapter we model planar map spaces only, but feature spaces with both line and ring topologies, so it was necessary to implement line and ring variants of the feature space metrics described in the following sections.

When defining a map, we assume that there are N units (neurons). The coordinates of the ith unit in map space (e.g. on the cortical sheet) are denoted $\boldsymbol{m}_i = (x_i, y_i)$, and the position in the one-dimensional feature space (e.g. the preferred stimulus) is denoted z_i.

Goodhill and Sejnowski (1997) described a mathematical framework that unifies a number of different measures of topography. The basis of this framework is the generic measure C:

$$C = \sum_{i=2}^{N} \sum_{j=1}^{i-1} F(i,j)G(i,j) \tag{3.1}$$

where F is a distance function in feature space and G is a distance function in map space. This form, the product of two corresponding pairwise distances summed over all possible pairings of neurons, is the basis for most of the measures described in this chapter.

3.1.1.1 Pearson distance correlation (PC)

The simplest variants of the C measure are based on Euclidean distances in both feature and map spaces, in this case:

$$F_E(i,j) = |z_j - z_i| \tag{3.2}$$

$$G_E(i,j) = ||\boldsymbol{m}_j - \boldsymbol{m}_i|| \tag{3.3}$$

It is useful to normalise the measure so that maps of different scales or with differing numbers of cells can be compared directly. This can be achieved by computing the Pearson correlation between pairwise distances in feature space and map space (Equation 3.4). This measure was mentioned by Bezdek and Pal (1995), but no results were reported. This measure is also related to the sample distance correlation proposed by Székely et al. (2007), but the latter measure uses centred distances (see reference for details). The Pearson distance correlation is given by:

$$C_{PC} = \frac{\sum_{i=2}^{N}\sum_{j=1}^{i-1}[F_E(i,j)-\bar{F}_E][G_E(i,j)-\bar{G}_E]}{\sqrt{\sum_{i=2}^{n}\sum_{j=1}^{i-1}[F_E(i,j)-\bar{F}_E]^2 \sum_{i=2}^{n}\sum_{j=1}^{i-1}[G_E(i,j)-\bar{G}_E]^2}} \tag{3.4}$$

Where $\bar{F}_E$ and $\bar{G}_E$ are the mean pairwise distances, for example:

$$\bar{F}_E = \frac{2}{N(N-1)} \sum_{i=2}^{N} \sum_{j=1}^{i-1} F_E(i,j) \tag{3.5}$$

With the Pearson correlation measure it is possible to combine data from different individuals as long as the scale of the map is consistent, as is often the case with subjects of the same age and species. With combined data, it does not make sense to compute distances in map space between cells from different subjects, as the coordinate systems may not be aligned and the map shape or orientation may be different. In this case we redefine C_{PC} as:

$$C_{PC} = \frac{\sum_{q=1}^{Q}\sum_{i=2}^{N_q}\sum_{j=1}^{i-1}[F_E(q,i,j)-\bar{F}_E][G_E(q,i,j)-\bar{G}_E]}{\sqrt{\sum_{q=1}^{Q}\sum_{i=2}^{N_q}\sum_{j=1}^{i-1}[F_E(q,i,j)-\bar{F}_E]^2 \sum_{q=1}^{Q}\sum_{i=2}^{N_q}\sum_{j=1}^{i-1}[G_E(q,i,j)-\bar{G}_E]^2}} \tag{3.6}$$

Where Q is the number of subjects, N_q is the number of neurons in the qth subject and the mean distances are also computed across all pairs in all subjects. The

revised distance functions are:

$$F_E(q,i,j) = |z_j^q - z_i^q| \tag{3.7}$$

$$G_E(q,i,j) = ||\boldsymbol{m}_j^q - \boldsymbol{m}_i^q|| \tag{3.8}$$

Here z_i^q and $\boldsymbol{m}_i^q$ denote the positions of the ith neuron from the qth subject.

3.1.1.2 Spearman distance correlation (SC)

As an alternative to the Pearson correlation, Bezdek and Pal (1995) used Spearman's rank correlation coefficient. This is sensitive to the ordering of data and not their absolute values, which means that the measure we denote as C_{SC} quantifies topology preservation and is not sensitive to distortion of the map unless it results in reordering of the neurons relative to their ordering in feature space. The Spearman distance correlation is given by:

$$C_{SC} = \frac{\sum_{i=2}^{N}\sum_{j=1}^{i-1}[f(i,j)-\bar{f}][g(i,j)-\bar{g}]}{\sqrt{\sum_{i=2}^{N}\sum_{j=1}^{i-1}[f(i,j)-\bar{f}]^2\sum_{i=2}^{N}\sum_{j=1}^{i-1}[g(i,j)-\bar{g}]^2}} \tag{3.9}$$

Where $f(i,j)$ and $g(i,j)$ are the tie-corrected ranks of $F_E(i,j)$ and $G_E(i,j)$ respectively, and $\bar{f}$, $\bar{g}$ are the mean ranks.

3.1.1.3 Topological correlation (TC)

The topological correlation (Doherty et al., 2006) is another closely related measure, but is based on graph theoretic rather than Euclidean distances. This makes it similar to C_{SC} in that it measures similarity of ordering rather than absolute position. To calculate the distances, it is necessary to construct Delaunay triangulations (see e.g. Lee and Schachter, 1980) in both map and feature spaces. The geodesic distance in map space $G_{graph}(i,j)$ between units i and j is the number of edges in the shortest path connecting them in the Delaunay triangulation. For the 1-D feature space the Delaunay triangulation is undefined, so rank difference is used instead:

$$F_{rank}(i,j) = |\zeta_j - \zeta_i| \tag{3.10}$$

Where ζ_i is the tie-corrected rank of z_i. The topological correlation C_{TC} is then defined as:

$$C_{TC} = \frac{\sum_{i=2}^{n}\sum_{j=1}^{i-1}[F_{rank}(i,j)-\bar{F}_{rank}][G_{graph}(i,j)-\bar{G}_{graph}]}{\sqrt{\sum_{i=2}^{n}\sum_{j=1}^{i-1}[F_{rank}(i,j)-\bar{F}_{rank}]^2\sum_{i=2}^{n}\sum_{j=1}^{i-1}[G_{graph}(i,j)-\bar{G}_{graph}]^2}} \tag{3.11}$$

Again, $\bar{F}_{rank}$ and $\bar{G}_{graph}$ are the mean distances over all pairs of cells.

3.1.1.4 Wiring length (WL)

The minimum wiring measure (C_{WL}) is designed to estimate the length of axonal 'wiring' required to connect all pairs of cells that are neighbours in feature space (e.g. are selective for neighbouring stimuli). This measure is a normalised version of the 'minimum wiring' objective function used by Goodhill and Sejnowski (1997). In this case the distance functions are defined as:

$$F_{neighbour}(i,j) = \begin{cases} 1 & : i,j \text{ neighbouring} \\ 0 & : \text{otherwise} \end{cases} \tag{3.12}$$

$$G_{E2}(i,j) = G_E(i,j)^2 = ||\boldsymbol{m}_j - \boldsymbol{m}_i||^2 \tag{3.13}$$

Neighbouring units are defined as those with identical or adjacent positions in feature space. C_{WL} is then defined as:

$$C_{WL} = \frac{N(N-1)}{2\sum_{i=2}^{N}\sum_{j=1}^{i-1} G_{E2}(i,j) \sum_{i=2}^{N}\sum_{j=1}^{i-1} F_{neighbour}(i,j)} \sum_{i=2}^{N}\sum_{j=1}^{i-1} F_{neighbour}(i,j)\, G_{E2}(i,j) \tag{3.14}$$

3.1.1.5 Path length (PL)

The path length measure C_{PL} is the same as wiring length, but the roles of map and feature spaces are reversed. The distance measures are:

$$F_{E2}(i,j) = |z_j - z_i|^2 \tag{3.15}$$

$$G_{neighbour}(i,j) = \begin{cases} 1 & : i,j \text{ neighbouring, i.e.} G_{graph}(i,j) = 1 \\ 0 & : \text{otherwise} \end{cases} \tag{3.16}$$

As the neurons are not located on a regular grid, as was the case when this measure was investigated by Goodhill and Sejnowski (1997), we define neighbouring in terms of the Delaunay triangulation. As for the wiring length, the path length measure is normalised such that it becomes independent of map size or measurement units. C_{PL} is defined as:

$$C_{PL} = \frac{N(N-1)}{2\sum_{i=2}^{N}\sum_{j=1}^{i-1} F_{E2}(i,j) \sum_{i=2}^{N}\sum_{j=1}^{i-1} G_{neighbour}(i,j)} \sum_{i=2}^{N}\sum_{j=1}^{i-1} F_{E2}(i,j)\, G_{neighbour}(i,j) \tag{3.17}$$

3.1.1.6 Zrehen measure (ZM)

This measure quantifies local consistency in maps and is a normalised version of the measure proposed by Zrehen (1993). It measures the separation in feature space of neurons that are neighbours in map space. Although originally applied to model neurons arranged in a regular grid, here we use the Delaunay triangulation

to determine which neurons are neighbours. The distance measures used are $G_{neighbour}$ (Equation 3.16) and a modified version of F_{rank} that counts the number of interposing 'intruders' in feature space between the neighbouring neurons:

$$F_{intruder}(i,j) = \begin{cases} 0 & : F_{rank}(i,j) \leq 1 \\ F_{rank}(i,j) - 1 & : \text{otherwise} \end{cases} \tag{3.18}$$

The measure C_Z is then defined as:

$$C_Z = \frac{1}{N \sum_{i=2}^{N} \sum_{j=1}^{i-1} G_{neighbour}(i,j)} \sum_{i=2}^{N} \sum_{j=1}^{i-1} F_{intruder}(i,j)\, G_{neighbour}(i,j) \tag{3.19}$$

3.1.1.7 Topographic product (TP)

The topographic product of Bauer and Pawelzik (1992) is the only measure examined in this chapter that does not fit into the C framework of Goodhill and Sejnowski (Equation 3.1). The topographic product P_T is a measure of the preservation of neighbour relations based on Euclidean distances. Bauer and Pawelzik first defined $n_k^G(i)$ as the index of the kth nearest neighbour of neuron i, in terms of distance in map space $G_E(i,j)$ (Equation 3.3), and $n_k^F(i)$ as the kth nearest neighbour of cell i in feature space (i.e. in terms of $F_E(i,j)$, Equation 3.2). They then defined the ratios:

$$Q_F(i,k) = \frac{F_E[i, n_k^G(i)]}{F_E[i, n_k^F(i)]} \tag{3.20}$$

$$Q_G(i,k) = \frac{G_E[i, n_k^G(i)]}{G_E[i, n_k^F(i)]} \tag{3.21}$$

The geometric mean over all neighbours within a given neighbourhood size k is given by:

$$P(i,k) = \left[\prod_{j=1}^{k} Q_F(i,j)\, Q_G(i,j) \right]^{1/2k} \tag{3.22}$$

If a map is perfectly ordered and all neighbourhood relations are preserved, then $P(i,k) = 1 \; \forall \, i,k$. The topographic product P_T is a measure of the deviation of P from 1 (by taking logarithms), averaged over all neurons and all possible neighbourhood sizes (Equation 3.23). The definition of P_T used here differs slightly from that of Bauer and Pawelzik in that we take the absolute value of $\log P(i,k)$ before averaging; this makes T_P non-negative and more suitable for use in a permutation test.

$$P_T = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{k=1}^{N-1} |\log P(i,k)| \tag{3.23}$$

A problem arises when two or more neurons have identical positions in either feature or map space, as this means that the order of neighbours is not always well defined. To resolve this issue, we use a Monte Carlo approach: the final value of T_P is taken to be the mean of 1000 samples in each of which the order of equidistant neighbours is randomly permuted.

3.1.2 Significance testing

For all of the measures described above, Monte Carlo permutation tests were used to calculate p-values. Taking the generic C measure as an example, the mth of M Monte Carlo samples is given by:

$$C_m = \sum_{i=2}^{N} \sum_{j=1}^{i-1} F^{shuf}(i,j) G(i,j) \tag{3.24}$$

$$\text{where } F^{shuf}(i,j) = F(r_i^m, r_j^m) \tag{3.25}$$

Where r^m is the mth randomly permuted instance of a vector containing the integers $\{1,\ldots,N\}$. In other words, for each sample the feature space positions were randomly shuffled and the measure computed using the shuffled values. We then compute M_{exceed}, the number of samples that are more ordered than the actual map:

$$M_{exceed} = \sum_{m=1}^{M} E_m \text{ where } E_m = \begin{cases} 1 & : C_m \text{ indicates greater order than C} \\ 0 & : \text{otherwise} \end{cases} \tag{3.26}$$

The p-value is given by:

$$p = \frac{M_{exceed} + 1}{M + 1} \tag{3.27}$$

All results in this chapter are based on Monte Carlo sample sizes of $M = 10^6$ unless it was faster to perform an exact permutation analysis (i.e. where $N \leq 9$). To control for multiple tests, the Benjamini–Hochberg step-up procedure (Benjamini and Yekutieli, 2001) was used to obtain corrected p-values. To test significance of the multiple-subject C_{PC} measure, the method above was modified so that feature space data were pooled across all subjects before shuffling.

3.1.3 Map models

To assess the sensitivity of the measures to different forms of topography, three generative map models were used (see Figure 3.2A–C). The map models were used to generate 501×501 arrays defining the ground truth tuning properties. This array was then sampled at N quasi-random points and noise was added to the samples. The map measures were then used to quantify the order in the noisy samples.

3.1.3.1 Linear map

The simplest map was a linear gradient intended to model maps with smooth large-scale structure (see Figure 3.2A). The linear map z_{lin} was defined as:

$$z_{lin}(x,y) = ax + by \tag{3.28}$$

Where a and b are drawn from a uniform distribution on the interval $[-1,1]$ and x, y are both in the interval $[0,1]$.

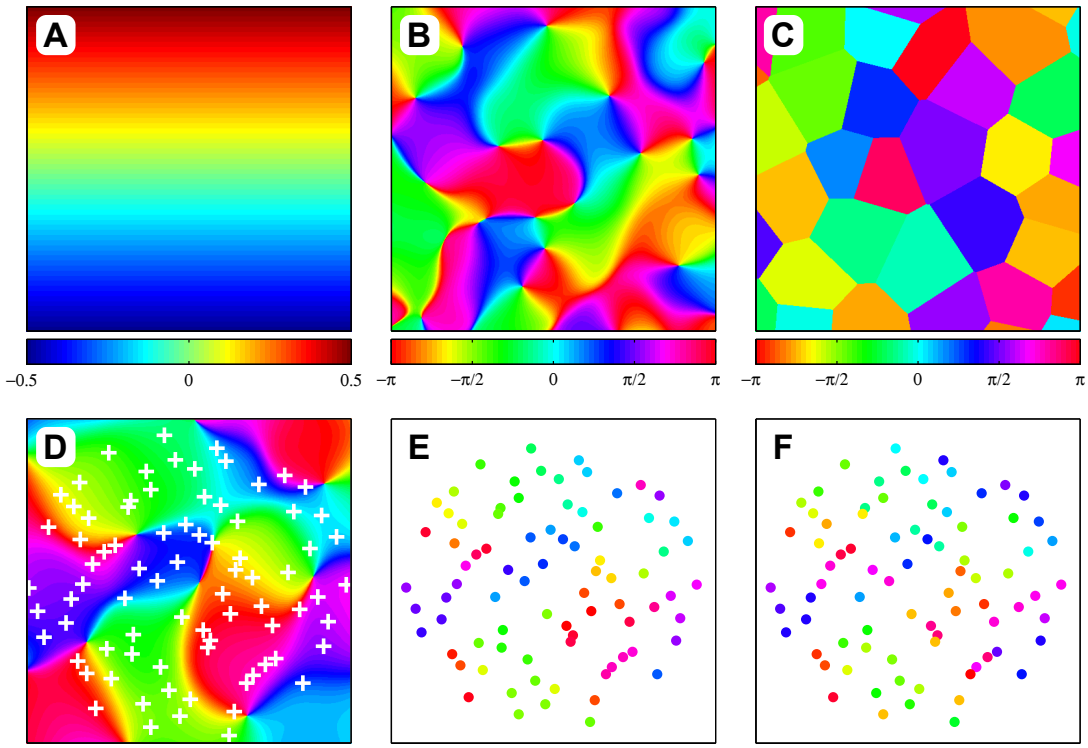

Figure 3.2: Map models and spatial sampling. Three map models were used to investigate the sensitivity of map measures to different forms of topography: (**A**) linear map, (**B**) angle map and (**C**) clusters (latter two with scale parameter $s = 0.2$). The sampling process is illustrated in the lower three panels: (**D**) raw angle map (scale parameter $s = 0.3$) with quasi-random sample locations marked (number of points $N = 80$), and sampled 'neurons' before (**E**) and after (**F**) noise was added (SNR = 3).

3.1.3.2 Angle map

To represent maps with a convoluted structure, such as visual cortex orientation maps (see Figure 3.2B), we used random angle maps derived from bandpass filtered white noise (Rojer and Schwartz, 1990; Macke et al., 2010). These maps are synthesised by generating 2-D arrays z_r and z_i of Gaussian white noise and convolving them with a 'Mexican hat' bandpass filter kernel (Equation 3.29). Treating the two arrays of filtered noise as the real and imaginary parts of an array of complex numbers, the angle map is found by taking the argument (Equation 3.30).

$$f_{mex}(x,y) = \frac{1}{2\pi s^2}\exp\left[-\frac{x^2+y^2}{2s^2}\right] - \frac{2}{\pi s^2}\exp\left[-\frac{2(x^2+y^2)}{s^2}\right] \tag{3.29}$$

$$z_{ang}(x,y) = \arg\left[(z_r * f_{mex}) + (z_i * f_{mex})i\right] \tag{3.30}$$

The filter scale parameter s determines the characteristic size of the aperiodic map features. As the feature space of this type of map is periodic, circular distance metrics and circular statistics (Berens, 2009) were used when computing all map measures of angle and cluster maps.

3.1.3.3 Clustered arrangement

This model was designed to test the sensitivity of measures to local consistency where there was no larger-scale topography. Clusters were generated by drawing $40^2 = 1600$ quasi-random seed points from a Halton sequence (Halton, 1964), and generating a Voronoi tessellation from these points. The seed point coordinates and tessellation were then rescaled by a factor of $40s$; this yields approximately equivalent scaling of angle and cluster maps for any given value of the scale parameter s. The z values for each seed point were then drawn from a uniform distribution and the pixel z values set to the z value of the nearest seed point, thus 'colouring' the Voronoi tessellation. Random variation of the tessellation was achieved by randomly setting the skip parameter (number of initial points in the Halton sequence to be discarded) when calling Matlab's `haltonset()` function.

3.1.3.4 Spatial sampling procedure

The process of measuring a biological map was modelled by quasi-random sampling of the maps and the addition of noise (see Figure 3.2D–F). The locations of the observation points were again drawn from a Halton sequence with a randomly chosen skip value. Points outside the unit disc were rejected to ensure that measure values were independent of map orientation. To simulate random neuronal variability and measurement error, Gaussian noise was added to the z values of each sample. The variance of the noise was defined in terms of the signal to noise ratio (SNR). For periodic feature spaces, the variance was computed using the CircStat toolbox (Berens, 2009).

$$z = z_{clean} + \eta \tag{3.31}$$

$$\text{where } \eta \sim \mathcal{N}(0, \frac{\sigma_z^2}{\text{SNR}^2}) \tag{3.32}$$

Where σ_z is the standard deviation of the feature space coordinate z across the whole map.

3.1.4 Pallid bat auditory cortical maps

The pallid bat echolocates for general orientation and obstacle avoidance and listens to prey-generated noise to localise and hunt terrestrial insects (Bell, 1982). A1 in the pallid bat consists of two subregions, one that is specialised for the processing of frequency modulated echolocation calls and a second that responds to broadband, noise-like sounds. This second region is likely to be important for passive detection and localisation of prey (Razak and Fuzessery, 2002). The passive hearing subregion is further divided into at least two clusters of neurons based on IID selectivity: the 'peaked' cluster and the binaural inhibition (EI) cluster

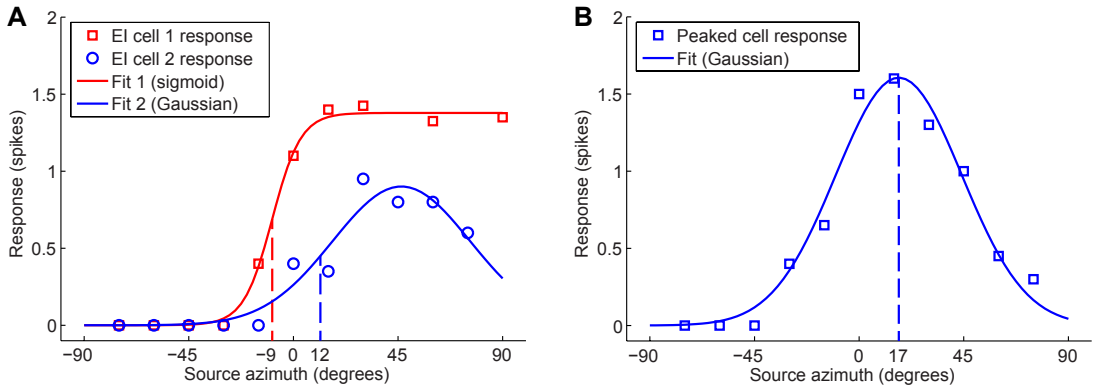

Figure 3.3: Tuning functions and characteristic stimuli. Examples of typical tuning functions of (**A**) EI cells and (**B**) peaked cells in pallid bat primary auditory cortex. Parametric tuning functions (solid lines) were fitted to the measured responses. EI neurons were assigned characteristic azimuth labels (indicated by dashed lines) where the fitted tuning function was equal to 50% of the maximum response. For Peaked neurons, the characteristic azimuth was defined as the peak of the fitted tuning function. IID tuning functions and characteristic stimuli were determined similarly.

(following the nomenclature of Razak, 2011). The peaked cluster is made up of neurons that respond to sounds arriving with similar amplitude at both ears, and have bell-shaped azimuth tuning functions, while the EI cluster consists of neurons that are excited by input from the contralateral ear and inhibited by the ipsilateral ear, which leads to sigmoidal azimuth tuning functions.

All data were collected as described by Razak (2011). In this chapter we analyse source azimuth, IID and frequency selectivity mapping data from four bats (corresponding to the maps shown in Figures 4–6 of Razak, 2011), together with source azimuth and frequency selectivity data from a further four bats. For the tonotopic maps, characteristic frequencies were determined as described by Razak (2011).

3.1.4.1 Azimuth labelling

To allow interpolation between the 15° azimuth spacing of the raw data, parametric tuning functions were fitted to the data (Figure 3.3). Two forms of parametric tuning curve were employed: Gaussian (Equation 3.33) and sigmoidal (Equation 3.34):

$$f_G(\psi) = A \exp\left[-\frac{(\psi-\psi_0)^2}{2\sigma_G^2}\right] \tag{3.33}$$

$$f_\varsigma(\psi) = \frac{A}{1+\exp\left(-\frac{\psi-\psi_T}{\sigma_\varsigma}\right)} \tag{3.34}$$

Where ψ is the source azimuth, ψ_0 and ψ_T are preferred azimuth (azimuth eliciting maximum response) and transition azimuth (azimuth of maximum gradient)

respectively, and A defines the maximum response (normalised spike count). The width parameters σ_G and σ_ς define the azimuth range over which the neuron responds in the Gaussian case, and the slope of the transition in the sigmoidal case respectively. Minimum values ($\sigma_G \geq 15^\circ$, $\sigma_\varsigma \geq 5^\circ$) were imposed on the width parameters to avoid over fitting.

Gaussian tuning curve models were used for peaked cells, while the better fitting of the two parametric models was used for EI cells. For peaked cells, the azimuth label is simply the azimuth eliciting maximum response ψ_0. For EI cells, the azimuth label ψ_{50} was defined as the ipsilateral (up-crossing) point where the tuning function is equal to 50% of its maximum. For sigmoidal tuning functions this is simply $\psi_{50} = \psi_T$. For EI cells with Gaussian fitted tuning functions ψ_{50} is given by:

$$\psi_{50} = \psi_0 - \sigma_G \sqrt{2 \log 2} \tag{3.35}$$

3.1.4.2 IID labelling

IID data was treated in a similar way to the source azimuth data. In this case the parametric tuning functions are:

$$f_G(\Delta I) = A \exp\left[-\frac{(\Delta I - \Delta I_0)^2}{2\sigma_G^2}\right] \tag{3.36}$$

$$f_\varsigma(\Delta I) = \frac{A}{1 + \exp\left(-\frac{\Delta I - \Delta I_T}{\sigma_\varsigma}\right)} \tag{3.37}$$

Where ΔI is the IID, and ΔI_0 and ΔI_T are the IIDs associated with maximum response (Gaussian) and maximum tuning function gradient (sigmoidal). The width parameters were constrained ($\sigma_G \geq 5$dB, $\sigma_\varsigma \geq 3$dB) to avoid over fitting. Again, the feature space labels for EI cells were defined as $\Delta I_{50} = \Delta I_T$ for sigmoidal and $\Delta I_{50} = \Delta I_0 - \sigma_G \sqrt{(2 \log 2)}$ for Gaussian tuning functions, and the labels for peaked cells were defined as ΔI_0, the IID corresponding to the maximum response.

3.2 Results

3.2.1 Statistical power of map measures for topography detection

We assessed the sensitivity of seven map measures by using them to quantify the topography in artificial 'electrophysiological' (i.e. spatially scattered) mapping data. The measures are: Pearson distance correlation (PC), Spearman distance correlation (SC), Zrehen measure (ZM), wiring length (WL), path length (PL), topographic product (TP) and topological correlation (TC); see section 3.1.1 for

definitions. Mapping data was generated by sampling an underlying map at spatially scattered locations, then adding random noise to the feature space coordinates of the samples (see Figure 3.2D–F). By varying the signal-to-noise ratio (SNR) and the number of points at which the map was sampled, we examined the relationship between the strength of the map (in terms of SNR) and the number of points needed for reliable topography detection using each measure. This corruption of the underlying map with random noise was intended to model departures of the measured map from an underlying trend; in real life this could be due to a local non-smoothness in the map, measurement error or some combination of these. We defined reliable detection as a statistical power of $\Pi = 0.8$ at a significance level of $\alpha = 0.05$ i.e. an 80% chance of correct detection, and the number of points needed to achieve this is denoted as N_{80}. The statistical power of the measures for a given map type, scale, SNR and number of points were estimated by analysing 400 randomly generated datasets: in each case 20 underlying maps were generated, each of these was spatially sampled 20 times and all measures, together with their p-values, were calculated for each sampled dataset. The statistical power Π was found by determining the proportion of trials in which the map was successfully detected ($p < \alpha$), and the standard error of this estimate was determined using the standard formula for the SD of the parameter of a binomial distribution: $\sqrt{\Pi(1-\Pi)/400}$. An iterative fitting procedure was used to approximate the function $\Pi(N)$ relating statistical power to the number of points, and hence to estimate N_{80}. The function $\Pi(N)$ was assumed to be sigmoidal with $\Pi(1) = \alpha$ and $\Pi(\infty) = 1$. This was then iteratively refined by estimating the power for a value of N close to the latest estimate of N_{80} (as described previously), then re-fitting. The standard error of the estimate of N_{80} was determined by Monte Carlo simulation.

Additional simulations (results not shown) showed that the findings described in this chapter are robust with respect to small changes in either significance level α or the detection threshold statistical power Π. N_{best} is the N_{80} of the most powerful measure for a given SNR and map type. Clearly, increasing the sample size increases the statistical power of any test and increasing the density of measurement points increases the ability to detect patterns at smaller scales. Here we focus on comparing the statistical powers of the map measures to identify which measures are most powerful and hence are likely to be most useful for detecting topographic organisation in experimental datasets.

3.2.1.1 Linear maps

Figure 3.4A shows the relationship between SNR and the number of points N_{80} required for reliable detection when the underlying map is a simple linear gradient (as in Figure 3.2A). The more powerful a measure is, the lower its line appears

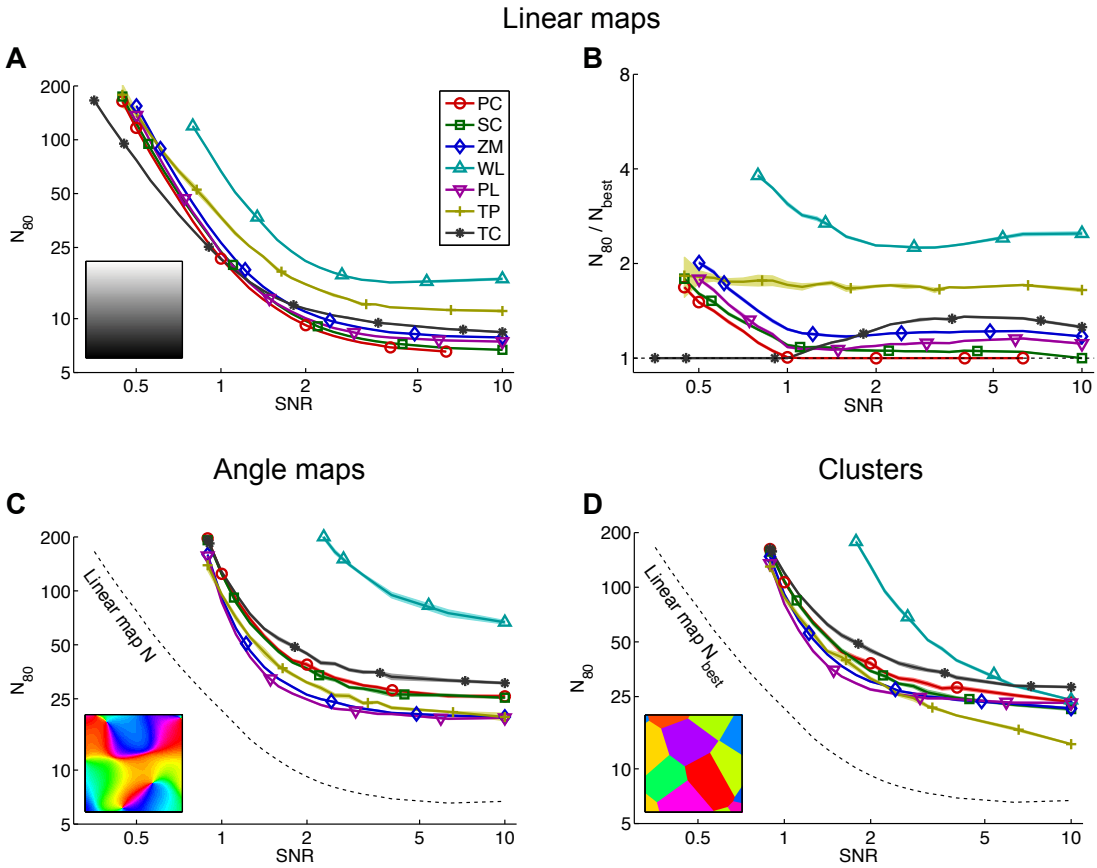

Figure 3.4: Comparison of the statistical power of seven map measures.
Statistical power of seven measures (PC: Pearson distance correlation, SC: Spearman distance correlation, ZM: Zrehen measure, WL: wiring length, PL: path length, TP: topographic product, TC: topological correlation) when detecting (**A**) linear maps, (**C**) angle maps and (**D**) clusters. Power is summarised by the quantity N_{80}, the mean number of points (e.g. neurons, voxels) required to achieve a statistical power of 80%; this is shown as a function of the SNR. (**B**) shows the relative powers of the measures for linear map detection; here N_{80} is normalised by N_{best}, the N_{80} of the most powerful measure for a given map type and SNR. For the angle maps and clusters the scale parameter $s = 0.4$ and the insets show examples of the corresponding map type and scale. All axes have logarithmic scales. Missing data indicate that N_{80} is outside the range $7 \leq N \leq 200$. Uncertainty is depicted by shaded regions of ± 1 StdErr.

on the plot; the most powerful measure at any given SNR is that which achieves reliable detection with the least data and hence is the lowest line on the plot. For linear maps, the topological correlation (TC) is the most powerful measure for detecting maps with weak topography (that is, maps heavily corrupted with noise: SNR < 0.8), while the Pearson distance correlation (PC) is the most powerful measure for maps with strong topography (SNR > 0.8). Four of the measures (PC, SC, ZM and PL) have similar power across the SNR range. WL and TP are consistently less powerful, requiring 1.6 to 4 times as many data as the most powerful measure at any given SNR to achieve the same statistical power

(Figure 3.4B).

3.2.1.2 Nonlinear maps

Neural maps generally have structures more complex than a linear gradient. We next assessed the statistical power of the same seven measures for detecting two forms of nonlinear map: convoluted angle maps similar to V1 orientation maps (Figure 3.4C), and clustered arrangements where there is no overall topography, but tuning properties are locally homogeneous (Figure 3.4D). The nonlinear maps are only locally consistent, so higher sampling densities are required in order to detect the map. When used for the detection of these nonlinear maps, the statistical power of the measures depends primarily upon the spatial scale of the map, as well as the SNR and number of points; the effect of map form (angle map versus clusters) is relatively minor (Figure 3.5). The parameter s controls the spatial scale of the model maps and hence the density of measurements required to resolve the map. Maps with larger features (greater s) can be detected with fewer measurements than smaller-scale maps, as can be seen in Figure 3.6. The most powerful measures for detecting smaller-scale nonlinear maps ($s = 0.4$) are the path length and topographic product (Figure 3.5A and B), despite the fact that the topographic product is one of the least powerful measures for detecting linear maps (see Figure 3.4B). For detecting both angle maps and clusters at larger scales ($s \geq 0.6$), the most powerful measures are the Pearson and Spearman distance correlations, path length, and the topographic product for cluster maps at very high SNR (Figure 3.5C–F). The Zrehen measure also has relatively high power for detecting nonlinear maps, particularly at smaller scales, but it is never the most powerful measure.

3.2.1.3 Summary

To summarise, the relative power of the measures varies according to both the type (mainly linear versus nonlinear) and scale of the map, as well as the SNR. The conventional correlative measures (Pearson and Spearman distance correlations) are the most powerful for detecting large scale topography i.e. linear maps and larger-scale nonlinear maps (approximately $s > 0.6$). An angle map on an infinitely large scale is equivalent to a linear map, so it is not surprising that the same measures are most effective at detecting linear and large-scale nonlinear maps. The path length and topographic product are the most powerful for detecting the localised topography in smaller-scale nonlinear maps ($s < 0.6$).

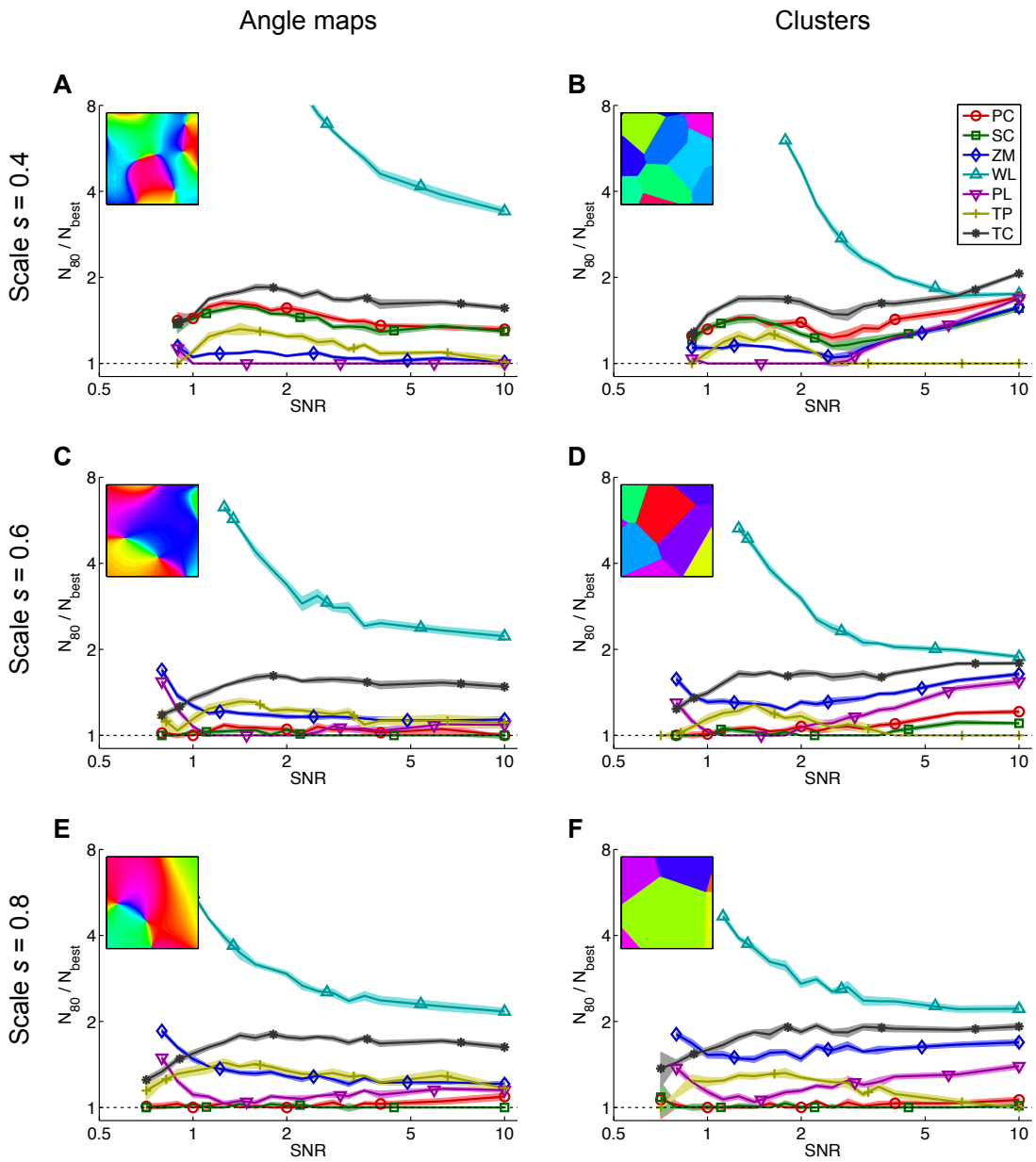

Figure 3.5: Relative power of measures for detecting maps of various scales and types. For each map measure, the plots show the number of data needed for reliable detection of angle maps (**A**),(**C**),(**E**) and clusters (**B**),(**D**),(**F**). To show the relative power more clearly, N_{80} is normalised by N_{best}, the N_{80} of the most powerful measure for a given map type and SNR. The more powerful the measure, the lower it appears on the plots. It can be seen that the map type i.e. angle map versus clusters, has little effect upon the relative powers of the measures; the ordering of the measures in terms of power is similar for both forms of map. All axes have logarithmic scales. Missing data indicate that N_{80} is outside the range $7 \leq N \leq 200$. Uncertainty is depicted by shaded regions of ± 1 StdErr.

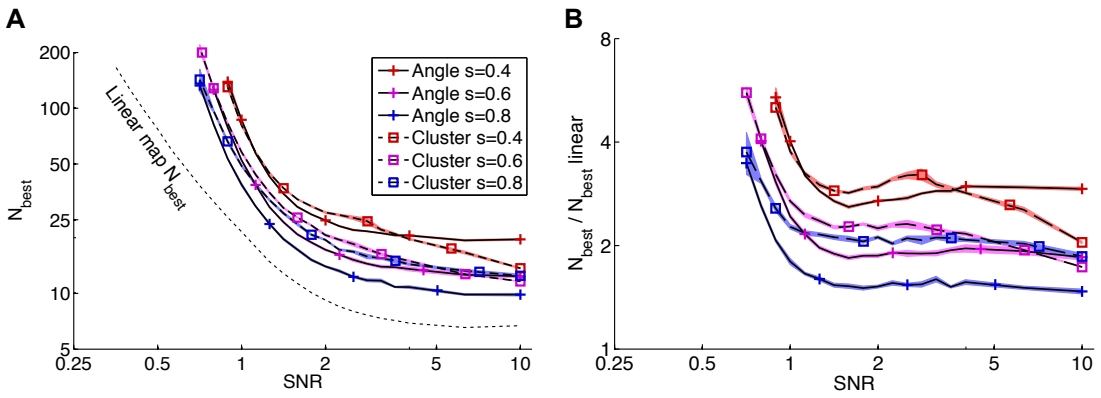

Figure 3.6: The effect of map scale on nonlinear map detection. Larger scale maps can be detected with fewer data. (**A**) shows N_{best} (N_{80} of the most powerful measure) for angle maps and clusters at three different scales: $s = \{0.4, 0.6, 0.8\}$. (**B**) shows the detectability of each type and scale of nonlinear map relative to a linear map with the same SNR i.e. N_{best} normalised by N_{best} for a linear map. All axes have logarithmic scales. Missing data indicate that N_{80} is outside the range $7 \leq N \leq 200$. Uncertainty is depicted by shaded regions of ± 1 StdErr.

3.2.2 Pallid bat A1 maps

To illustrate how statistical tests can be used to objectively determine the existence of neural topographic maps, we quantified the topography of three different neuronal tuning properties in pallid bat A1. Using data gathered from 211 cells in the EI and peaked clusters of eight bats (see section 3.1.4 and Razak, 2011), we tested for the existence of significant maps of frequency (tonotopy), IID and source azimuth. IID and azimuth labels corresponding to steeply sloping regions of the tuning curves, rather than maxima, were chosen because cells in the EI cluster generally have sigmoid-like tuning curves without clearly defined maxima. This means that the IID and azimuth maps in the EI cluster differ from typical place maps (as in e.g. the superior colliculus) where the locus of activity directly reflects the value of the stimulus variable. In the peaked cluster, the azimuth labels were located at the peaks of the tuning curves. The pallid bat data was analysed directly; there was no sub-sampling or other preprocessing of the data aside from determining the characteristic stimulus labels (see section 3.1.4).

3.2.2.1 Tonotopy

Significant topographic maps of characteristic frequency were detected in all 8 bats (Table 3.1). In all 8 bats significant tonotopy was also detected when the EI clusters were considered in isolation. Tonotopy was also significant within the peaked cluster in 3 of 4 bats for which data from the peaked cluster was available. These results are consistent with the tonotopic arrangement of auditory cortex found in

Measure	Frequency			IID		Azimuth		All
	EI	Peaked	EI+Peaked	EI	Peaked	EI	Peaked	
PC	5/8	2/4	3/4	2/4	0/4	5/8	1/4	52.5%
SC	6/8	2/4	4/4	2/4	0/4	5/8	0/4	55.0%
ZM	5/8	2/4	4/4	3/4	0/4	7/8	1/4	62.5%
WL	6/8	2/4	3/4	1/4	0/4	2/8	0/4	37.5%
PL	6/8	2/4	4/4	3/4	0/4	6/8	3/4	70.0%
TP	8/8	3/4	4/4	3/4	0/4	4/8	0/4	62.5%
TC	5/8	2/4	4/4	3/4	0/4	5/8	0/4	57.5%
Any measure	8/8	3/4	4/4	3/4	0/4	8/8	3/4	

Table 3.1: Proportion of bats with significant tonotopic, IID and azimuth maps. Results of map detection analysis of pallid bat data. Each table cells shows the number of bats in which significant maps were detected / total number of bats from which data were available. The 'any measure' row shows the number of bats where significant topography was detected by at least one measure. Each column relates to a given tuning property (e.g. frequency) and group of neurons (e.g. cells from the EI cluster). The 'all cell groups' column gives combined detection rates for each measure across all maps in all bats; this is a coarse indication of the relative power of the measures.

many species.

3.2.2.2 IID and source azimuth maps

Systematic cortical maps of IID and source azimuth selectivity are present within the EI cluster in the pallid bat (Razak and Fuzessery, 2000; Razak, 2011). Figure 3.7 shows an example of an azimuth map and gives the corresponding measure values. These results confirm the presence of a topographic arrangement of IID and azimuth tuning within the EI cluster. Significant topography in IID maps was detected in the EI cluster in 3 of 4 bats for which IID data were available, but no significant IID topography was found in the peaked cluster (Table 3.1). Significant azimuth maps in EI were detected in all eight bats, but in the peaked cluster topography was much weaker (as in Razak, 2011), being marginally significant ($0.02 < p < 0.05$ after Benjamini–Hochberg correction) in 3 of 4 animals from which data were available (Table 3.1).

3.2.2.3 Population analysis

The Pearson distance correlation measure allows us to combine data from multiple animals into a single statistic to assess the strength of topography across the population (see section 3.1.1.1). This population analysis provides additional confirmation of highly significant tonotopy and highly significant topographic

	Frequency map			IID map		Azimuth map	
	EI	Peaked	EI+Peaked	EI	Peaked	EI	Peaked
C_{PC}	0.27	0.52	0.34	0.25	-0.019	0.30	0.33
p	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$	0.0017	3.9	$< 10^{-4}$	0.0024
n	156	49	205	71	42	156	49

Table 3.2: Map detection analysis of combined map data from all 8 bats. Results of the analysis of combined data from all bats. Columns indicate the tuning property (e.g. characteristic frequency) and cell class (e.g. EI). Pearson distance correlation C_{PC}, p-value (Bonferroni corrected, 7 tests) and number of neurons n are given for each candidate map i.e. combination of tuning property and cell class. The Bonferroni method of correcting for multiple tests can lead to corrected p-values greater than 1, as is the case for IID tuning in the peaked cluster.

arrangement of IID and azimuth selectivity (see Table 3.2). Although the azimuth map is only marginally significant in 3 of 4 animals for which peaked cluster data is available, it is clearly significant ($p = 0.0024$) when the data from the four bats are combined. This analysis is also useful because the Benjamini–Hochberg procedure used to correct for multiple tests in the individual analysis is not very conservative and can, at best, be expected to give a false discovery rate of 0.05, equivalent to approximately 8 tests wrongly identified as significant. Combining the data into a single statistical test, or a much smaller number of tests, avoids the difficulties associated with correcting for large numbers of comparisons.

3.2.2.4 IID and azimuth maps are independent of tonotopy

One possible explanation for the topographic arrangements of IID and azimuth selectivity is that they are somehow a consequence of tonotopy. To test this hypothesis we calculated the correlation between characteristic frequency and 50% IID, and between characteristic frequency and 50% azimuth. In the EI cluster, only one significant correlation was found, between characteristic frequency and azimuth in bat PAL28 (Pearson $\rho = 0.67$, $p = 0.0006$). There was no significant correlation between characteristic frequency and azimuth in the EI cluster when data from all 8 animals was combined, or between IID and characteristic frequency in any bat or across all bats. In the peaked cluster there was no significant correlation between azimuth and characteristic frequency. There is, however, a significant negative correlation between IID and characteristic frequency in the peaked cluster at the population level (Pearson $\rho = -0.31$, $p = 0.035$). Interestingly, this correlation does not result in the significant tonotopy also manifesting as a significant IID map (see Table 3.2). In summary, the topographic arrangements of azimuth and IID selectivity do not appear to be a consequence of tonotopy.

3.2.2.5 Can IID and azimuth maps be explained by a linear gradient?

It is useful to compare the pallid bat data to the map models discussed in the previous section. Both the linear and cluster map models are plausible candidates for the underlying form of the pallid bat azimuth and IID maps (tonotopy is locally and approximately linear). If it is assumed that only frontal space is represented, or that frontal space predominates, then both the azimuth and IID feature spaces are non-periodic and the space maps could be linear (perhaps oriented near-perpendicular to the tonotopic gradient). Alternatively, the space map could take the form of clusters as this is a known organisational principle of A1 (see e.g. Cohen and Knudsen, 1999). One approach to resolving this question is to fit the models to the experimental data. Both the cluster and angle map models are under-constrained by the data; for any possible set of mapping data, there are an infinite number of possible angle or cluster maps that would explain the data perfectly. Fitting the linear model, however, was straightforward and allowed us to estimate the SNR of the bat data based on the assumption of an underlying linear map. To do this, we fitted a bilinear function that predicts the selectivity feature z for a given location on the cortex defined by x and y. The SNR of the data was then estimated by calculating the proportion of the standard deviation of z that was explained by the bilinear fit. The best frequency maps had estimated SNRs between (approximately) 0.9 and 2, while the IID and azimuths maps had estimated SNRs between 0.4 and 2. The estimated SNR indicates how well the pallid bat mapping data is explained by a linear model, and the broad range of observed SNRs suggests that the linearity of the maps varies considerably between bats. While the results of the permutation tests show that azimuth and IID tuning is organised non-randomly, it is not possible to say conclusively what form the azimuth and IID maps take; this question can only be addressed by further mapping using a technique with higher spatial resolution.

3.3 Conclusion

We have shown that topography in the anatomical layout of neuronal tuning properties can be quantified using measures that do not rely on any prior knowledge about the form of the map. These measures can be used to perform statistical tests for the existence of significant topography. This provides an objective method for detecting topographic maps that are unclear, for instance where data are available from only a small number of neurons, or the scale of map features are close to the spatial resolution of the measurement technique.

We assessed the sensitivity of seven measures (Pearson distance correlation, Spearman distance correlation, Zrehen measure, wiring length, path length, topographic product and topological correlation; see section 3.1.1 for definitions)

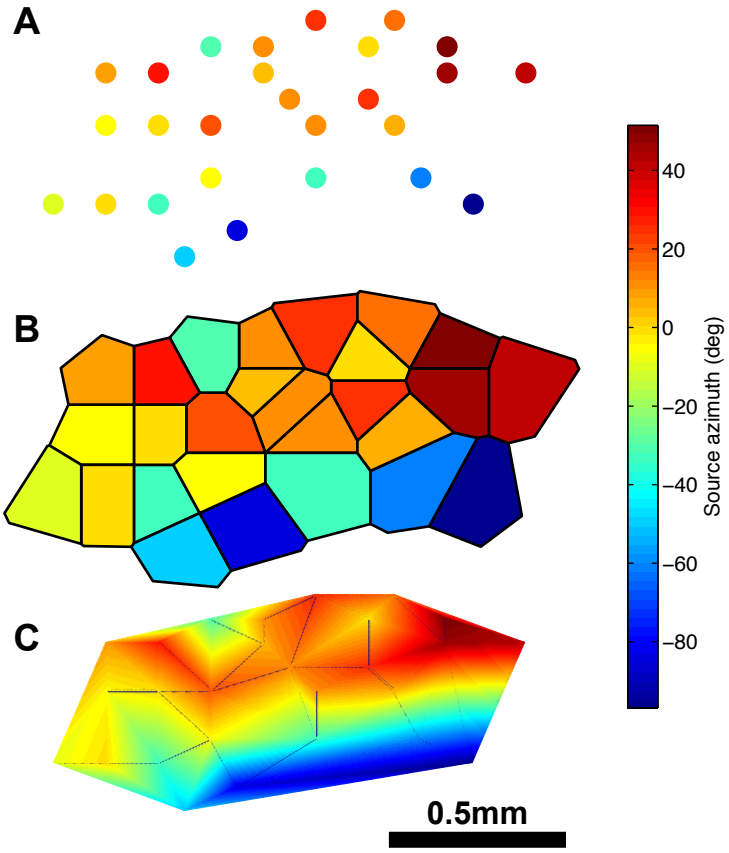

Figure 3.7: Pallid bat azimuth map example.
The perceived orderliness of a topographic map can vary depending on how the data is presented. This figure shows identical mapping data (an azimuth map in pallid bat A1) plotted in three ways: (**A**) no interpolation, (**B**) Voronoi tessellation (nearest neighbour interpolation), and (**C**) linear interpolation (as Figure 1.2). The method used in (**A**) might be considered advantageous as its lack of interpolation avoids any implicit assumptions about the properties of neurons between the measured locations. Analysis results for this map (measure values and Benjamini–Hochberg corrected p–values; see Results): $C_{PC} = 0.25, p = 0.013$; $C_{SC} = 0.27, p = 0.0039$; $C_Z = 0.23, p = 0.0039$; $C_{WL} = 0.88, p = 0.28$; $C_{PL} = 0.59, p = 0.002$; $P_T = 0.2, p = 0.032$; $C_{TC} = 0.22, p = 0.0053$.

to linear and nonlinear model maps obscured by adding noise to the characteristic stimulus values. Sensitivity was quantified by calculating the statistical power, for map detection, of permutation tests based on each measure. The sensitivity of measures depended on the form (in particular linear versus nonlinear) and scale of smoothness in the map, and on the SNR of the characteristic stimulus labels; no one measure was the best at detecting all maps. For detecting linear maps the Pearson and Spearman distance correlations and the topological correlation were the most powerful. For larger-scale nonlinear maps the Pearson and Spearman distance correlations are among the most powerful, while the path length and topographic product have more power when detecting smaller-scale nonlinear maps. It is unsurprising that more data are required to detect smaller scale maps. The smaller the scale of the map features, the smaller the spatial extent of local smoothness in the map and therefore the greater the spatial density of samples needed to detect that local trend. This is related to the Nyquist–Shannon sampling theorem: smaller-scale maps contain higher spatial frequencies, and greater sampling densities are therefore needed to resolve the map.

The SNR parameter in the models discussed in this chapter is intended to model all departures of the map data from idealised map forms, including local variability of tuning properties and measurement errors. Although it may be possible to estimate the SNR of measured map data as we have done in section 3.2.2.5, such estimates require some prior assumptions about the form of the map, so it is doubtful that this approach would be useful for determining the number of recording sites required in a map detection experiment.

These results confirm the presence of topographic organisation of spatial (azimuth) and binaural (IID) selectivity in pallid bat A1. It has been suggested previously that topographic arrangements of IID and azimuth selectivity exist within the EI cluster (Razak and Fuzessery, 2000, 2002; Razak, 2011), however these maps were identified only subjectively, and the strength or clarity of the maps varies considerably between bats. Razak showed that the overall level of activation of the EI cluster varied systematically with source azimuth (Figure 7 of Razak, 2011), but did not give any quantitative evidence for a systematic relationship between tuning properties and locations of neurons. The identification of topographic arrangements of IID and source azimuth selectivity in the pallid bat raises the question as to whether similar maps are also present in other species.

One feature of the pallid bat data that is revealed in the results of this analysis is the variability in the apparent orderliness of the maps between animals; in some bats highly significant topography is detected by many measures, but in a few cases the map measures show only weak topography. The results of the population analysis (Table 3.2) show that there is significant topography in azimuth tuning in the the EI (Pearson distance correlation $= 0.30$, $p < 10^{-4}$, $n = 156$) and peaked (Pearson distance correlation $= 0.33$, $p = 0.0024$, $n = 49$) clusters, and in

IID tuning in the EI cluster (Pearson distance correlation $= 0.25$, $p = 0.0017$, $n = 71$) when the data from all bats is considered together. There are a number of possible reasons for the observed differences between bats: random sampling variability, measurement error in recording and extracting tuning curves and characteristic stimuli, individual differences in the strength of the map or in the form of the map (e.g. warped or fractured maps).

3.3.1 Summary and recommendations

Topography in neural maps can be objectively quantified using measures that compare the pairwise anatomical (map space) distances between neurons with the pairwise distances in some feature space, for example the difference in preferred stimulus. Correlation between these two distances indicates a tendency toward topographic arrangement of the feature. By applying a permutation test, these measures can be used to determine whether a suspected neural topographic map is statistically significant; this is valuable where the topography is weak or unclear, the measurements are noisy, the number of data is limited, or the characteristic scale of map features is close to the spatial resolution of the measurement technique. The way in which map space and feature space distances are quantified determines the type of map that the measure is most sensitive to. Some measures (particularly the Pearson and Spearman distance correlation) are more effective, relative to other measures, at detecting the large-scale smoothness found in linear or larger-scale nonlinear maps than they are at detecting localised topography in smaller-scale nonlinear maps (see Figure 3.5). The opposite is true for other measures, particularly the topographic product. The wiring length and, to a lesser extent, topological correlation measures had relatively low statistical power for map detection in general.

The approach used to test for the presence of significant topography might be guided by the investigator's prior knowledge about the form of the map. If the map is thought to be linear (e.g. a tonotopic or retinotopic map), or convoluted on a scale where map features are many times larger than the distance between recording sites, either the Pearson or Spearman distance correlation would be a good choice of measure. For nonlinear maps with smaller features, the topographic product or path length are likely to be a good choice. If the form of the map is unknown, more than one measure might be used (e.g. the Pearson or Spearman distance correlation together with the topographic product) and a suitable method used to correct for multiple tests. If the Pearson distance correlation is used, data from multiple subjects can be combined in a single permutation test for detecting topography, without the need for registering or otherwise preprocessing the data. This approach offers the possibility of detecting maps at the population level when the topography is too weak, or insufficient data are available to support detection

in individual subjects.

Chapter 4

Fisher and Shannon Information in Finite Neural Populations

The work presented in this chapter was previously published as 'Fisher and Shannon Information in Finite Neural Populations' in *Neural Computation* (Yarrow et al., 2012).

The Matlab code used to obtain the results described in this chapter is available for download from ModelDB:
http://senselab.med.yale.edu/modeldb/ShowModel.asp?model=142990.

This chapter describes theoretical modelling work investigating the properties of measures of information used to analyse sensory population coding; in particular, the Fisher information and the stimulus-specific information. Both Fisher information and information theoretic measures, including the SSI, are now widely used for the study of neural codes. These tools are no longer the preserve of theorists, and are being applied by experimental research groups in the analysis of empirical data. Fisher information is a particularly accessible tool for experimentalists, as it is generally easier to calculate than information theoretic measures, in terms of both data requirements and computational complexity.

While both families of measure are widely used, previous studies have almost invariably made use of either Fisher information (when measuring whole populations) or information theory (for studying single neurons). This leads to difficulties in comparing the findings of studies based on different measures, since they are rarely applied to the same cases. Are the two families of measure interchangeable; do they ultimately yield the same insights as to which stimuli are best encoded? The answer to this question is: sometimes (Butts and Goldman, 2006), but to date this issue has only been examined for very small populations (number of neurons $N = 4$). For most biologically relevant population codes, the

relationship between Fisher and Shannon information is unclear. Resolving this ambiguity is of crucial importance to bridge the gap between the Fisher information and information theoretic strands of the literature.

In this chapter we employ numerical models of simplified, but broadly biologically realistic, populations to clarify the link between Fisher and Shannon information. We also examine in detail the limits of applicability of Fisher information. How many neurons are required before I_{Fisher} provides a good working estimate of the mutual information (MI)? How does Fisher information relate to information theoretic measures? We go on to show, through numerical simulation, that Fisher information can be used to obtain the asymptotic value of SSI, in the same way that it can provide the asymptotic value of the MI.

4.1 Methods

We consider a population of N sensory neurons encoding a unidimensional circular stimulus variable θ, which represents a direction e.g. of a moving bar. Each experiment consists of a number of virtual trials, in which the spike count r_i of each neuron over a time interval τ is computed. Each presentation of a stimulus θ is therefore associated with a response vector $\boldsymbol{r} = [r_1 \dots r_N]$. For the purposes of this study, information is assumed to be encoded exclusively by the spike counts; the timing of individual spikes within the measurement window is disregarded. Although this represents a simplification, the rate coding model is frequently employed for its tractability and has been shown to be valid in a number of contexts (Heller et al., 1995; Tovée et al., 1993).

The response of each neuron can be represented by a deterministic component (the tuning curve) that defines the mean response over many trials, and a random component that models the trial-to-trial variability or noise; these are described in the following sections. The model framework described below (sections 4.1.1–4.1.2) was used in all experiments described in this chapter, except for those based on the cricket cercal interneuron model described by Theunissen and Miller (1991), which is examined in section 4.1.4.

4.1.1 Tuning curves

The mean firing rates of each neuron were modelled by a circular Gaussian function, given here for the ith neuron:

$$f_i(\theta) = f_{bg} + f_{mod} \exp\left[-\frac{1-\cos(\theta-\phi_i)}{(\frac{\pi}{180}\omega)^2}\right] \tag{4.1}$$

Where f_{mod} and f_{bg} are the stimulus-dependent firing rate modulation depth and stimulus-independent background firing rate respectively, both measured in

spikes/s; ϕ_i is the characteristic (preferred) stimulus of the ith neuron; θ is the stimulus angle and ω is a width parameter. Unless otherwise stated, the following parameter values were used in all simulations involving this tuning function: $f_{mod} = 50\,\text{spikes/s}$, $\omega = 30°$. In all simulations the neurons' characteristic stimuli were uniformly distributed around the 360°range of the stimulus angle.

4.1.2 Trial-to-trial variability

Trial-to-trial variability was modelled by a multivariate Gaussian distribution:

$$\boldsymbol{r} \sim \mathcal{N}[\tau \boldsymbol{f}(\theta), Q(\theta)] \tag{4.2}$$

Where $\boldsymbol{r}$ is the vector of spike counts recorded in response to stimulus θ, $\boldsymbol{f}(\theta)$ is the vector of mean neuronal responses defined in the preceding section and τ is the integration time over which spike counts are recorded in each trial. In order to construct the inter-neuronal covariance matrix $Q(\theta)$, it is first necessary to establish the trial-to-trial variance of each individual neuron and any correlations in trial to trial variability.

A multiplicative model of neuronal variability was used:

$$\sigma_i^2(\theta) = F\tau f_i(\theta) \tag{4.3}$$

Where F is the Fano factor, the ratio of the spike count variance σ_i^2 to the mean spike count $\tau f_i(\theta)$ over the time interval τ. This type of model can be viewed as a generalisation of Poisson noise. The Poisson distribution is rather inflexible as it has only a single parameter, with the Fano factor fixed at unity. By using a Gaussian noise model, we gain an extra parameter and with it the flexibility to adjust the Fano factor. In addition to this, the Fisher information can still be found analytically, without having to resort to time consuming numerical methods. For this reason, negative spike counts have not been rectified to zero, as this would render the variability non-Gaussian. Using a non-zero background firing rate helps to prevent the occurrence of negative spike counts, and a value of $f_{bg} = 10\,\text{spikes/s}$ has been used in most simulations.

Inter-neuronal correlations in the trial-to-trial variability are defined by a correlation matrix C. Three forms of the correlation matrix are examined in this chapter:

- Independent trial-to-trial variability i.e. uncorrelated noise: C is the identity matrix:

  $$C_{ij} = \delta_{ij} \tag{4.4}$$

  Where δ is the Kronecker delta function.

- Localised correlations, specifically correlations that decay exponentially as a function of the difference in characteristic stimuli. In this case C is given by:

$$C_{ij} = \delta_{ij} + (1 - \delta_{ij})\, c \exp\left(-\frac{|\phi_i - \phi_j|}{\rho}\right) \tag{4.5}$$

  Where c is a correlation scaling coefficient and ρ is a correlation range coefficient. The correlation scales examined here ($0 \leq c \leq 0.3$) cover most biologically realistic scenarios. Unless otherwise stated, a range value of $\rho = 30°$ was used, meaning that the extent (in stimulus space) of the noise correlations and tuning curves were matched.

- Uniform correlations, where every pair of neurons has a trial-to-trial variability correlation coefficient of c:

$$C_{ij} = \delta_{ij} + (1 - \delta_{ij})\, c \tag{4.6}$$

Once the correlation matrix has been defined, the covariance matrix is given by:

$$Q_{ij}(\theta) = F\, [\tau f_i(\theta)]^{0.5}\, C_{ij}\, [\tau f_j(\theta)]^{0.5} \tag{4.7}$$

A number of factors determine the coding precision of a population, principally: the maximum and minimum (background) mean firing rates, the level of trial-to-trial variability, and the integration time over which spike counts are recorded. The Fano factor variability model was used as it allowed a convenient simplification to be made; increasing F clearly increases the variability of the response, while increasing τ means that we average the response over a longer time window and hence reduce the effective level of variability. In the case of Gaussian noise where the variance is determined by a Fano factor, F and τ have exactly equal and opposite effects, so we can fully capture the effect of both parameters by considering only their ratio F/τ, which has units of spikes/s^2 (further details are given in Appendix A.1).

We explore F/τ initially in the interval $[10^{-4}, 10^3]$ spikes/s^2, but most of the analyses extend only up to $F/\tau = 100$ spikes/s^2. Since F is commonly thought to be in the range $[1,3]$, the highest values of F/τ can be thought of as corresponding to recording time windows in the region of 10–30 ms. Some caution is required when applying this model with high values of F/τ. Very short integration times lead to low mean spike counts, and bring the model into a regime where the Gaussian distribution is no longer a good approximation of the Poisson-like distribution of real neuronal responses. For this reason we have restricted most of the analyses to $F/\tau \leq 100$ spikes/s^2.

4.1.3 Gain modulation

To examine the effect of adaptation-like gain changes on precision, we used the following gain modulation model (Seriès et al., 2009):

$$f^i_{mod} = f_{mod}\left[1 - \beta \exp\left(-\frac{1-\cos(\phi_i - \phi_{mod})}{(\frac{\pi}{180}\omega_{mod})^2}\right)\right] \tag{4.8}$$

Where f^i_{mod} is the post-adaptation stimulus-dependent firing rate modulation depth of the ith neuron and f_{mod} is the original peak firing rate common to all neurons. The 'adapting stimulus' and extent of adaptation (centre and width of the modulation profile) are defined by ϕ_{mod} and ω_{mod} respectively, while $\beta = [0,1]$ is a gain modulation scaling factor.

4.1.4 Cricket cercal system model

We also re-implemented a model of the cercal interneurons in the cricket first described by Theunissen and Miller (1991). The formulation given here is that used by Butts and Goldman (2006).

The stimulus model is as described above, and the population consists of four neurons with characteristic directions evenly spaced at 90°intervals around the 360°stimulus space. The mean response is given by a rectified cosine tuning curve:

$$f_i(\theta) = \frac{\cos(\theta - \phi_i) - 0.14}{0.86} \tag{4.9}$$

The standard deviation of the neuronal response is defined as a linear function of the mean response, hence the variance is a quadratic function of the mean (c.f. Equation 4.3, where the variance is a linear function of the mean). The parameter A is a variability scaling factor.

$$\sigma_i = A[0.048 + 0.052 f_i(\theta)] \tag{4.10}$$

The cosine tuning curve and noise are added together and negative values are rectified to zero, yielding the response spike counts:

$$r_i(\theta) = [f_i(\theta) + \eta]_+ \quad \text{where} \quad \eta \sim \mathcal{N}(0, \sigma_i^2) \tag{4.11}$$

The rectification has the effect that the variability becomes non-Gaussian; note that this is in contrast to the other simulations described in this chapter, where negative spike counts are not rectified in order to preserve Gaussianity. For the cricket cercal system model (Figure 4.4) only, Fisher information is calculated by Monte Carlo integration in order to take into account the non-Gaussian response distribution. The SSI calculations for this model assume a Gaussian response distribution and are therefore an approximation.

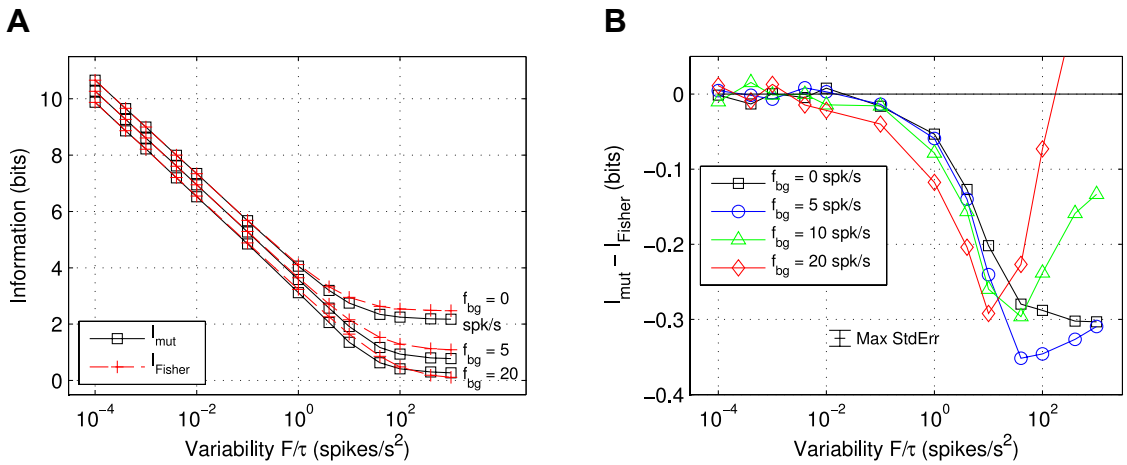

Figure 4.1: Mutual and Fisher information in a small population
(**A**) MI and I_{Fisher} as a function of variability, for several levels of background activity. Increasing the background activity increases the signal to noise ratio and reduces information. I_{Fisher} diverges from the MI with increasing F/τ; background activity accelerates this divergence. At high levels of variability, both the MI and I_{Fisher} flatten out and do not reduce with further increases in F/τ. Error bars too small to plot; see B.
(**B**) Difference between the MI and I_{Fisher} for the same data shown in A, together with an additional case $f_{bg} = 10\,\text{spikes/s}$. The maximum standard error across all points is shown on the plot.

4.2 Mutual information and I_{Fisher}

While it is known that the mutual information and I_{Fisher} are equal for infinite populations, how they are related in finite populations is less clear. As discussed in section 2.1.3, it has been shown that I_{Fisher} forms an upper bound on the MI, and that the MI approaches this bound asymptotically as N tends to infinity (Brunel and Nadal, 1998). To verify this numerically, and to establish the population size required for I_{Fisher} to provide an accurate estimate of the MI, a series of population models were examined. In addition, a four-neuron population model was used to assess the effect of trial-to-trial variability and background activity in very small populations.

Figure 4.1 shows how the MI and I_{Fisher} for a very small population (four neurons) vary as a function of the trial-to-trial variability. Both the MI and I_{Fisher} decrease with increasing variability, and are almost logarithmically proportional to F/τ (see Figure 4.1A). The MI is almost equal to I_{Fisher} when the noise level is low, even in a population of only four cells, and the difference between the two measures increases as the variability increases (see Figure 4.1B).

Background activity has a similar effect to variability. Since background activity is uniformly present and gives no information about the stimulus it is essentially noise, therefore increasing the background activity reduces the signal to noise

ratio, and this drop in SNR results in lower information values. Increased background activity also contributes to the divergence of MI and I_{Fisher}, leading to greater differences between the two measures for a given level of trial-to-trial variability.

For large values of F/τ (roughly corresponding to integration times of less than 10 ms with a Fano factor of 1), both the MI and I_{Fisher} flatten out and do not reduce with further increase in variability. In terms of Fisher information, this can be understood as the regime within which the trace term in Equation 2.7 is dominant (Shamir and Sompolinsky, 2004). In this regime, information is encoded primarily by the stimulus-dependent response variances, as opposed to the mean responses.

In most cases, the MI converges towards I_{Fisher} from below as the variability is decreased. However, for very high levels of variability combined with background activity, I_{Fisher} can be less than the MI (e.g. when $f_{bg} = 20\,\text{spikes/s}$ in Figure 4.1), and can even become negative (unlike the Fisher information, which is inherently non-negative). This occurs because the amount of noise in the system is such that the overall entropy of the response becomes significantly greater than the stimulus entropy[1], while the derivation of I_{Fisher} relies on the assumption that the entropies of stimulus and response are approximately equal. I_{Fisher} is therefore best at predicting the MI when I_{Fisher} is non-negative and within the logarithmically proportional regime with respect to F/τ.

Figure 4.2 shows the effect of population size upon the MI and I_{Fisher} under a number of different correlation structures. Figure 4.2A shows that I_{Fisher} is essentially proportional to $\log N$ over the range of population sizes examined. The asymptotic approach of the MI to the bound formed by I_{Fisher} is evident in Figure 4.2B, which shows the difference between the two measures plotted against N. Increasing the variability F/τ increases the difference between the MI and I_{Fisher} for a given population size. Despite this, even for the highest level of variability modelled (e.g. equivalent to supra-Poisson variability $F = 3$ with a time window of 30 ms), there is a difference of only 3.5% between the two measures for a population of 50 neurons. For $\tau = 300\,\text{ms}$, $F = 3$, the same relative error is achieved with less than 20 neurons.

From a decoding perspective, increasing the population size means that there are more parallel 'channels' carrying information about the stimulus. With a greater number of channels, a decoder can better average out the variability of these channels, hence coding precision is increased. The information carried by each channel becomes increasingly redundant as N increases, so the gain in coding precision diminishes; this is why we observe that information is approximately proportional to $\log N$ rather than N.

[1] Since both the stimulus and response variables in our model are continuous, all the entropies calculated in this chapter are differential entropies. These are largely equivalent to discrete entropy as described here, but are obtained by integrating over a continuous distribution rather than summing over a discrete distribution. See Appendix A.3.1 for further details.

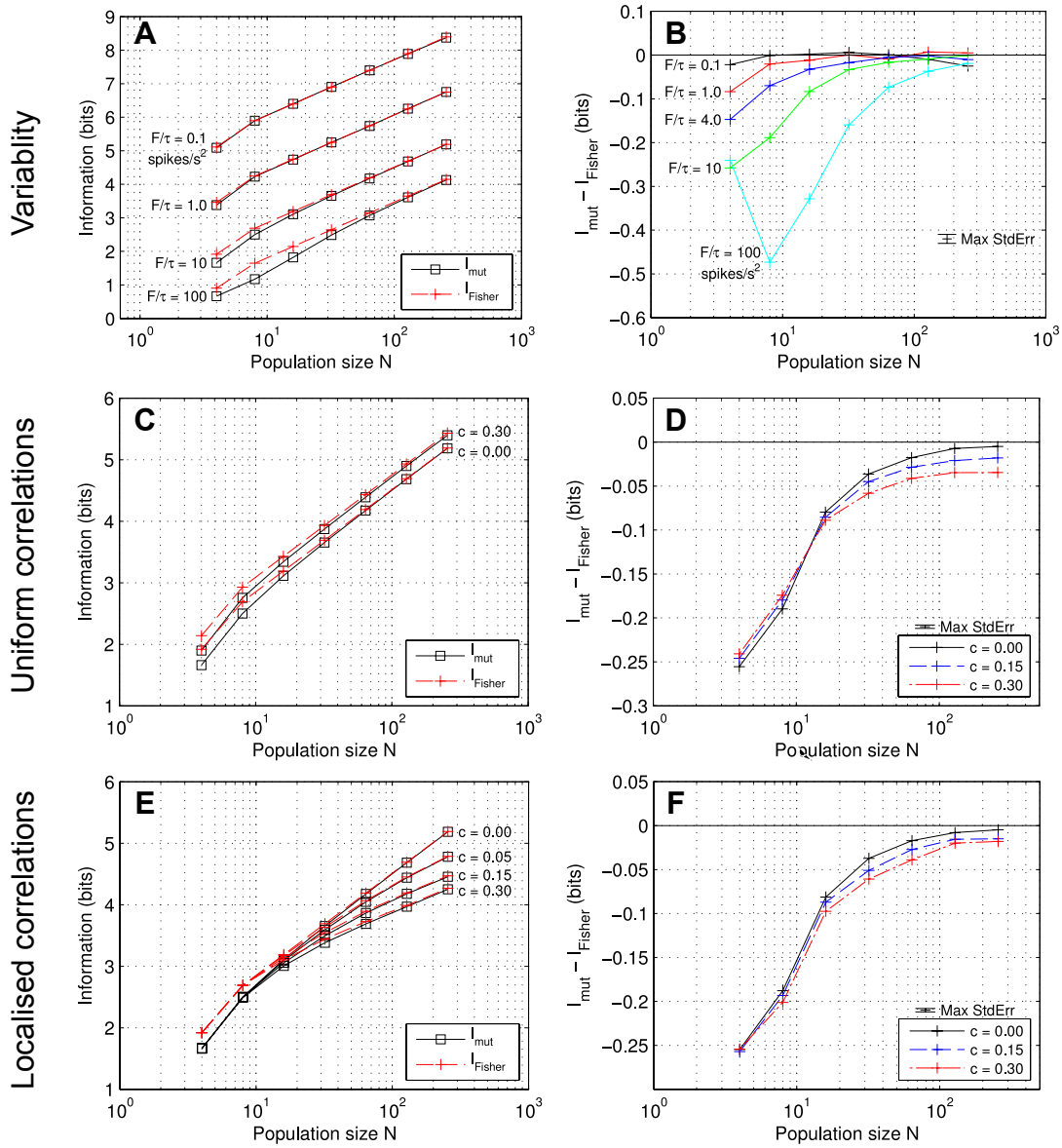

Figure 4.2: Mutual information and I_{Fisher} in large populations.
This figure summarises the relationship between the MI and I_{Fisher} in populations varying in size from 4 to 320 neurons. I_{Fisher} is, in most cases, a good approximation of the MI. The two measures diverge only for small populations ($N < 100$). Errors are shown as in Figure 4.1.
(A) MI and I_{Fisher} for various levels of independent variability. The MI converges towards I_{Fisher} from below with increasing N. Parameters: $f_{bg} = 10\,\text{spikes/s}$.
(B) This plot shows the difference between the MI and I_{Fisher} for the same cases as **(A)**.
(C),**(D)** Absolute values, and difference between, the MI and I_{Fisher} for various values of c with uniform correlation structure. Uniform correlations increase coding precision, but delay convergence of the MI and I_{Fisher}. Parameters: $F/\tau = 10\,\text{spikes/s}^2$, $f_{bg} = 10\,\text{spikes/s}$.
(E),**(F)** As per **(C)** and **(D)**, but with a localised correlation structure. Localised correlations reduce coding precision and delay convergence between the MI and I_{Fisher}. Parameters: $F/\tau = 10\,\text{spikes/s}^2$, $f_{bg} = 10\,\text{spikes/s}$.

The relationship between the MI and I_{Fisher} is complicated slightly when there are inter-neuronal correlations in the trial-to-trial variability. Panels C and D of Figure 4.2 show the effect of uniform correlations. The presence of uniform correlations slightly increases the information conveyed by the population, but, at the correlation levels simulated, the effect is much less than that of altering the level of variability. The information increase due to uniform correlations is effectively independent of population size. The reason for this increase in coding precision can be understood by considering the extreme case of $c = 1$. In this scenario, the noise correlation coefficient for every pair of neurons is 1, therefore every cell in the population exhibits exactly the same random noise. The relative firing rates of the neurons (the profile of activity across the whole population, determined by the tuning curves) are thus perfectly preserved, allowing very accurate decoding (see Averbeck et al., 2006, for further explanation of how noise correlations affect the precision of population codes).

Panels E and F of Figure 4.2 illustrate the effect of localised correlations. In contrast to uniform correlations, these act to reduce coding precision, although this effect is again small in comparison to that of variability. In large populations, the presence of localised correlations can have a marked effect, as it greatly reduces the rate with which both Fisher and the MI increase with $\log N$ (Wilke and Eurich, 2002).

In general, both uniform and localised correlations act to increase the difference between the MI and I_{Fisher}, although this effect is small in comparison to that of changing the level of variability. For very small populations ($N < 10$), however, uniform correlations actually reduce the difference. The effects of localised correlations vanish as the population size decreases, as the increasing spacing between tuning curves leads to a general reduction in pairwise correlation coefficients across the population. The effect of correlations, both uniform and localised, on the difference between the MI and I_{Fisher} is greatest for large populations, in contrast to the effect of variability, which diminishes with increasing N. As a result of this, correlations reduce the rate of convergence of the two measures, whereas variability itself does not.

The three noise correlation scenarios examined here can be seen as lying on a single continuum, where uniform correlations are equivalent to localised correlations with infinite range, and independent variability corresponds to zero range. The correlation range parameter ρ can be varied continuously, allowing the change in coding precision across this continuum to be explored. It is most useful to consider the correlation range relative to the width of the tuning curves, as the tuning curve width determines the extent of activity and the range of signal (as opposed to noise) correlations present in the population. Figure 4.3 shows how coding precision varies across the correlation range continuum. As shown perviously for Fisher information by Sompolinsky et al. (2001), the worst case scenario in terms of precision is when the correlation range matches the tuning curve width. It is

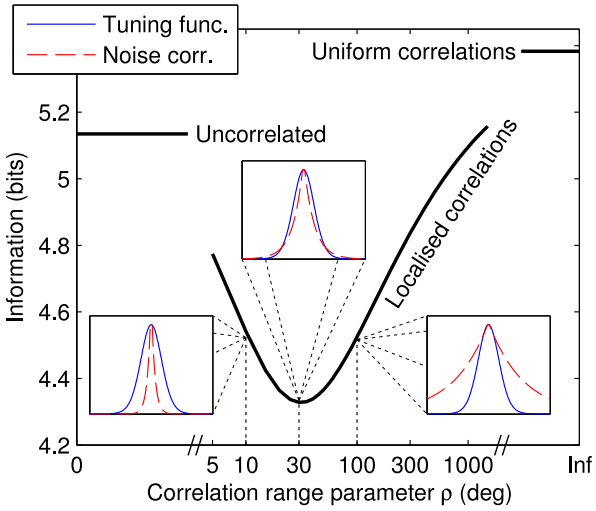

Figure 4.3: The effect of localised correlation range on precision.
The main plot shows how, given a constant correlation strength, I_{Fisher} is dependent on the correlation range parameter ρ. The MI is very similar and has been omitted for clarity. Introducing short-range correlations reduces coding precision relative to the uncorrelated noise case (top left), which is equivalent to $\rho = 0$. Precision decreases as the noise correlation range increases until it reaches a minimum, before increasing and converging towards the precision of the uniform correlation case as $\rho \to \infty$. The insets show the normalised tuning curve (solid line) and the correlation coefficients (i.e. a slice through the correlation matrix C; dashed line) for one neuron. Minimum information occurs where the noise correlation profile is most closely matched to the tuning curve i.e. where $\rho = \omega = 30°$.

well known that signal and noise correlations with the same sign, as is the case with localised noise correlations, degrade coding precision (Latham and Nirenberg, 2005; Averbeck et al., 2006). The worst case correlation range can be seen as a logical extension of that principle: maximal degradation of both Fisher and Shannon information occurs when the signal correlations (tuning curves) and noise correlations have not only the same sign, but the same extent and shape.

4.3 Which stimuli are most precisely represented by a neuron?

Tuning curves are commonly used to characterise the selectivity of neurons, but it is not always clear how they should be interpreted. Which stimuli does a neuron represent; which does it convey the most information about? Those at the peak of the tuning curve, where the activity of the neuron is most prominent? Or those at the steep flanks of the tuning curve, where the level of activity is most strongly modulated by small changes in the stimulus? To address these questions, Butts and Goldman (2006) calculated the stimulus-specific information for small populations

of model neurons ($N \leq 4$) and showed that the stimuli that are best encoded by a neuron depend upon the level of variability. For neurons operating within a low noise regime (see Figure 4.4B), the best encoded stimuli lie on the flanks of the tuning curve ('flank coding'), while those operating in the high noise regime (see Figure 4.4D) have a single best encoded stimulus coinciding with the peak of the tuning curve ('peak coding'). This property is not unique to the SSI; the specific surprise also gives similar predictions. This is in contrast to Fisher information, which always predicts that the best encoded stimuli lie on the flanks of the tuning curve. This finding was potentially troublesome to the field as it suggested that the interpretation of the tuning curve depends on the measure used to determine the stimulus-specific precision. To investigate the extent of this issue and its implications for the analysis of experimental data, we used the SSI to further investigate how trial-to-trial variability, and also population size, affect which stimuli are most precisely encoded.

When determining the best-encoded stimuli for a neuron within a population, both the marginal SSI and singleton SSI are relevant. The meaning of the singleton and marginal SSI can be intuitively understood by considering a scenario where a population is constructed progressively by introducing one neuron at a time. The singleton SSI and marginal SSI are the contributions to the population SSI from the first and last neurons respectively. Because there is redundancy in the information encoded by each neuron, the actual informational contribution from a single neuron within a population lies somewhere between these bounds (e.g. the shaded regions in Figure 4.4).

4.3.1 The effect of variability and integration time in small populations

As reported by Butts and Goldman (2006), the stimuli most precisely represented by a neuron, according to the SSI, can lie at either the peak or flanks of the tuning curve, depending on the amount of noise present. Figure 4.4 illustrates this by showing the marginal and singleton SSI of the cricket cercal interneuron model for three different noise levels. The tuning curves and variability are shown in Figure 4.4A.[2] Panels B–D of Figure 4.4 show how the best encoded stimuli shift from the flanks of the tuning curve to the peak of the tuning curve as the noise level is increased. Both the singleton and marginal SSI undergo this transition, with the marginal SSI transitioning between peak and flank regimes at a higher noise level (thus, if the singleton SSI is known to be in the flank coding regime, we can infer that the marginal SSI, which is more difficult to calculate, is also in the flank regime). The difference in the peak/flank transition point is due to the fact

[2] Figure 4.4 is based on results obtained from a re-implementation of the cricket cercal interneuron model used by Butts and Goldman (2006), and its layout is based on that of Figure 3 from their article.

that the marginal SSI relates to a four-neuron population, while the singleton SSI is based only upon a single neuron. The presence of more neurons in the population increases the coding precision, reducing the effective noise level of the code; this will be examined further in the following section on the effect of population size. It is important to note that in all three cases the predictions of Fisher information and SSI differ; the shapes of the curves are different, and indicate that different stimuli are most precisely encoded.

Under the flank coding regime, stimulus values can be read out by matching the firing rate of the neuron with the flanks of the tuning curve. Under the peak coding regime, it is not the precise level of activity, but the fact that the neuron's activity stands out from the background noise that conveys most of the information. This is a more robust, but coarser, indicator of the stimulus value—we know only that it is somewhere close to the neuron's characteristic stimulus—and this is reflected in the lower absolute SSI values.

To further investigate the peak/flank transition in small populations, we used a four-neuron population model with circular Gaussian tuning curves and Fano factor variability, as described in sections 4.1.1–4.1.2. We calculated both the marginal SSI and marginal specific surprise (I_{sur}) for several levels of variability (F/τ), so that the predictions of these closely-related measures could be compared (see Figure 4.5A). Both measures have similar shapes and absolute values, and both exhibit a transition from the flank regime to the peak regime with increasing F/τ, although for I_{sur} the transition occurs at a higher value of F/τ i.e. its flank regime is more extensive. It is important to note that the quantity F/τ represents both noise level (Fano factor) and integration time; a transition from peak to flank regime could be caused by an decrease in the Fano factor, or equivalently by an increase in the time over which spikes are counted in each trial. At low levels of variability (probably unrealistically low in biological terms), SSI and I_{sur} are practically indistinguishable. Although the two measures differ more at higher F/τ values, their shapes are qualitatively similar. Fisher information differs from both SSI and I_{sur} in all four cases. While the shape of the singleton Fisher information, and hence its indication of best-encoded stimulus, remains identical across the four levels of variability, its absolute value varies by two orders of magnitude (not shown). Even in the lowest variability case ($F/\tau = 0.1\,\text{spikes}/\text{s}^2$), where all three measures indicate flank coding, the best encoded stimuli predicted by Fisher information and the Shannon information measures differ.

Figure 4.5B shows the effect of altering the background firing rate upon the level of variability at which the peak/flank transition occurs. The shape of the marginal SSI is summarised by its peak to flank ratio (PFR); this is defined as the ratio of the SSI at the characteristic stimulus (tuning curve peak) to its value at the maxima of the Fisher information (flanks of the tuning curve).[3] A PFR value of one indicates

[3]The SSI_{flank} value does not necessarily correspond to the local maximum of the SSI, as the peaks

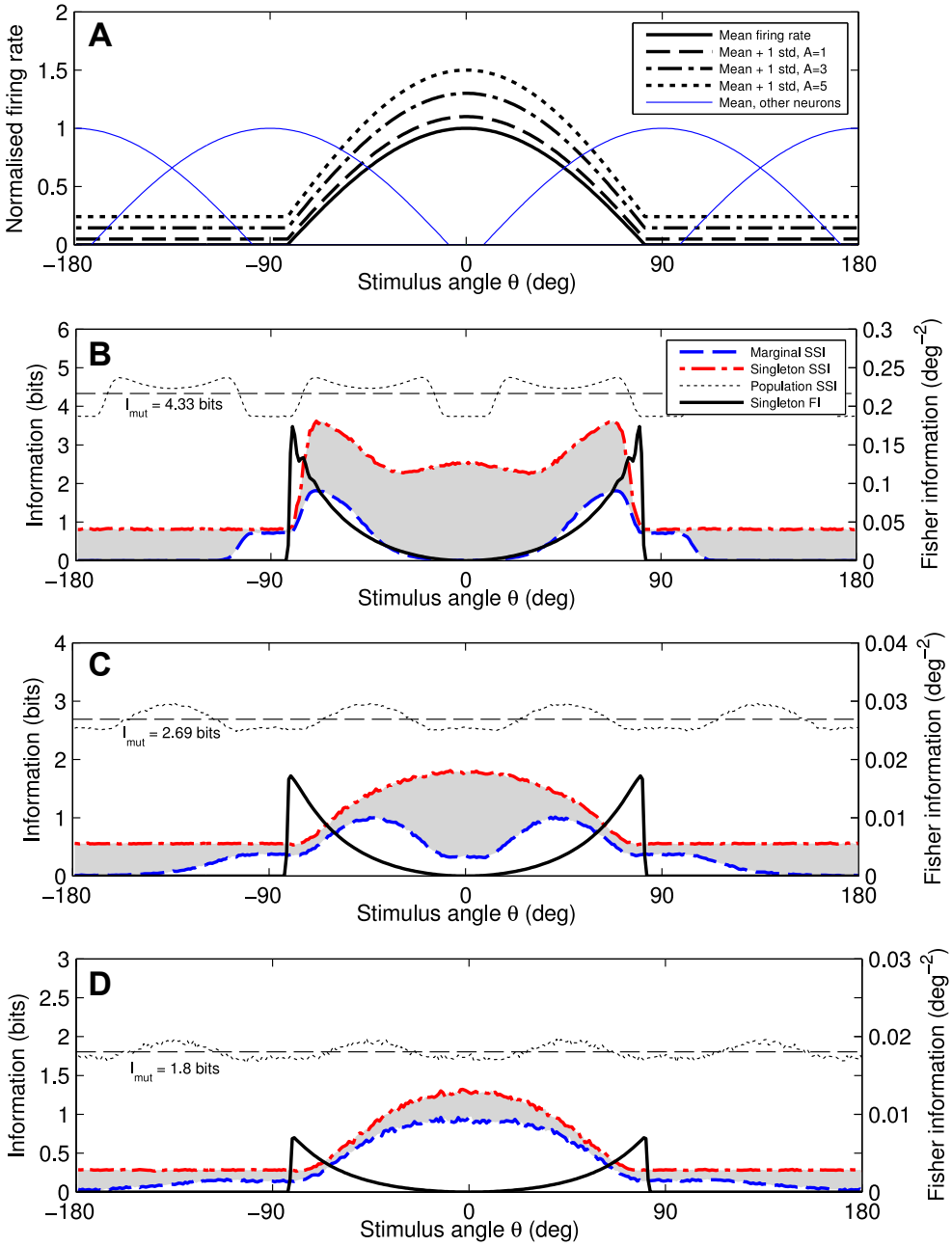

Figure 4.4: SSI example: cricket cercal interneurons

Model four-neuron population of cricket cercal interneurons as described by Theunissen and Miller (1991). This is a re-implementation of the simulations reported in Figure 3 of Butts and Goldman (2006).

(**A**) Tuning curves (mean responses) and three levels of trial to trial variability (illustrated as curves of mean + one standard deviation).

(**B**) Stimulus-specific information for the low noise case ($A = 1$). The dotted line shows the SSI of the whole population, while the shaded region shows the potential range of the information contribution of a single neuron, bounded from above by the singleton SSI and from below by the marginal SSI. Fisher information for a single neuron is shown for comparison. Both the singleton and marginal SSI indicate that the best-encoded stimuli lie on the flanks of the tuning curve.

(**C**) Intermediate noise case ($A = 3$). Here the singleton SSI is greatest at the peak of the tuning curve while the mSSI is greatest on the flanks.

(**D**) High noise case ($A = 5$). In this case both the singleton and marginal SSI are greatest at the peak of the tuning curve.

the point at which the SSI has three peaks of approximately equal value, and is therefore at the transition between the peak and flank regimes. PFR values of less than one correspond to the flank coding regime and values greater than one indicate the peak coding regime.

In the absence of background activity ($f_{bg} = 0$) the population remains within the flank coding regime up to $F/\tau \approx 30\,\text{spikes/s}^2$ (equivalent to $\tau \approx 33\,\text{ms}$ for $F = 1$). Introducing a small amount of background activity ($f_{bg} = 5\,\text{spikes/s}$, 10% of f_{mod}) has a pronounced effect, with a transition to the peak coding regime now occurring at $F/\tau \approx 3.5\,\text{spikes/s}^2$ (equivalent to $\tau \approx 285\,\text{ms}, F = 1$). Further increases in baseline activity continue to shift the peak/flank transition to lower F/τ values. This is line with the findings of Wilke and Eurich (2002), who noted a rapid decrease in Fisher information at low levels of background activity. When $f_{bg} = 0$, neurons with characteristic stimuli that differ from θ by more than about 3ω have essentially zero activity and hence zero variance. Increasing f_{bg} causes these neurons—approximately half of the population in this case—to fire at $f_{bg}\,\text{spikes/s}$ and to have a rate variance of $F f_{bg}\,(\text{spikes/s})^2$, thus substantially increasing the variability of the population as a whole.

Figure 4.5C illustrates the effect of uniform correlations in trial to trial variability upon the peak/flank transition. Uniform correlations improve coding precision and shift the regime transition to greater F/τ values relative to the uncorrelated case, although this is less pronounced than the shift caused by small levels of background activity.

4.3.2 The effect of population size

As demonstrated by Butts and Goldman and described in the preceding section, in very small populations the SSI can predict either flank or peak coding, depending on the amount of noise, noise correlation and the time over which spikes are counted. To date, this has not been investigated in populations larger than four neurons. Here we examine the effect of population size upon the stimuli that are best encoded by a neuron. Do both peak and flank regimes occur in larger populations? How does population size affect the transition between regimes?

By using Monte Carlo integration (Metropolis and Ulam, 1949) to compute the SSI and specific surprise (see Appendix A.3.2), we were able to extend the analysis of Butts and Goldman to populations of up to 256 neurons. Such a sampling approach is necessary because the dimensionality of the response distribution is equal to the number of neurons, so any algorithm that exhaustively integrates over this

of the SSI and Fisher information only become aligned as $N \to \infty$. The PFR can therefore be subject to fluctuations as parameter sweeps cause local SSI features to move across the stimulus value at which SSI_{flank} is calculated. The advantage of calculating the PFR in this way is that it is only necessary to compute the SSI at two predetermined points, as opposed to over the entire range of the stimulus variable.

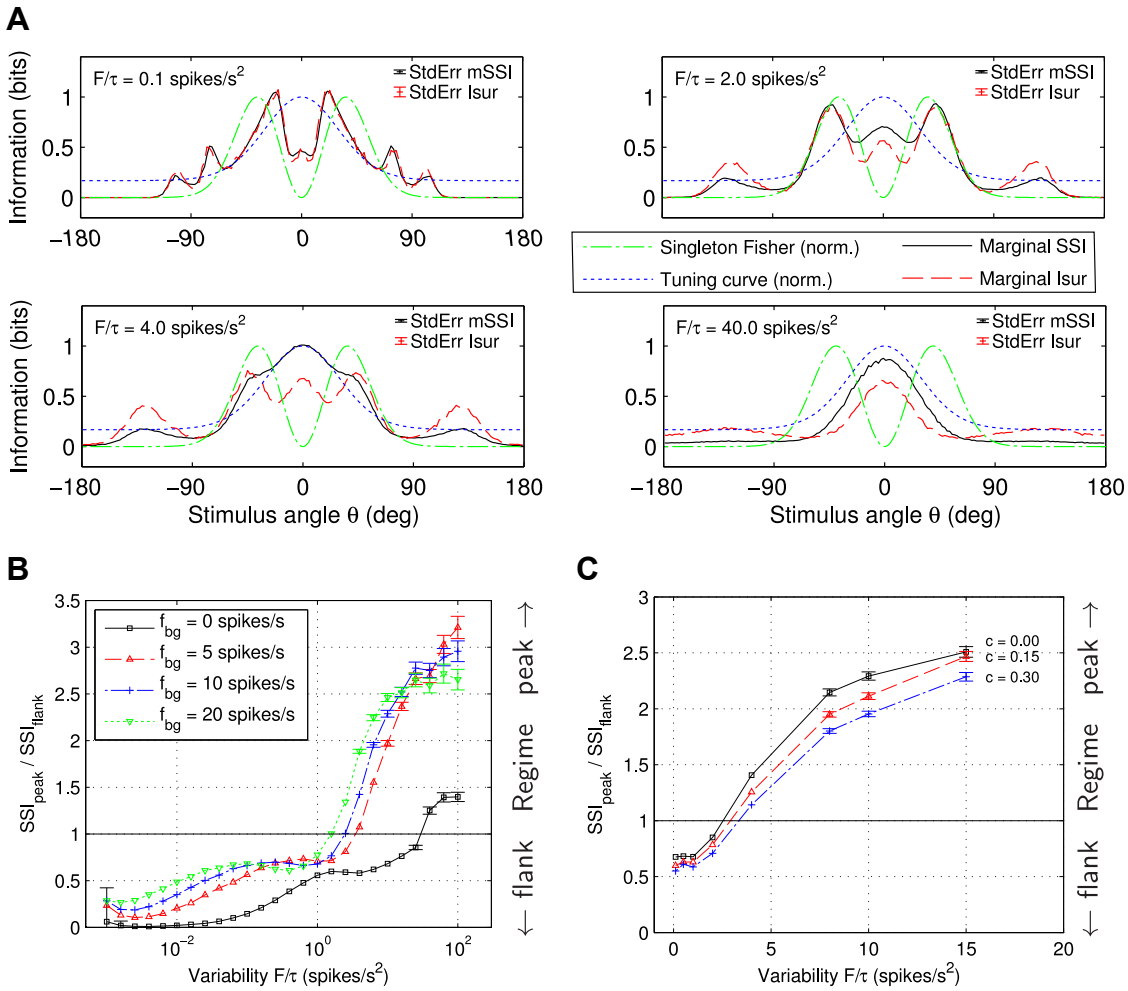

Figure 4.5: Best-encoded stimuli in small populations.
In small populations (here $N = 4$), SSI and specific surprise (I_{sur}) can be either unimodal or bimodal, depending on the level of trial-to-trial variability.
(A) Marginal SSI and marginal I_{sur} for several levels of variability. SSI and I_{sur} are very similar at low F/τ values, but diverge to some extent as the variability increases. Both measures undergo a transition from bimodal (greatest on the flanks of the tuning curve) to unimodal (greatest at the peak of the tuning curve) as the variability increases; this occurs slightly earlier for the SSI. Parameters: independent variability, $f_{bg} = 10\,\text{spikes/s}$. Error bars show the worst case standard error for each measure.
(B) Marginal SSI peak to flank ratio (PFR) for several levels of background activity f_{bg}, with independent variability. Altering the level of background activity has a pronounced effect on the transition between low noise ($I_{peak}/I_{flank} < 1$) and high noise (> 1) regimes, with higher f_{bg} causing the transition to occur at a lower level of variability. For clarity, error bars have been omitted where the standard error is less than 0.02 bits.
(C) Marginal SSI PFR for various values of c, with uniform correlation structure. Uniform correlations improve coding precision, delaying the transition from flank to peak regime to greater levels of variability compared to the independent case. Parameters: $f_{bg} = 10\,\text{spikes/s}$. Error bars omitted when StdErr $< 0.02\,\text{bits}$.

distribution quickly becomes intractable as the population size increases. In order to validate the Monte Carlo approach we first replicated (see Figure 4.4) the results shown in Figure 3 of Butts and Goldman (2006), which describe the SSI for the cricket cercal interneurons, and were obtained via quadrature integration. Because of the similarity between the SSI and specific surprise, the unique advantages of the specific information (on which the SSI is based), and because the Monte Carlo estimate converges more rapidly for the SSI than for the specific surprise (due to its averaging over the stimulus ensemble), we leave aside the specific surprise and focus on the SSI for the remainder of the chapter.

We used the SSI to examine how the best-encoded stimulus of a neuron is affected by the size of the population that it exists within. Figure 4.6A shows the marginal SSI for populations of various sizes; all curves are normalised to allow comparison. The mSSI shows a transition from the peak coding to the flank coding regime with increasing N, and the shape of the mSSI approaches the shape of the Fisher information as N becomes larger, although the units and absolute values of the two measures are different.

The transition between peak and flank regimes with N is shown in Figure 4.6B for several levels of variability, using the peak to flank ratio to summarise the shape of the marginal SSI. For very low levels of variability ($F/\tau = 0.1\,\text{spikes}/\text{s}^2$), the population operates in the flank regime at all population sizes, but at more realistic noise levels a transition occurs. The population size at which this happens depends on the level of variability: more noise means that a larger population size is required before the population moves into the flank coding regime.

For sufficiently large populations (approximately $N > 50$), variability no longer determines the coding regime and no (qualitative) discrepancy exists between the measures; both predict that neurons operate in the flank coding regime (for τ in the range $[10, 30]\,\text{ms}$ given F in the range [1,3]). Population size, along with trial-to-trial variability, is therefore an important determinant of the coding properties of individual neurons within a population.

Panels C and D of Figure 4.6 show the effect of correlations. In line with other findings, uniform correlations increase precision and hence drive the population towards the flank coding regime, while localised correlations have the opposite effect. Figure 4.6D shows the effect on the PFR and transition point; localised correlations decrease the PFR and shift the peak/flank transition to lower N, while localised correlations have the opposite effect. The effect of localised transitions on the PFR is greatest at moderate population sizes around the regime transition, while uniform correlations have the greatest effect in small populations. To understand this difference, recall that the effect of localised correlations on precision is negligible in very small populations and increases with population size (see Figure 4.2E), whereas the effect of uniform correlations does not vary with population size (see Figure 4.2C).

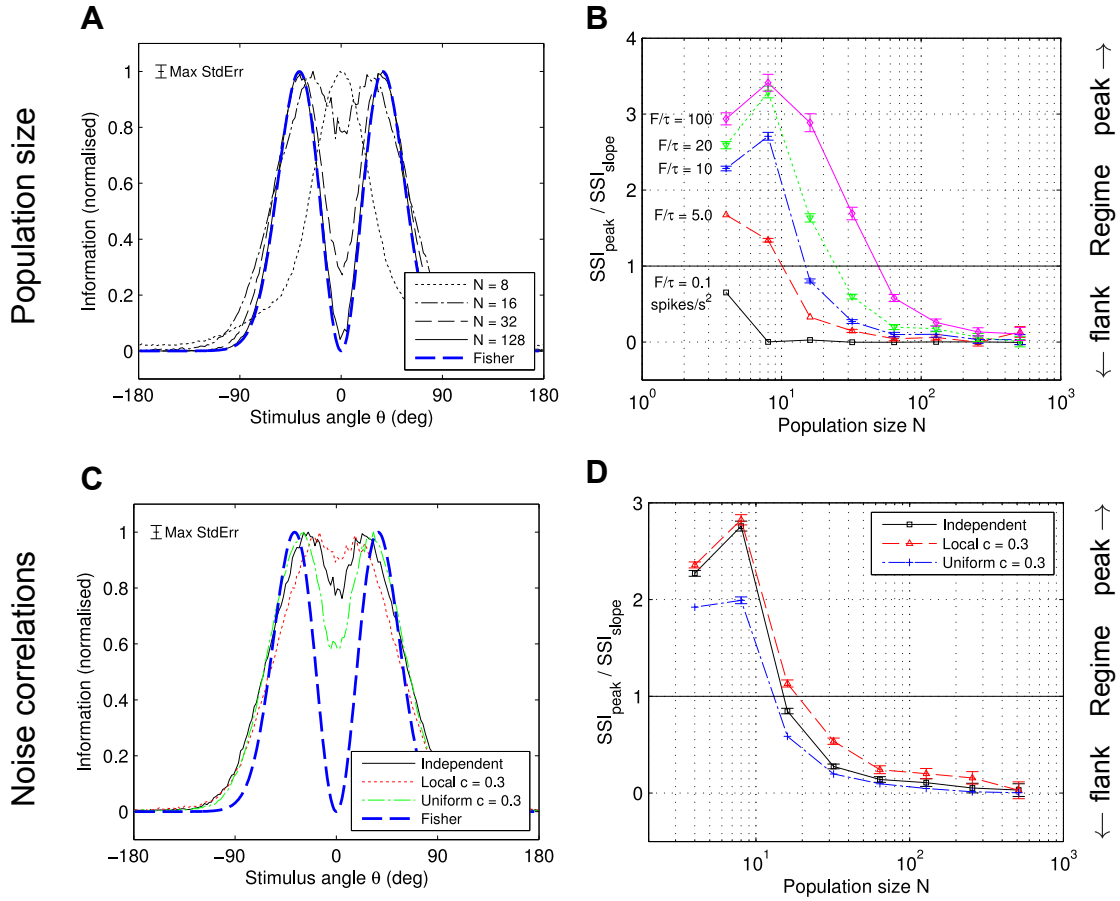

Figure 4.6: The effect of population size upon mSSI and PFR.
Marginal SSI (mSSI) and marginal SSI peak to flank ratio (PFR) in populations of various sizes. The stimulus value that is most precisely encoded by a neuron varies with population size. Maximum information occurs on the flanks of the tuning curve in large populations, but can occur at the peak or flanks in small populations, depending on the level of variability.
(A) Marginal SSI for various population sizes with independent variability. This plot illustrates the transition of greatest SSI from peak to flank of the tuning curve, and towards the shape of the singleton Fisher (heavy dashed line). Parameters: $F/\tau = 10\,\text{spikes/s}^2$, $f_{bg} = 10\,\text{spikes/s}$. Error bar: worst case across all N, θ.
(B) PFR versus N for various levels of independent variability. The $F/\tau = 10\,\text{spikes/s}^2$ case corresponds to the SSI curves in A. Increasing variability delays the transition ($SSI_{peak}/SSI_{flank} = 1$) to greater population sizes. Parameters: $f_{bg} = 10\,\text{spikes/s}$. Error bars $< 0.02\,\text{bits}$ omitted.
(C) The effect of correlated variability on marginal SSI; localised correlations bring the neuron closer to the peak regime, while uniform correlations have the opposite effect. Singleton Fisher information shown for comparison; at this population size the mSSI has not yet converged to the shape of the Fisher information. Parameters: $N = 16$, $F/\tau = 10\,\text{spikes/s}^2$, $f_{bg} = 10\,\text{spikes/s}$. Error bar: worst case across all N, c.
(D) PFR curves showing the effect of correlated variability. Localised correlations increase PFR, corresponding to reduced coding precision, while uniform correlations have the opposite effect. Parameters: $F/\tau = 10\,\text{spikes/s}^2$, $f_{bg} = 10\,\text{spikes/s}$. Error bars $< 0.02\,\text{bits}$ omitted.

4.3.3 SSI_{Fisher}

As described in section 2.1.3, I_{Fisher} allows us to make quantitative comparisons between Fisher information and Shannon mutual information when considering overall coding precision, but when dealing with stimulus-specific precision only qualitative comparisons have previously been possible. Qualitatively, Figure 4.6C suggests that the shape of the marginal SSI converges towards the shape of the singleton Fisher information as the population size goes to infinity. To allow this convergence to be investigated quantitatively (i.e. using the same units), we propose a new measure: SSI_{Fisher}. SSI_{Fisher} is a stimulus-specific decomposition of I_{Fisher}; more specifically it is the SSI of an optimal Gaussian-distributed estimator that saturates the Cramér-Rao bound (a formal definition is given in Appendix A.2). SSI_{Fisher} is an approximation of the SSI, in the same way that I_{Fisher} is an approximation of the MI. Here we consider the marginal SSI_{Fisher} (mSSI_{Fisher}), which is calculated in the same way as the marginal SSI, but is based upon SSI_{Fisher} rather than the SSI itself.

Figure 4.7A shows mSSI, mSSI_{Fisher}, and Fisher information together, for several population sizes. Fisher information is shown on a separate scale for ease of comparison and the scales are adjusted such that the maximum of $\text{SSI}/\text{SSI}_{Fisher}$ is level with the maximum Fisher information. It can be seen that the three curves converge with increasing N; in the case of SSI and SSI_{Fisher} this convergence is to the same absolute value, which equals Fisher information up to a multiplicative constant. The SSI_{Fisher}, like the SSI, undergoes a peak to flank transition with increasing N. Interestingly, the transition occurs later in SSI_{Fisher} than in the SSI itself, which is surprising as SSI_{Fisher} is derived from Fisher information, which relates to an upper bound on coding precision.

The convergence of mSSI and mSSI_{Fisher} roughly parallels that of the MI and I_{Fisher}, but the latter converge more quickly. Panels B and C of Figure 4.7 show the difference between Shannon and Fisher information based measures, as proportion of the Shannon information, for stimulus-specific (SSI, SSI_{Fisher}) and overall (MI, I_{Fisher}) quantities respectively. It can be seen that convergence occurs at approximately the same rate for both sets of measures, and that the relationships between the four variability cases are similar on both plots. Note that the scales on the two plots are different; the proportional difference between MI and I_{Fisher} is less than that between SSI and SSI_{Fisher}. This is due to differences in what is being measured: the marginal measures compared by ΔmSSI relate to the rate of change of overall information with respect to N, rather than the absolute value.

4.3.4 Summary

For large populations (more than around 50 neurons, for integration times down to around 10–30 ms) both Fisher information and the marginal SSI indicate that neurons provide information primarily about stimuli on the flanks of their tuning curves. Even for large populations, there is some difference between the shapes of Fisher information and marginal SSI, but this diminishes as the population size increases. Smaller populations, however, can operate in either the flank coding regime or a peak coding regime where neurons convey most information about stimuli at the peaks of their tuning curves. Here, the regime depends upon the level of trial-to-trial variability (noise), integration time window, the amount and structure of noise correlations, and the population size. Increased noise, the presence of localised noise correlations, and reduced population size all drive the system towards the peak coding regime. Conversely, decreased noise, uniform noise correlations, and larger population sizes have the opposite effect, moving the population towards the flank coding regime. For small, noisy populations—populations operating in the peak coding regime—Fisher information gives a misleading indication of which stimuli are best represented by a neuron. This discrepancy between best-encoded stimulus predictions (mSSI versus Fisher) parallels the divergence of overall coding precision (MI versus I_{Fisher}) in small populations described in section 4.2.

4.4 Gain Modulation

Gain modulation due to adaptation or attention-like processes is an often observed phenomenon in sensory neurons (see e.g. Wark et al., 2007). We applied the principles introduced above to examine the functional consequences of adaptation-like localised negative gain modulation, using the model described in section 4.1.3. Reducing the overall activity of the population is equivalent to reducing the signal to noise ratio, so the negative gain modulation causes a reduction in the MI and I_{Fisher}. This is well understood, so the investigation focussed on the stimulus-specific precision: does adaptation affect the representation of the adapting stimulus itself or adjacent stimuli on the flanks of the affected tuning curves? Is the coding precision of the adapted stimulus increased, decreased or unchanged?

Population Fisher information and population SSI were calculated for 108 model populations with different combinations of population size, tuning curve width, modulation width and modulation depth. The shape of both Fisher information and SSI for the population depends mainly upon the relative width of the tuning curves and modulation profile; the function defining the height of the tuning curve peaks (see Figure 4.8). When the modulation profile is narrower than the tuning

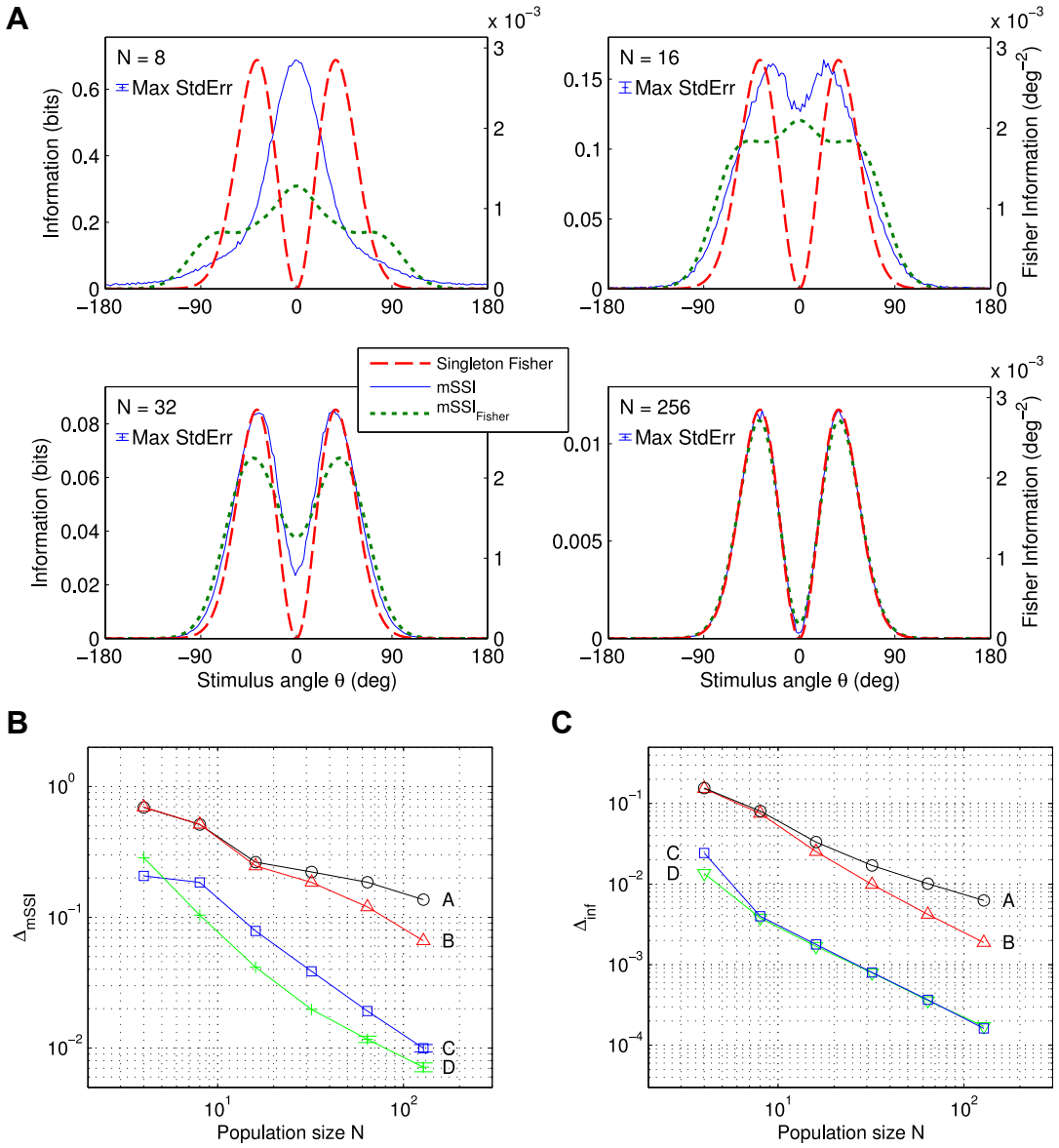

Figure 4.7: Marginal SSI converges towards marginal SSI_{Fisher} as population size increases. **(A)** Marginal SSI, marginal SSI_{Fisher} and singleton Fisher information for populations of different sizes. The scales of the y axes are adjusted so that the maximum value of mSSI, mSSI_{Fisher} is aligned with the maximum value of the Fisher information. Parameters: independent variability, $F/\tau = 10\,\text{spikes/s}^2$, $f_{bg} = 10\,\text{spikes/s}$. Error bar: worst case across θ.
(B) Convergence of marginal SSI and marginal SSI_{Fisher} for various parameter values. The y axis quantity is defined as $\Delta_{mSSI} = \frac{\text{RMS}(mSSI - mSSI_{Fisher})}{\text{RMS}(mSSI)}$ where RMS denotes the root mean square. Case A: localised correlations, $c = 0.3$, $F/\tau = 10\,\text{spikes/s}^2$, $f_{bg} = 10\,\text{spikes/s}$. Case B: independent variability, $F/\tau = 10\,\text{spikes/s}^2$, $f_{bg} = 10\,\text{spikes/s}$. Case C: independent variability, $F/\tau = 1\,\text{spikes/s}^2$, $f_{bg} = 10\,\text{spikes/s}$. Case D: independent variability, $F/\tau = 1\,\text{spikes/s}^2$, $f_{bg} = 0\,\text{spikes/s}$. Error bars $< 5\%$ relative error omitted.
(C) Convergence of the MI and I_{Fisher} roughly parallels the convergence of mSSI and mSSI_{Fisher}. $\Delta_{inf} = \frac{|I_{mut} - I_{Fisher}|}{I_{mut}}$ Parameter values are the same as in **(B)** for each case. Error bars $< 5\%$ relative error omitted.

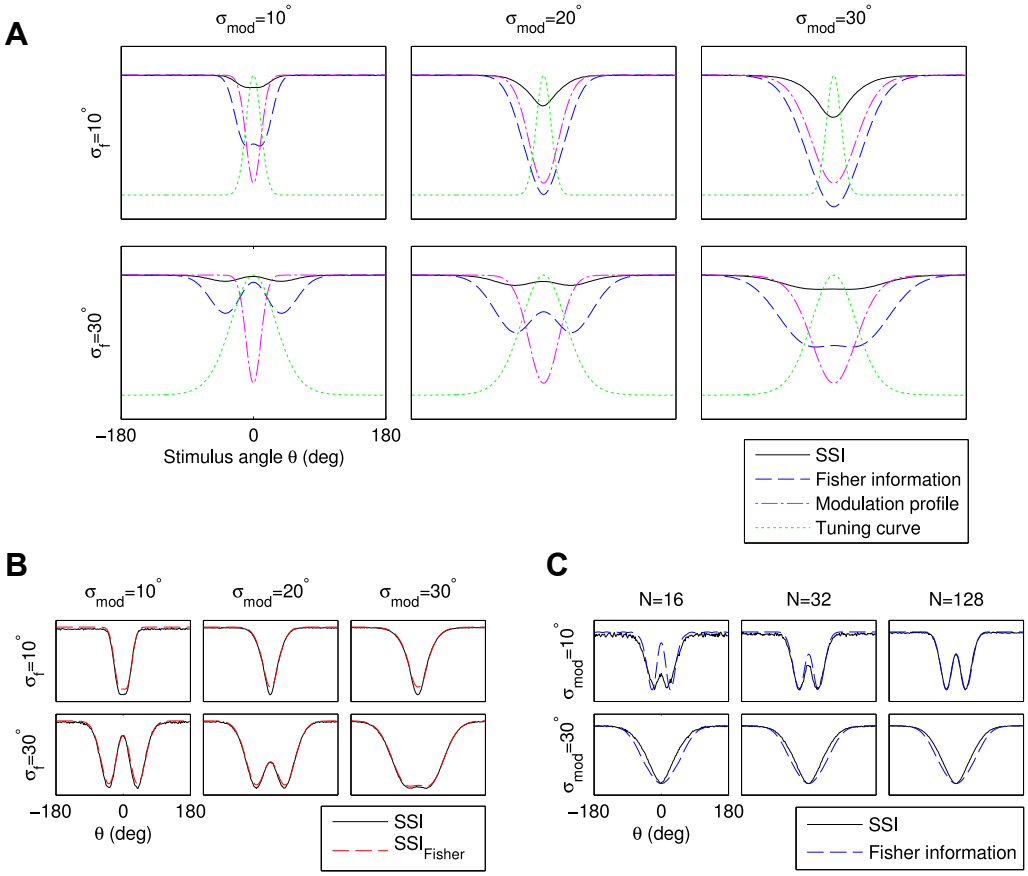

Figure 4.8: The effect of adaptation-like gain modulation on the SSI.
(**A**) Population SSI and Fisher information for various combinations of tuning curve width (ω) and modulation profile width (ω_{mod}). SSI and Fisher are normalised so that their maximum values coincide on the plots. Similarly, zero for both measures coincide at the bottom of each plot. A single unmodulated tuning curve is shown, along with the modulation profile, to allow the widths to be visualised. The shape of both SSI and Fisher information is dependent upon the relative width of modulation and tuning curves. When $\omega_{mod} < \omega$ a double trough shape is produced. The closer the two widths, the less pronounced the central peak. For $\omega_{mod} > \omega$, there is a single trough with the least precise coding occurring at the centre of the modulation (the 'adapting' stimulus). Parameters: modulation depth 90%, independent variability, $N = 128$, $F/\tau = 5\,\text{spikes/s}^2$, $f_{bg} = 10\,\text{spikes/s}$.
(**B**) Population SSI and SSI_{Fisher} for the same cases shown in (**A**). Both measures are plotted on the same scale. Although the values and shapes of SSI and Fisher shown in A differ, they indicate the same precision: SSI_{Fisher} and SSI are almost equal across all stimulus values.
(**C**) Trough shapes for population SSI and Fisher converge with increasing N, but the shapes of both functions are largely independent of N for large populations. Measures are normalised such that maximum and minimum values are aligned. Parameters as above, $\omega = 20°$.

curves, both Fisher information and SSI have a double trough shape, with the coding precision of stimuli adjacent to the adapting stimulus reduced, while the representation of the adapting stimulus itself remains relatively unaffected (see Figure 4.8A, bottom left axes).[4] Hol and Treue (2001) observed a similar effect in a human psychophysical study; they observed that adaptation had no effect on the discrimination threshold at the adapting stimulus, but increased the threshold for neighbouring stimuli on both sides.

As the modulation width is increased relative to the tuning curve width, the representation of the adapting stimulus becomes progressively less precise relative to neighbouring stimuli (bottom centre panel, in this example $\omega_{mod} = \omega$). Ultimately, as the modulation width is further increased beyond the tuning curve width, the adapting stimulus becomes the least precisely represented stimulus, with coding precision increasing monotonically with distance from the adapter (top right panel).

Although the shapes and values of the population Fisher information and SSI in Figure 4.8A are different, the coding precision that they imply is essentially the same. Figure 4.8C shows the population SSI and SSI_{Fisher} for the same six cases. In all cases the SSI and SSI_{Fisher} are very similar. The slight difference between SSI and SSI_{Fisher} in the immediate neighbourhood of the adapting stimulus (the bottom of the trough) may be due to low response firing rates, which locally increase the signal to noise level and delay convergence of the two measures.

Figure 4.8C illustrates the convergence of the shape of the population SSI and Fisher information as the population size is increased. In the same way that singleton Fisher information is usually in agreement with the marginal SSI, we find here that population Fisher information in most cases predicts the same pattern of stimulus-specific precision as the population SSI. For large populations ($N > 32$) the shapes of the two measures are very similar. Differences are observed only within a restricted domain where N is small and the modulation width is narrow relative to the tuning curve width. In these cases Fisher information overestimates, relative to the SSI, the representation precision for the adapting stimulus.

4.5 Conclusion

The results presented in this chapter show that it is feasible to compute the SSI for populations consisting of hundreds of neurons via Monte Carlo integration,

[4] The extreme case for this scenario is an infinitesimally narrow modulation profile. In this case, the gain modulation would only affect a single neuron and the singleton Fisher information of this cell would effectively be subtracted from the population Fisher information, leading to the double trough shape with the representation precision of the adapting stimulus unchanged and that of neighbouring stimuli (those lying on the slopes of the modulated tuning curve) reduced. As the modulation width is increased, more neurons are affected and the troughs and central peaks begin to cancel out, eventually resulting in a single trough with the maximum reduction in coding precision occurring at the adapting stimulus itself.

even in the presence of correlated variability. This means that the SSI has the potential to be used to analyse experimental results at the population level, as well as for single neurons. Although the full set of results presented in this chapter represents considerable computational effort, calculating the SSI for a single empirically-determined model, even with 200 neurons, requires at most a day or so of computing time on a modern desktop computer.

The predictions of the SSI and Fisher information converge rapidly as a function of the number of neurons in the population. The exact pattern of convergence depends on the parameters of the chosen model. However, we found that for populations larger than around 50 neurons, they are qualitatively identical, even with high levels of variability and/or short integration times. The stimuli that are best encoded are then always those falling on the flanks of the tuning curves. This indicates that there is no need for populations to be very large for the SSI and the Fisher information to yield similar predictions. The marginal SSI and the Fisher information differ only over a restricted domain (small temporal windows, small populations, high noise), which seems to roughly correspond to the range where Fisher Information 'fails', i.e. where the Cramér-Rao Bound is not saturated by maximum-likelihood or other optimal decoders (Bethge et al., 2002; Xie, 2002).

Correlations in the trial to trial variability (noise correlations) have a relatively minor effect upon the convergence of information theoretic and Fisher-based measures. The 50-neuron guideline threshold for qualitative convergence holds in the presence of biologically realistic levels of correlation, whether uniform or localised.

An important direction for future research is to examine how coding accuracy and best-encoded stimuli depend on the coarseness of discrimination. In this chapter, we have shown that large populations probably operate in the flank coding regime for fine discrimination tasks, and it is clear that the peak coding regime is relevant for very broad discrimination, where entirely separate groups of neurons are activated in response to the stimuli. What happens between these two edge cases has yet to be investigated. The peak and flank coding regimes discussed in this chapter are specific to unimodal tuning curves. For monotonic tuning functions, high Fisher information occurs at the steeply sloping region of the curve; this is equivalent to the flank coding regime in unimodal tuning curves. It is not clear what the best encoded stimulus is for monotonic tuning curves in tasks other than fine discrimination. This question is addressed in Chapter 5, which extends the analysis described in this chapter to monotonic tuning curves and other behavioural tasks besides fine discrimination.

Chapter 5

What Determines the Best-Encoded Stimulus?: Monotonic Tuning Curves and the Influence of Behavioural Task

The work presented in this chapter has been submitted for publication in *Frontiers in Computational Neuroscience* as 'The Influence of Population Size, Noise and Behavioural Task on Best-Encoded Stimulus for Neurons with Unimodal or Monotonic Tuning Curves' (Yarrow et al., 2014).

The Matlab code used to obtain the results described in this chapter is available in the form of a toolbox from https://github.com/StuYarrow/Popcode.

This chapter describes theoretical modelling work that examines the effect of trial-to-trial variability, population size and behavioural task on the best-encoded stimulus. In this chapter, the analyses presented in Chapter 4 are extended to monotonic tuning curves, and the effect of stimulus alternative (choice) spacing in forced choice tasks on the best-encoded stimulus is also examined. First, we use the SSI to determine the best-encoded stimuli for populations of neurons with monotonic tuning curves and show that, as with unimodal tuning curves, the SSI and the Fisher information predict similar best-encoded stimuli for large populations. In smaller populations, we show that the best-encoded stimulus depends on the level of trial-to-trial variability. We then go on to

show how Chernoff distance can be used to quickly estimate the best-encoded stimulus for two-alternative forced choice tasks, and to examine how the best-encoded stimulus for monotonic tuning curves is affected by the behavioural task, specifically the distance between stimuli and the number of stimuli in discrimination and classification tasks. We show that, as with unimodal tuning curves, when the stimulus distribution is primarily defined by the task it is the task, together with the tuning curves and variability, that determines the best-encoded stimulus.

While bell-shaped tuning curves have been widely studied (see e.g. Paradiso, 1988; Zhang and Sejnowski, 1999; Sompolinsky et al., 2001; Wilke and Eurich, 2002; Butts and Goldman, 2006), sigmoidal monotonic tuning curves have received little attention from the theoretical community (Guigon, 2003; Salinas, 2006; McDonnell and Stocks, 2008). Studies using the SSI have shown that, for rich stimulus ensembles with many possible stimuli, the best-encoded stimulus for peaked tuning curves depends on the level of trial-to-trial variability and the number of neurons in the population (Butts and Goldman, 2006, and Chapter 4). Population codes involving monotonic tuning curves have received comparatively little attention, but are also important in sensory neuroscience as they are often found where the intensity of neuronal activity reflects the intensity of the stimulus, for example relative luminance in the visual system (e.g. Sakmann and Creutzfeldt, 1969), sound intensity in the auditory system (e.g. Sachs and Abbas, 1974), and pressure of touch in the somatosensory system (Adrian and Zotterman, 1926b). For monotonic tuning curves, the Fisher information predicts that the best-encoded stimulus lies on the sloping flank of the tuning curve, but the SSI has not previously been computed for such tuning curves. Building upon the earlier studies of unimodal tuning curves, this chapter focusses first upon on monotonic tuning curves and asks: (1) what are the best-encoded stimuli for monotonic tuning curves? and (2) are the best-encoded stimuli dependent on the level of variability and the population size?

Studying monotonic tuning curves introduces some additional complexity, as monotonicity requires a linear stimulus variable rather than a circular one. Most theoretical studies of population coding (including Chapter 4 of this thesis and Paradiso, 1988; Seung and Sompolinsky, 1993; Zhang and Sejnowski, 1999; Wilke and Eurich, 2002; Butts and Goldman, 2006; Berens et al., 2011) use angular stimuli, as the endless nature of the stimulus space is mathematically convenient and orientation tuning in the visual cortex is a popular subject for experimental work. Whereas a uniform stimulus distribution on a periodic variable (such as edge orientation in natural scenes) can be a reasonable approximation of reality, linear stimulus parameters such as luminance or sound intensity are far from uniformly distributed in nature. The popularity of circular stimulus spaces in theoretical work has meant that relatively little is known about how non-uniformity of the stimulus

distribution might affect which stimuli are best encoded by a neuron. To address this, the SSI was computed for uniform and non-uniform stimulus distributions in order to show how best-encoded stimuli for both unimodal and monotonic tuning curves are affected by local non-uniformity in the stimulus distribution.

The SSI can also be used to determine the best-encoded stimulus in the context of any arbitrary task by manipulating the stimulus distribution; for example, a two-alternative forced choice task can be modelled by a stimulus ensemble consisting of only two stimuli. When the stimulus ensemble is restricted by the task in this way, the spacing between stimuli determines the best-encoded stimulus for unimodal tuning curves (Butts and Goldman, 2006). For closely spaced stimuli, the difference in the neuronal response distributions that they elicit is dominated by the tuning curve gradient and the SSI predicts that the best-encoded stimuli are on the flanks of the tuning curve. This is the same as the prediction of the Fisher information—which is unsurprising as the Fisher information is specific to fine discrimination and is based on the derivative of the tuning curve. Conversely, if the spacing between stimuli is large, the SSI predicts that the best-encoded stimulus lies at the peak of the tuning curve. Building upon this work, we ask: how does the best-encoded stimuli for monotonic tuning curves depend on the behavioural task?

5.1 Methods

5.1.1 Model framework

The simulations described in this chapter are based upon a model population of rate-coding neurons representing an abstract one-dimensional stimulus S. We consider both discrete and continuous stimuli and, unless otherwise stated, the stimulus is uniformly distributed across a finite non-periodic interval.

$$\text{Discrete:} \qquad P(S=s) = \frac{1}{k} \tag{5.1}$$

$$\text{Continuous:} \qquad p(S=s) = \frac{1}{s_{max} - s_{min}} \tag{5.2}$$

Where $P(S=s)$ and $p(S=s)$ are probability mass and density respectively at the stimulus value s (upper case characters represent ensembles and lower case characters represent concrete values). In the discrete case the ensemble consists of k stimulus values, and in the continuous case stimuli can take any value in the interval $[s_{min}, s_{max}]$.

The stimulus is encoded in the firing rates of N neurons with response spike counts $\boldsymbol{r} = \{r_1, r_2, \ldots, r_N\}$; the responses of each neuron are conditionally independent given S. The response R_i of the ith neuron is Poisson distributed with the expected value

defined by the product of a tuning function $f_i(s)$ and integration time τ:

$$R_i \sim \text{Pois}\left[\tau f_i(s)\right] \tag{5.3}$$

The conditional response distribution for the ith neuron, given s, is therefore:

$$P(R_i = r_i | S = s) = \frac{[\tau f_i(s)]^{r_i} \exp\left[-\tau f_i(s)\right]}{r_i!} \tag{5.4}$$

Because the responses of each neuron are conditionally independent, the population response distribution, conditioned upon s, is given by:

$$P(R = \boldsymbol{r} | S = s) = \prod_{i=1}^{N} \frac{[\tau f_i(s)]^{r_i} \exp\left[-\tau f_i(s)\right]}{r_i!} \tag{5.5}$$

Two forms of tuning curve are investigated in this chapter: Gaussian unimodal and sigmoidal monotonic. The tuning curve of the ith neuron in a population with unimodal tuning curves is defined as:

$$f_U^i(s) = f_{bg} + f_{mod} \exp\left[-\frac{(s-\sigma_i)^2}{2\omega^2}\right] \tag{5.6}$$

Where f_{bg} and f_{mod} are the background firing rate and firing rate modulation depth ($f_{mod} = f_{max} - f_{bg}$), both in spikes/s, σ_i is the characteristic (peak) stimulus and ω is a width parameter. The monotonic tuning curve of the ith neuron is similarly defined as:

$$f_M^i(s) = f_{bg} + \frac{f_{mod}}{1 + \exp\left(-\frac{s-\sigma_i}{\omega}\right)} \tag{5.7}$$

For this type of tuning curve the characteristic stimulus σ is the midpoint of the sloping flank of the tuning curve. Figure 5.1 shows examples of both types of tuning curve.

5.1.2 Fisher information

The Fisher information $J(s)$ for a neuron with rate tuning curve $f(s)$, integration time τ, and Poisson variability (Seung and Sompolinsky, 1993; Bethge et al., 2002) is given by:

$$J(s) = \tau \frac{f'(s)^2}{f(s)} \tag{5.8}$$

The units of the Fisher information are A^{-2}, where A are the arbitrary units of the stimulus variable. The tuning curve derivatives $f'_U(s)$ for unimodal turning curves and $f'_M(s)$ for monotonic tuning curves are given by:

$$f'_U(s) = f_{mod} \frac{\sigma_i - s}{\omega^2} \exp\left[-\frac{(s-\sigma_i)^2}{2\omega_i^2}\right] \tag{5.9}$$

$$f'_M(s) = \frac{f_{mod}}{2\omega\left[\cosh\left(\frac{\sigma_i - s}{\omega}\right) + 1\right]} \tag{5.10}$$

Figure 5.1 includes examples of the Fisher information for both unimodal and monotonic tuning curves.

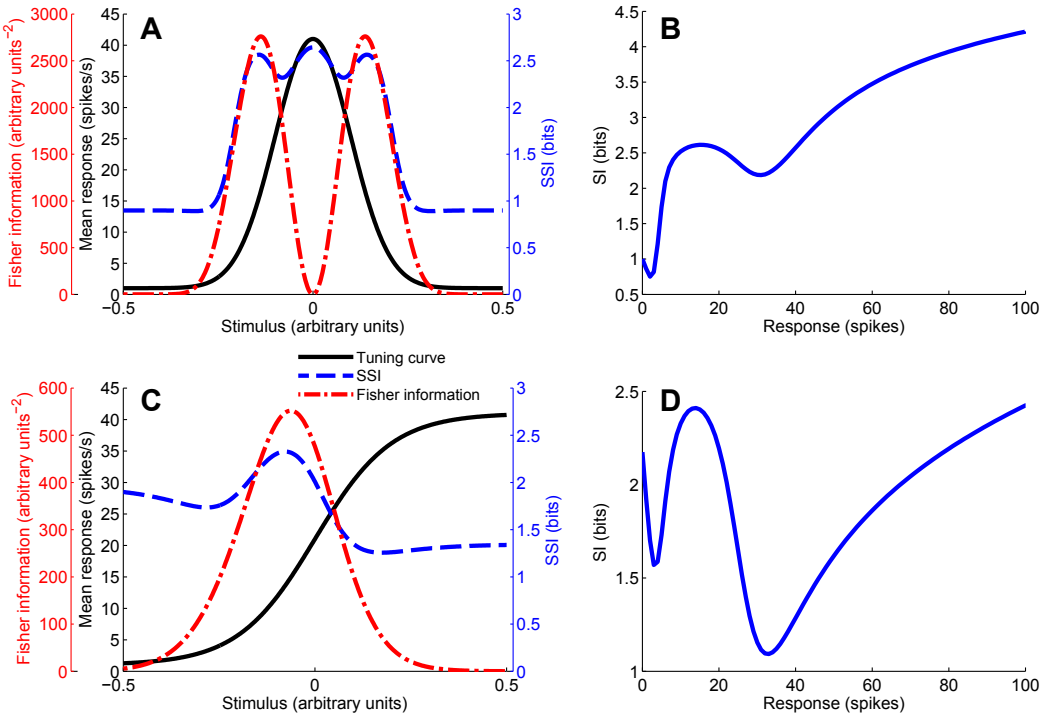

Figure 5.1: Examples of tuning curves, SSI, SI, and Fisher information. Unimodal tuning curve for a single neuron (**A**), together with the SSI and Fisher information. Note that the Fisher information is greatest at the flanks of the tuning curve, while the SSI at the peak and flanks is roughly equal in this case (close to the transition between peak and flank coding regimes). (**B**) shows the response specific information for the tuning curve shown in (**A**). Example of a monotonic tuning curve for a single neuron (**C**), along with its SSI and Fisher information; here both the SSI and Fisher information predict similar best-encoded stimuli. (**D**) gives the response specific information for the tuning curve shown in (**C**). Parameters: $N = 1$, $f_0 = 1\,\text{spike/s}$, $f_{mod} = 40\,\text{spikes/s}$, $\tau = 1\,\text{s}$, $\omega = 0.1$, $\sigma = 0$, stimulus ensemble consists of 401 equally probable discrete stimuli regularly spaced in the interval $[-1, 1]$.

5.1.3 Stimulus-specific information

The specific information (SI; DeWeese and Meister, 1999) is the average reduction in uncertainty about the stimulus (reduction in stimulus entropy) associated with observing a neuronal response:

$$I_{SI}(r) = \sum_{s \in S} p(s|r) \log p(s|r) - p(s) \log p(s) \tag{5.11}$$

$$= H(S) - H(S|R = r) \tag{5.12}$$

It is often more useful to be able to quantify the information associated with a given stimulus, rather than a response. The stimulus-specific information $I_{SSI}(s)$ (Butts, 2003) is the expected value of the SI associated with a given stimulus i.e. the SI averaged over the conditional response distribution. Examples of the specific information and stimulus-specific information for unimodal and

monotonic tuning curves are shown in Figure 5.1. The SSI is defined as:

$$I_{SSI}(s) = \sum_{r \in R} p(r|s) I_{SI}(r) \tag{5.13}$$

$$= \sum_{r \in R} p(r|s) \left[\sum_{s' \in S} p(s'|r) \log p(s'|r) - p(s') \log p(s') \right] \tag{5.14}$$

The precision of the modelled population codes was always set (by choosing appropriate f_0, f_{mod} and τ) such that the mutual information between stimulus and response did not saturate the stimulus entropy, as this would flatten the SSI, making it difficult to identify the best-encoded stimulus. We compute the SSI using the Monte Carlo method described in Appendix A.3.2. This approach removes the necessity to exhaustively integrate over the high-dimensional response ensemble and makes it feasible to compute the SSI for populations of hundreds of neurons.

5.1.4 Chernoff distance

We use the Chernoff distance (Chernoff, 1952, Equation 2.13) to quantify the dissimilarity between the response distributions elicited by two stimuli, which gives an indication of the discriminability of the stimuli. The Chernoff distance is linked to both the mutual information between stimulus and response and the error rate in a two-alternative discrimination task (Kang and Sompolinsky, 2001). The Chernoff distance is also related to the Fisher information, as it has been shown that they are approximately proportional for fine discrimination tasks (Cover and Thomas, 2006; Kang et al., 2004). Although the Chernoff distance has been used to quantify the precision of population codes as a function of the distance between stimuli in a discrimination task (Kang et al., 2004), it has not previously been used to predict best-encoded stimuli; in this thesis we explore the latter application of the Chernoff distance.

Much of the computational complexity in evaluating the Chernoff distance arises from the need to integrate over the response ensemble for every iteration of the maximisation on the exponent α. This means that, as the population size increases, it rapidly becomes very time consuming to compute the Chernoff distance. However, it is possible in many cases to avoid this expensive calculation. Firstly, a closed-form solution exists for the Chernoff distance between two univariate Poisson distributions (Johnson and Sinanović, 2001; Nielsen, 2013), here with parameters λ_1 and λ_2:

$$D_C(\lambda_1, \lambda_2) = \lambda_1 \frac{(\Lambda - 1)\left(\log \frac{\Lambda - 1}{\log \Lambda} - 1\right) + \log \Lambda}{\log \Lambda} \qquad \text{where } \Lambda = \frac{\lambda_2}{\lambda_1} \tag{5.15}$$

The multivariate case, however, is a little more complex. A recently proposed information geometric method (Nielsen, 2013) provides a way to perform an alternative optimisation with reduced computational complexity; this method

involves computing Bregman divergences in the natural parameter space of the distributions. Following Nielsen's univariate Poisson example (Nielsen, 2013), the natural parameters for the joint distribution of N independent Poisson variables with means $\boldsymbol{\lambda} = \{\lambda_1 \dots \lambda_N\}$ are $\boldsymbol{\theta}(\boldsymbol{\lambda})$, where $\theta_i(\lambda_i) = \log \lambda_i$. The Bregman divergence between two such distributions with natural parameters $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$ is given by:

$$B(\boldsymbol{\theta}, \boldsymbol{\theta}') = \sum_{i=1}^{N} \exp \theta_i' - \sum_{i=1}^{N} \exp \theta_i - \sum_{i=1}^{N} (\theta_i' - \theta_i) \exp(\theta_i) \tag{5.16}$$

Considering the distributions of responses to two stimuli s_1 and s_2, with corresponding natural parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, there exists a point $\boldsymbol{\theta}$ on the line joining $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ such that $B(\boldsymbol{\theta}_1, \boldsymbol{\theta}) = B(\boldsymbol{\theta}_2, \boldsymbol{\theta}) = D_C(s_1, s_2)$. The intermediate point $\boldsymbol{\theta}$ can be expressed as a weighted average, where α controls the weighting: $\boldsymbol{\theta} = \alpha \boldsymbol{\theta}_1 + (1 - \alpha)\boldsymbol{\theta}_2$. Computing the Chernoff distance is thus reduced to a bisection search involving the calculation of two Bregman divergences (Equation 5.16) per iteration. The implementation was verified by cross checking against directly calculated Chernoff distances for univariate test cases.

5.1.5 Quantifying similarity in the shape of information measures

It is often useful to compare the shapes of informational quantities that are both functions of the stimulus, but have different units, e.g. the SSI and the Fisher information. To do this, we first discretise the stimulus space (in every case 201 points were used) then compute the value of both measures at each stimulus value. We can then treat the resulting discretised functions of the stimulus as vectors, where the shape of the function is equivalent to the direction of the vector. Similarity of shape can then be quantified by normalising the vectors to unit length and taking the dot product. Using the SSI as an example, the normalised SSI $\hat{I}_{SSI}(s)$ is given by:

$$\hat{I}_{SSI}(s) = \frac{I_{SSI}(s)}{\sqrt{\sum_{s \in S} I_{SSI}^2(s)}} \tag{5.17}$$

If, for example, the Fisher information $J(s)$ is similarly normalised, then the normalised dot product is given by:

$$\hat{I} \bullet \hat{J} = \sum_{s \in S} \hat{I}_{SSI}(s) \hat{J}(s) \tag{5.18}$$

This is a scalar measure of shape similarity, where a value of one indicates that the two functions have identical shape, i.e. that they are directly proportional to one another.

5.1.6 Task modelling

Two-alternative forced choice tasks are frequently used in experimental neuroscience because of their simplicity. In this chapter, two-alternative stimulus

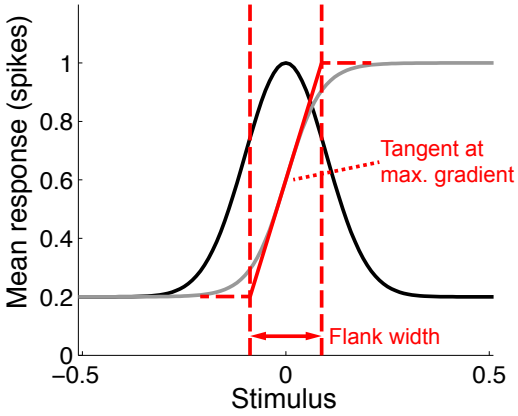

Figure 5.2: Unimodal and monotonic tuning curves of equivalent width. The width parameter ω must be set differently for each form of tuning curve model to obtain equal flank gradients, and hence equal Fisher information maxima and flank widths. The flank width, as indicated, is used when calculating the normalised stimulus spacing δs for forced choice tasks. Parameters: $\omega = 0.1$ (unimodal), $\omega = 0.044$ (monotonic).

ensembles were modelled in order to find out which stimuli were best encoded by a neuron for a given stimulus spacing Δs. In the model the absolute values of the stimuli (the choices) are not fixed, only the distance between choices Δs appears as a parameter. This can be considered equivalent to an experimental setup where the stimulus choices are fixed relative to one another, but their position can vary with respect to the tuning curves, such as in a visuospatial task without predefined gaze fixation or an auditory source discrimination task where the head is free to move.

Generalising from two-alternative to K-alternative tasks, we assume that the stimulus ensemble consists of K distinct stimuli regularly spaced at intervals of Δs, and that the set of K stimuli can be translated anywhere in the stimulus space. Any concrete stimulus value s can be a member of the stimulus ensemble when the ensemble is translated to K different positions, as s can be the first, second, third, etc., stimulus in the set. We denote each SSI component as $I_k(s)$, where s is the kth member of the stimulus ensemble. For example, if $K = 2$ we have two components of the SSI:

$$I_1(s) = I(s; S = \{s, s + \Delta s\}) \tag{5.19}$$

$$I_2(s) = I(s; S = \{s - \Delta s, s\}) \tag{5.20}$$

We assume that each of the cases is equally likely, and compute the SSI as the simple average over the K components:

$$I_{SSI}(s) = \frac{1}{K} \sum_{k=1}^{K} I_k(s) \tag{5.21}$$

The same method was used to compute the Chernoff distance as a function of the stimulus (see Figure 5.3 for an example of how the combined Chernoff distance is

computed), but here the number of stimuli in the set is fixed at $K = 2$ due to the inherent limitations of the measure.

Note that this is slightly different to the method of averaging used by Butts and Goldman. In Figure 4 of Butts and Goldman (2006), the SSI at stimulus s is given by $I_{SSI}(s) = \frac{I(s-\Delta s/2)+I(s+\Delta s/2)}{2}$, where the stimulus ensemble is $S = \{s-\Delta s/2, s+\Delta s/2\}$. As this is an average over the stimulus ensemble it is equal to the mutual information between the stimulus and responses, given that stimulus ensemble. The two methods of averaging give very similar results for small Δs, but differ at larger Δs. For example, the Butts and Goldman method of averaging always yields zero information at the peak of a symmetric tuning curve, as the responses to $s-\Delta s/2$ and $s+\Delta s/2$ are identical regardless of the value of Δs. The method of averaging used here has the advantage that it gives the information I(s) for the exact stimulus s, rather than stimuli offset from s by some amount. It is therefore more strictly linked to s, which makes it better suited for assessing the best-encoded stimulus when Δs is large, while remaining equally valid for small values of Δs.

When modelling forced choice tasks it is sometimes useful to give the interval between stimuli Δs in a normalised form δs, such that it is expressed relative to the width of the tuning curve flank:

$$\delta s = \Delta s \frac{f_{mod}}{\max f'(s)} \tag{5.22}$$

Where f_{mod} and $\max f'(s)$ are, respectively, the modulation depth and maximum gradient of the tuning curve. $\Delta s = |s_2 - s_1|$ is the stimulus interval and s_1, s_2 are adjacent stimuli.

5.1.7 Visualising uncertainty in the neural code

In any probabilistic code, information is lost when symbols (here neural responses) are ambiguous, when it is unclear which stimuli caused them. When studying best-encoded stimuli and the shape of the SSI and similar measures, it is useful to be able to visualise this ambiguity; the 'confusion' between stimuli that is introduced by the code. We do this by considering a hypothetical observer of the population response with no direct knowledge of the stimulus, who tries to infer s from the responses $\boldsymbol{r}$ to a single stimulus presentation. We represent the observer's knowledge of the stimulus by the random variable Z, and assume that the observer has full knowledge of the stimulus distribution $P(S)$ and the stochastic encoding scheme $P(R|S)$. For simplicity, we assume that the observer's prior $P(Z)$ is equal to the true stimulus distribution $P(S)$ (although this need not be the case; any prior could be modelled). We are interested in the distribution $P(Z|S)$, as this tells us what the hypothetical observer (or, equivalently, any downstream neural information processing element) can infer on average about the stimulus, following the presentation of a given stimulus. Although the SSI is closely related to the

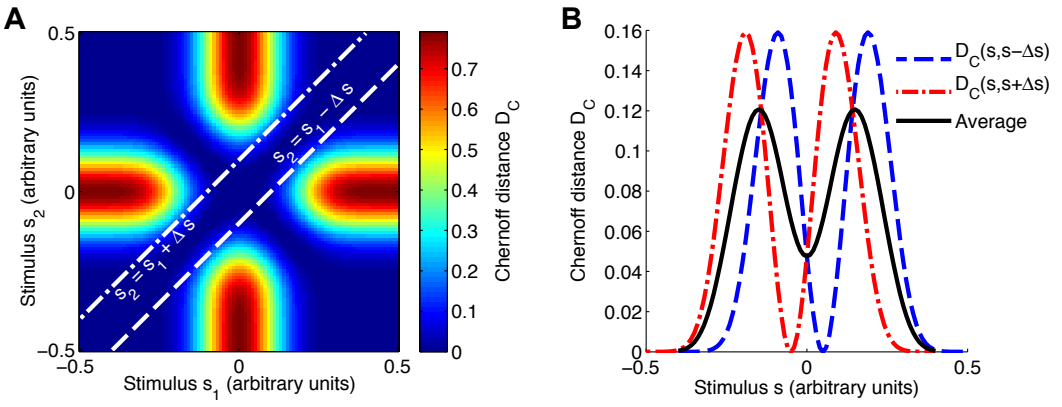

Figure 5.3: Evaluating the best-encoded stimulus for a 2AFC task using the Chernoff distance. Panel (**A**) shows a map of the Chernoff distance between the response distributions for every possible pair of stimuli. The white lines overlaid on the heat map indicate where the difference between stimuli s_1 and s_2 is equal to the stimulus spacing of interest, in this case $\Delta s = 0.1$. In a 2AFC task with a given stimulus spacing Δs, a stimulus s can occur in two ways: paired with $s - \Delta s$ or with $s + \Delta s$; these two possibilities correspond to the two white lines on (**A**). The interrupted red and blue curves in (**B**) correspond to slices through the map shown in (**A**); $D_C(s, s+\Delta s)$ corresponds to the line $s_2 = s_1 + \Delta s$ and $D_C(s, s-\Delta s)$ to $s_2 = s_1 - \Delta s$. We construct an information tuning curve (solid black line) based on the Chernoff distance by taking the average of these two components. Parameters: $N = 1$, $f_0 = 1\,\text{spike/s}$, $f_{mod} = 40\,\text{spikes/s}$, $\tau = 50\,\text{ms}$, $\omega = 0.1$, $\sigma = 0$, stimulus ensemble consists of 101 equally probable discrete stimuli regularly spaced in the interval $[-0.5, 0.5]$.

uncertainty in $P(Z|S)$—i.e. the conditional entropy $H(Z|S = s)$—it is important to note that these measures are not the same (see Appendix A.4 for further details).

$P(Z|S)$ is computed as follows.

$$P(z|s) = \sum_{\boldsymbol{r} \in R} P(z|r)P(r|s) \tag{5.23}$$

Where $P(z|r) = P(s|r)$, since our observer has full knowledge of the encoding process—the model is symmetrical. $P(s|r)$ is obtained by applying Bayes' theorem:

$$P(s|r) = \frac{P(r|s)P(s)}{P(r)} = \frac{P(r|s)P(s)}{\sum_{s \in S} P(r|s)P(s)} \tag{5.24}$$

For single neurons or very small populations the sum over R is straightforward, but it becomes intractable as the population size increases. To overcome this, we used a Monte Carlo method identical to that used when calculating the SSI (Appendix A.3.2).

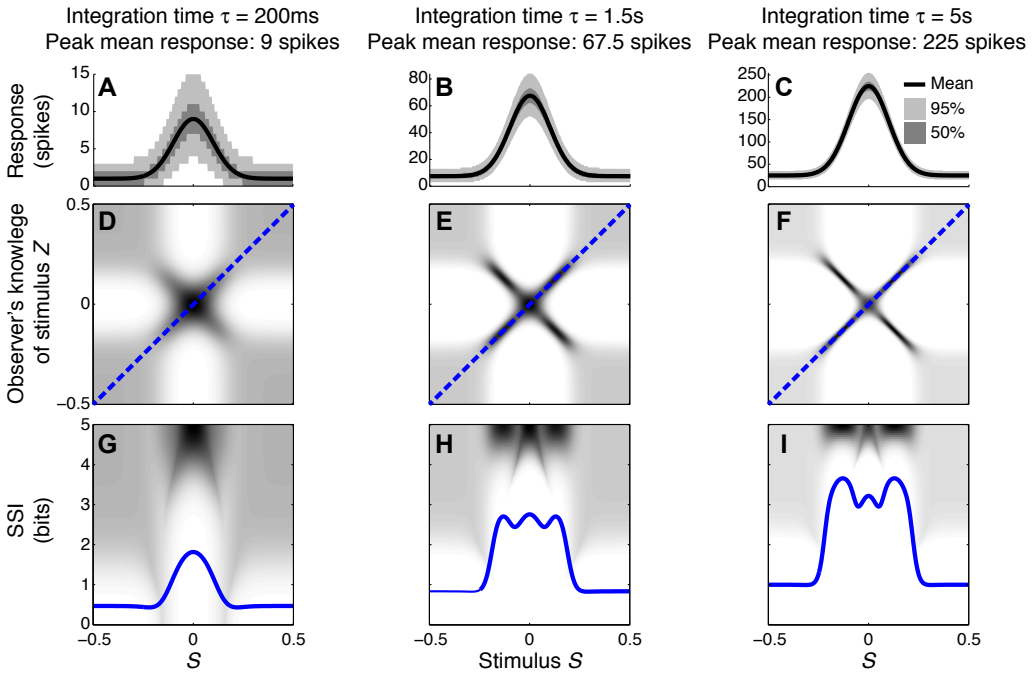

Figure 5.4: Bayesian reconstruction with a unimodal tuning curve. **(A)**–**(C)** Tuning curves and trial-to-trial variability for a single neuron at three different integration times and hence three different levels of trial-to-trial variability. Variability is illustrated by shading: the dark region contains half of the probability mass (25th to 75th percentile), while the light region extends down to the 2.5th percentile and up to the 97.5th percentile and, together with the dark region, contains 95% of the probability mass. **(D)**–**(F)** $P(Z|S)$, the average posterior stimulus distribution after observing the response to a single trial, conditioned on the true stimulus, for the tuning properties shown in **(A)**–**(C)**. Darker shading indicates higher probability and, for clarity, the shading density is scaled independently for each panel and does not span the full $[0,1]$ interval. This distribution can be read like a lookup table: if the true stimulus is selected on the horizontal axis then the average posterior distribution is given by that column. The more information conveyed by the neuron, the more precise the reconstruction and hence the more the response probability mass is concentrated close to the blue dashed line that corresponds to exact reconstruction ($Z = S$). **(G)**–**(I)** The SSI for the same three cases overlaid upon the distributions shown in **(D)**–**(F)**, where each column of the distribution is sorted so that the highest probability bins are uppermost. This illustrates more clearly the relationship between the SSI and the amount of uncertainty (i.e. the entropy $H(Z|S=s)$) in the distribution $P(Z|S)$; greater SSI corresponds to more precise reconstruction and hence lower posterior entropy. The transition between peak coding and flank coding, according to the SSI, occurs close to the case shown in panels B,E and H where the uncertainty associated with the low-gradient peak region of the tuning curve is equal to that due to ambiguity between the two symmetric flanks. Parameters: $N = 1$, $f_{mod} = 40\,\text{spikes/s}$, $f_0 = 5\,\text{spikes/s}$, $\omega = 0.1$, $\sigma = 0$, stimulus ensemble consists of 401 equally probable discrete stimuli regularly spaced in the interval $[-0.5, 0.5]$.

5.2 Results

5.2.1 The effect of trial-to-trial variability in single neurons

5.2.1.1 Unimodal tuning curves

In neurons with unimodal tuning curves, the best-encoded stimulus predicted by the SSI depends both on the level of trial-to-trial variability (noise) and on the number of neurons in the population (Butts and Goldman, 2006, and Chapter 4). The best-encoded stimuli according to the SSI can lie on the flanks of the tuning curves, or at the peaks of the tuning curves (in small populations with high noise or short integration times), suggesting that there may be two distinct coding regimes.

To illustrate how and why the best-encoded stimuli predicted by the SSI change with the trial-to-trial variability, we simulated a single unimodally-tuned neuron and manipulated the level of variability by changing the integration time. Under the Poisson variability model, as time passes following the presentation of the stimulus the mean spike count increases in proportion to the elapsed time τ, but the standard deviation only increases as $\sqrt{\tau}$, so the signal to noise ratio also increases in proportion to $\sqrt{\tau}$ (Figure 5.4A–C). Panels D–F of Figure 5.4 show the expected posterior distribution on the stimulus conditioned on the true stimulus ($P(Z|S)$; we use the variable Z to represent the observer's knowledge of the stimulus to avoid confusion with the true stimulus S; see section 5.1.7 for further details). This posterior distribution describes the knowledge that a hypothetical observer (e.g. a downstream neuron) has, on average, about the stimulus after receiving the output from our model neuron in a single trial. The symmetric cross shape of the distribution is due to the symmetry of the unimodal tuning curve: because of the ambiguity between the two halves of the tuning curve, it is not possible to distinguish stimuli on one side of the peak from those on the other based on the responses of that neuron alone. In the context of a population, however, the ambiguity between tuning curve flanks tends to be resolved by the responses of the other neurons. Although the gradient of the tuning curve is greatest on its flanks, the inter-flank ambiguity greatly increases the conditional entropy of the posterior. While the distribution is visibly more concentrated around $Z = S$ and $Z = -S$ (i.e. precisely correct decoding) when the variability is lower, there is no obvious qualitative change in the distribution between Figure 5.4D and F that accounts for the change in best-encoded stimulus. The reason for the difference in best-encoded stimuli becomes clearer when each column of the distribution is sorted so that the amount of uncertainty (i.e. the conditional entropy $H(Z|S = s)$) can be seen more easily (Figure 5.4G–H). Through this visualisation, we can see that the transition between peak and flank coding occurs where the uncertainty associated with the two flanks is equal to that of the flat, but narrow, peak (Figure 5.4B, E, and H).

The peak of a tuning curve is associated with a uniquely high mean response, which is more different from the background response than any other point on the tuning curve. Responses to stimuli around the peak are therefore always informative as they allow coarse discrimination of the peak from untuned background activity, and this contribution to the overall information dominates when the variability is high (e.g. Figure 5.4A, D, and G). In order to discriminate between closely neighbouring stimuli and hence estimate the stimulus more precisely, the responses associated with those stimuli must be different. This fine discriminability is quantified by the Fisher information and is maximal where the gradient of the tuning curve is greatest. When the variability is low, fine discrimination dominates, with the result that the best-encoded stimulus shifts to the flanks of the tuning curve (Butts and Goldman, 2006, and Figure 5.4C, F, and I).

5.2.1.2 Monotonic tuning curves

To investigate whether similar distinct coding regimes could also exist for monotonic tuning curves, we calculated the SSI for a single sigmoidally-tuned model neuron with Poisson trial-to-trial variability. The tuning curve parameters were fixed and the level of variability was manipulated by changing the integration time. Figure 5.5 shows the SSI and the SI for the model neuron at several different integration times from 5 ms to 1 s. At very short integration times (very high trial-to-trial variability) the SI increases monotonically with increasing spike count (Figure 5.5A) and the SSI is almost proportional to the tuning curve (Figure 5.5F). The mean responses to all stimuli are less than one spike, therefore the Poisson response distributions for all stimuli are monotonically decreasing with maxima at zero spikes. At this point in most trials no spikes have occurred yet, regardless of the stimulus, so zero spike counts are uninformative, as indicated by their low SI. Non-zero responses, however, are more likely to have been caused by stimuli at the high-responding region of the tuning curve (we refer to this as the tuning curve plateau), so these have higher SI. As the integration time increases and the mean responses to plateau stimuli increase above one spike (Figure 5.5B and C), these response distributions are no longer monotonic, but peaked around the expected response. This means that very low spike counts are indicative of stimuli on the non-selective, low-responding 'baseline' region of the tuning curve, while high responses are more likely to be caused by stimuli on the plateau. Both high and very low spike counts therefore have relatively high SI, and there is a trough in the SI curve at intermediate responses (Figure 5.5B and C) and a corresponding trough in the SSI at the flank (sloping region) of the tuning curve (Figure 5.5G and H). This is the coarse discrimination regime for monotonic tuning curves, analogous to the peak coding regime for unimodal tuning curves. Here the neuronal responses support discrimination of baseline from plateau stimuli, but are too noisy

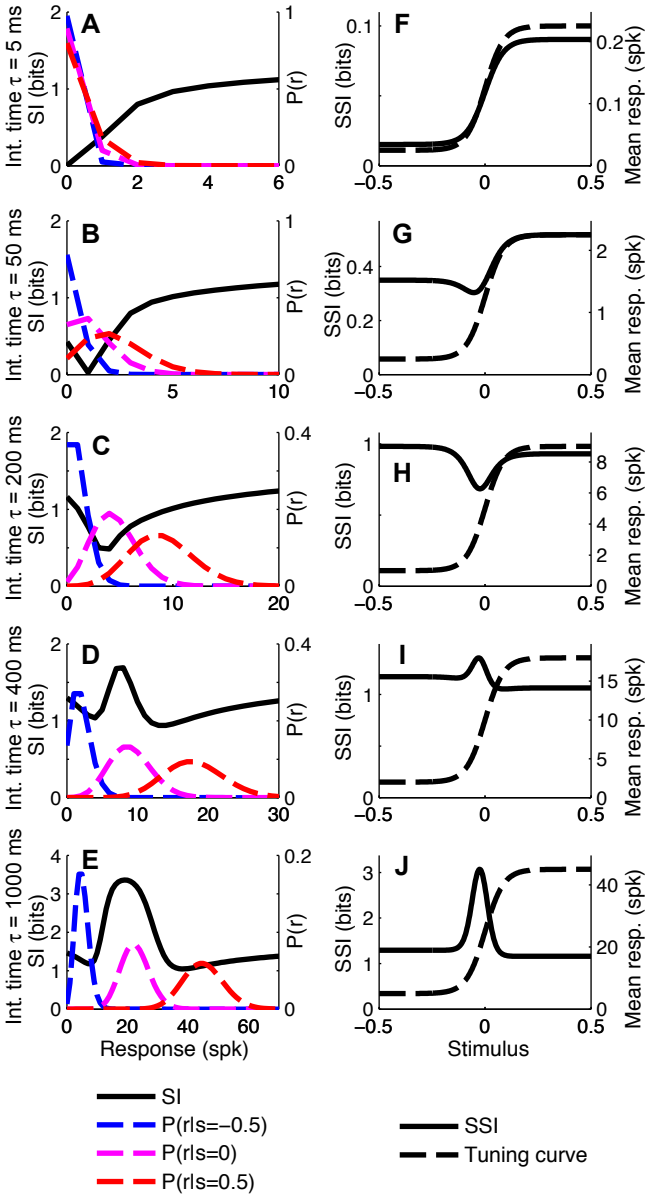

Figure 5.5: SSI for a sigmoidal monotonic tuning curve. (**A**)–(**E**) Specific information (SI), together with response distributions for low, mid and high responding regions of the tuning curve. These are shown for five different levels of trial-to-trial variability determined by the integration time τ; short integration times lead to high variability and vice versa. The differences in trial-to-trial variability are best illustrated by the increasing separation between the high, mid and low response distributions. (**F**)–(**J**) The SSI for the same five sets of parameter values. Parameters: $N = 1$, $f_{mod} = 40\,\text{spikes/s}$, $f_0 = 5\,\text{spikes/s}$, $\omega = 0.044$, $\sigma = 0$, stimulus ensemble consists of 401 equally probable discrete stimuli regularly spaced in the interval $[-0.5, 0.5]$.

to allow neighbouring stimuli to be distinguished from one another. This can be seen in the conditional posterior distribution $P(Z|S)$; the chequerboard pattern in Figure 5.6D means that stimuli can be decoded as being from either the baseline or plateau regions of the tuning curve, but stimuli from within either region are indistinguishable. The distributions of responses to stimuli on the flank of the tuning curve overlap with those of the baseline and plateau, so these intermediate responses could have been caused by any stimulus and are therefore uninformative. This is clearly visible as the dark central band in Figure 5.6G, which shows that the posterior distribution for stimuli on the flank of the tuning curve is broad rather than tightly peaked around the correct stimuli. Within the coarse discrimination regime the relative values of the SSI for the baseline and plateau regions are largely determined by the characteristic stimulus of the neuron: the position of the tuning curve flank within the stimulus space. If the plateau is relatively narrow then high spike counts identify a smaller range of stimuli than low spike counts and the SSI is at its greatest on the plateau. Conversely, and perhaps unintuitively, if the plateau is broader than the baseline region then low spike counts are more informative and the maximum SSI occurs on the baseline region (results not shown). When the flank is located centrally in the stimulus space, the SSI of the plateau and baseline regions are similar, as in Figure 5.5H.

As the integration time is increased and the trial-to-trial variability consequently decreases, the SI for intermediate spike counts (Figure 5.5D) and the SSI at the flank of the tuning curve (Figure 5.5I), increase and become maxima (both were minima at shorter integration times). The best-encoded stimulus is now at the flank of the tuning curve. This is the emergence of the fine discrimination regime; here the trial-to-trial variability is sufficiently low to allow discrimination between stimuli on the flank of the tuning curve. The gradient of the tuning curve is greatest in the centre of the flank, and it is here that the Fisher information is greatest and the response distributions of adjacent stimuli are most different. The relatively high precision of decoding on the flank can be seen in the conditional posterior distributions (Figure 5.6E and H) as a concentration of posterior probability mass around the correct stimulus ($Z = S$). Further increases in integration time increase the maximum values of the SI (Figure 5.5E) and SSI (Figure 5.5J; see also Figure 5.6F and I), but the best-encoded stimulus remains the same.

5.2.1.3 Summary

Monotonic tuning curves exhibit distinct coding regimes analogous to the peak and flank regimes of unimodal tuning curves. When the trial-to-trial variability is low, fine discrimination dominates and the maximum SSI is on the flank of the tuning curve, where the Fisher information is also maximal, and hence both measures predict similar best-encoded stimuli. For high variability, coarse discrimination

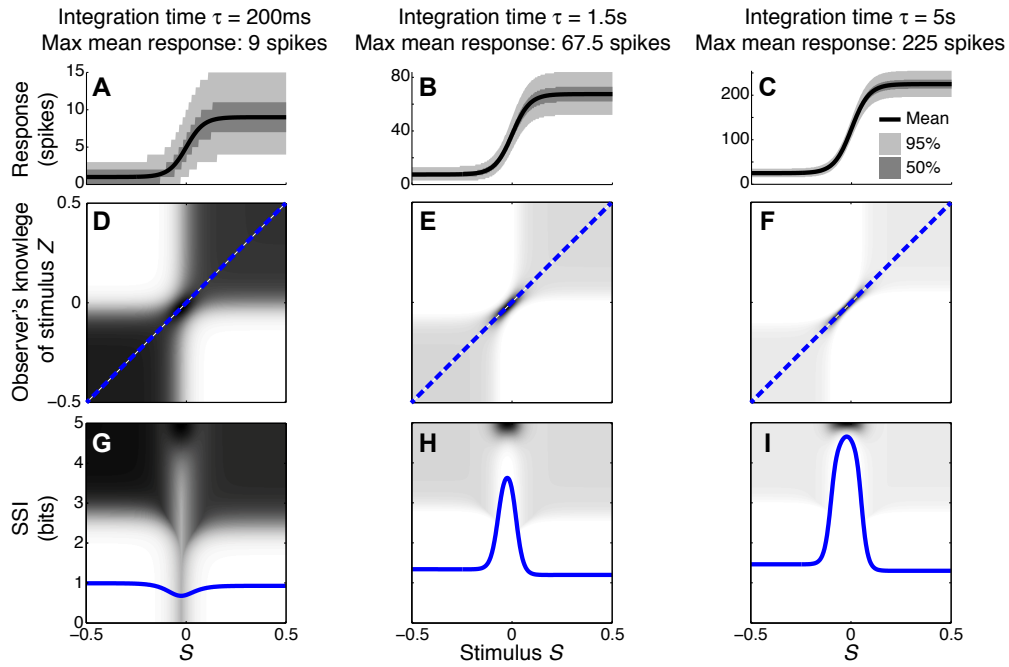

Figure 5.6: Bayesian reconstruction with a monotonic tuning curve. **(A)**–**(C)** Tuning curves and trial-to-trial variability for a single neuron at three different integration times and hence three different levels of trial-to-trial variability. Variability is illustrated by shading: the dark region contains half of the probability mass (25th to 75th percentile), while the light region extends down to the 2.5th percentile and up to the 97.5th percentile and, together with the dark region, contains 95% of the probability mass.
(D)–**(F)** $P(Z|S)$, the average posterior stimulus distribution after observing the response to a single trial, conditioned on the true stimulus, for the tuning properties shown in **(A)**–**(C)**. Darker shading indicates higher probability and, for clarity, the shading density is scaled independently for each panel and does not span the full $[0,1]$ interval. This distribution can be read like a lookup table: if the true stimulus is selected on the horizontal axis then the average posterior distribution is given by that column. The more information conveyed by the neuron, the more precise the reconstruction and hence the more the response probability mass is concentrated close to the blue dashed line that corresponds to exact reconstruction ($Z = S$).
(G)–**(I)** The SSI for the same three cases overlaid upon the distributions shown in **(D)**–**(F)**, where each column of the distribution is sorted so that the highest probability bins are uppermost. This illustrates more clearly the relationship between the SSI and the amount of uncertainty (i.e. the entropy $H(Z|S=s)$) in the distribution $P(Z|S)$; greater SSI corresponds to more precise reconstruction and hence lower posterior entropy. Two contrasting coding regimes can be seen: a Fisher-like regime (best-encoded stimulus at the steepest part of the tuning curve (H and I), and a regime where the SSI is relatively flat, with minimum information occurring at the steep region of the tuning curve (**G**).
Parameters: $N = 1$, $f_{mod} = 40\,\text{spikes/s}$, $f_0 = 5\,\text{spikes/s}$, $\omega = 0.044$, $\sigma = 0$, stimulus ensemble consists of 401 equally probable discrete stimuli regularly spaced in the interval $[-0.5, 0.5]$.

dominates and the best encoded stimulus is less easy to predict. Monotonic tuning curves differ from unimodal tuning curves in that the high-response region is broad rather than localised. This means that, as described above, strong responses are not always informative (as is the case for peaked tuning curves), as any stimulus on the upper plateau of the tuning curve is likely to generate a strong response. The best-encoded stimulus for a single monotonically-tuned neuron is not clearly defined, as maximum SSI can occur over the whole of the plateau or the whole of the baseline region. Another important difference between monotonic and unimodal tuning curves is symmetry: as monotonic tuning curves have only a single flank, each stimulus on the flank is unambiguously associated with a unique response distribution. This means that, for a given tuning curve gradient and variability (equal Fisher information), responses to stimuli on the tuning curve flank are more informative than they would be for a single neuron with a unimodal tuning curve. For example, compare the SSI peaks close around the threshold stimulus in Figure 5.6H and I (approximately 3.5 and 4.5 bits respectively) with the SSI peaks at the tuning curve flanks in Figure 5.4H and I (approximately 2.5 and 3.5 bits).

5.2.2 The effect of population size

Population size is an important determinant of the magnitude and shape of the SSI for population coding neurons, and hence of the best-encoded stimulus. To quantify the information contributed by a neuron in the context of a population code, we use the marginal SSI (mSSI). The mSSI of a neuron is the difference between the SSI of the entire population and the SSI of the population without the neuron of interest. It is useful to think of the singleton SSI and marginal SSI as upper and lower bounds, respectively, on the informational contribution of a neuron to a population code. One way to understand this is to imagine building up a population by adding one neuron at a time. Assuming that the code is redundant, each successive neuron will result in a smaller and smaller additions to the population SSI: the first increment is the singleton SSI and the last is the marginal SSI.

5.2.2.1 Unimodal tuning curves

Previous research on unimodal tuning curves showed that increasing the number of neurons in the population shifts the code towards the asymptotic regime, where Fisher information accurately predicts both the mutual information between stimulus and response (Brunel and Nadal, 1998), and the best-encoded stimulus predicted by the mSSI (see Chapter 4). In large populations the best-encoded stimuli predicted by the mSSI lie on the flanks of the tuning curves and coincide

with the Fisher information maxima. In small populations with high trial-to-trial variability, the best-encoded stimuli according to the mSSI are at the peaks of the tuning curves. The number of neurons at which the transition between the peak and flank coding regimes occurs depends upon the variability: if the variability is low, each neuron conveys a correspondingly higher amount of information and the peak to flank transition occurs at relatively low N. The higher the variability, the larger the population must be before it enters the asymptotic regime. Other factors, including the strength and structure of correlations in the trial-to-trial variability, also affect the population size at which the asymptotic regime is reached, but these effects are relatively small (see Chapter 4 for details).

5.2.2.2 Monotonic tuning curves

To study how population size affects the best-encoded stimulus for monotonic tuning curves, we calculated the mSSI for populations of N model neurons with stereotypical sigmoidal tuning curves and Poisson trial-to-trial variability (again controlled by the integration time τ). The results show that, as for unimodal tuning curves, the shape of the mSSI converges towards the shape of the singleton Fisher information as the number of neurons in the population increases (Figure 5.7A). In the asymptotic regime, the best-encoded stimulus is on the flank of the tuning curve (Figure 5.7H). The rate with which the shape of the mSSI converges to the shape of the Fisher information with increasing population size depends on the level of variability; as for unimodal tuning curves, the higher the variability, the slower the convergence and the larger the population at which the asymptotic regime is reached (Figure 5.7A). Like the SSI for single neurons, the shape of the mSSI in small populations with high variability is close to that of the tuning curve (Figure 5.7B) and the best-encoded stimuli are those on the tuning curve plateau.

Between these extremes of a small population with high variability and a large population with low variability, the mSSI for a unimodal tuning curve can take on variety of different shapes and consequently the best-encoded stimuli vary widely. The mSSI goes through a predictable progression of shapes as the trial-to-trial variability is decreased (by increasing the integration time), or the population size increased. As we have already seen, the mSSI is sigmoidal for small populations and short integration time (Figure 5.7B). If τ or N is increased slightly, the mSSI for the baseline region of the tuning curve increases and an mSSI peak forms at the lower end of the tuning curve flank (Figure 5.7C–E). This peak in the mSSI becomes sharper and moves towards the the characteristic stimulus (the centre of the tuning curve flank) as τ or N increase (Figure 5.7D–F). At the same time, the mSSI around the tuning curve plateau changes from the flat upper part of a sigmoid into a second peak, which also becomes sharper and migrates towards the characteristic stimulus as τ or N is increased (Figure 5.7C–F). At these intermediate, transitional

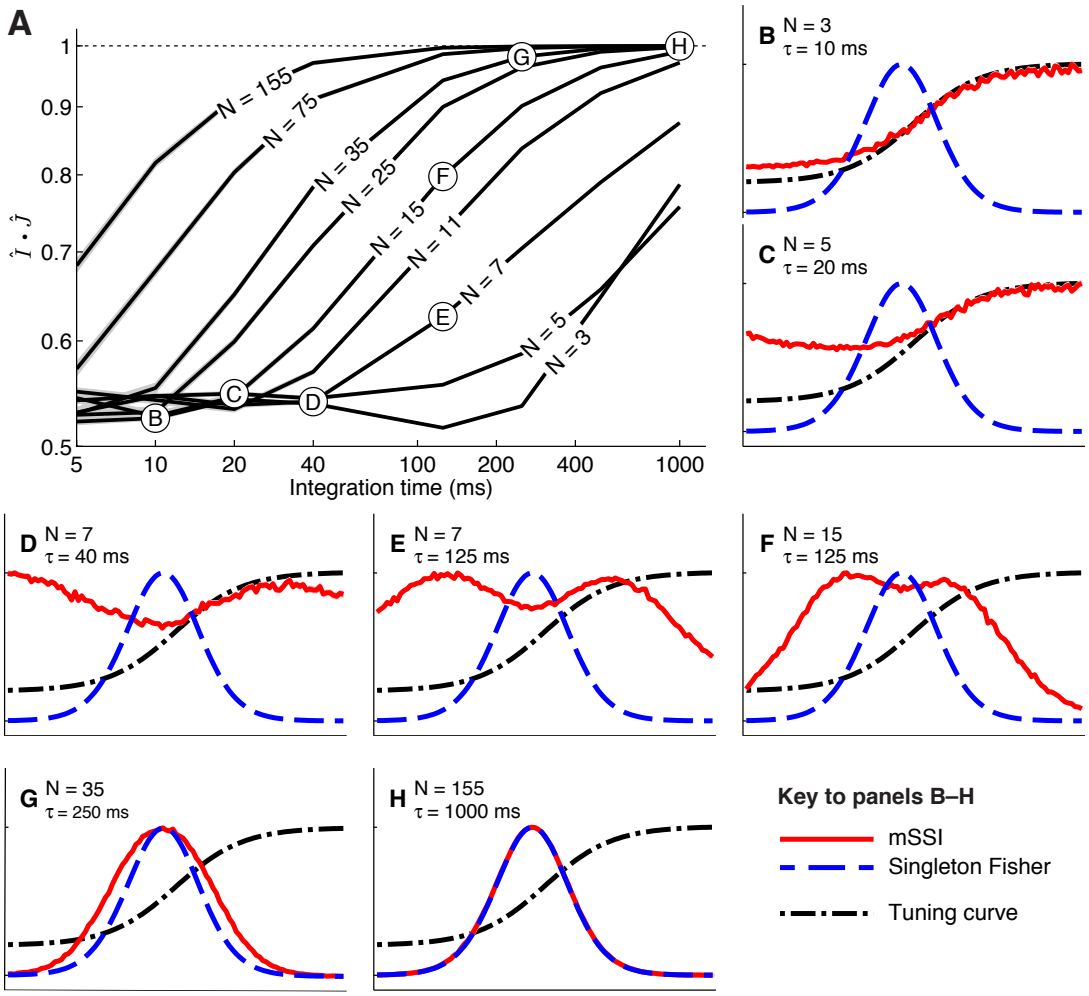

Figure 5.7: Marginal SSI in populations of sigmoidally tuned neurons. (**A**) The normalised dot product of the mSSI and the singleton Fisher information as a function of integration time τ for various population sizes N. This indicates the similarity in shape of the mSSI and the Fisher information; a value of 1 indicates that the mSSI is directly proportional to the Fisher information, while low values indicate that the two measures have dissimilar shapes. (**B**)–(**H**) mSSI as a function of the stimulus for parameter values corresponding to the points labelled in (**A**). The Fisher information and tuning curve for the neuron of interest are also shown. All quantities in (**B**)–(**H**) are normalised. Parameters: $f_{mod} = 40\,\text{spikes/s}$, $f_0 = 10\,\text{spikes/s}$, $\omega = 0.1$, $\sigma = 0$, stimulus variable is continuous in the interval $[-1, 1]$.

values of τ and N, the mSSI is bimodal, with both peaks having approximately equal magnitude. As τ or N is increased further, the two mSSI peaks ultimately merge into a single peak located at the tuning curve flank, coincident with the maximum Fisher information (Figure 5.7E–G). This convergence of the best-encoded stimuli predicted by the mSSI and the Fisher information occurs at population sizes that are modest in the context of the mammalian brain: the shapes of the two measures are close even in populations of the order of 100 neurons at biologically relevant integration times (of the order of 100 ms; Figure 5.7A).

5.2.2.3 Summary

The single neuron SSI for monotonically-tuned neurons is dependent upon the level of trial-to-trial variability and the marginal SSI is dependent upon the variability and the population size. For large populations the best-encoded stimulus lies on the sloping region of the tuning curve, close to the point of maximum gradient, as predicted by the Fisher information. In smaller populations the best-encoded stimulus is dependent on the variability and can be anywhere on, or close to, the sloping part of the tuning curve. These findings are in agreement with those for unimodal tuning curves (Butts and Goldman, 2006, and Chapter 4), but the shape of the SSI—and hence the best-encoded stimulus—is more varied in the case of monotonic tuning curves.

5.2.3 Non-uniform stimulus distributions

For most of the simulations described in this chapter we have assumed that stimuli are uniformly distributed, but this is rarely the case in nature. Even stimulus variables that seem uniformly distributed may not be; for example edge orientation in natural scenes is non-uniformly distributed, with horizontal and vertical contours occurring more frequently than other orientations (Coppola et al., 1998). In particular, monotonic tuning curves imply a linear (as opposed to circular) stimulus space and this, in turn, implies that the stimuli are non-uniformly distributed such that they fall within some finite interval; a uniform distribution within sharply defined limits is extremely unlikely. As the space of possible stimulus distributions is essentially limitless, we did not attempt to examine a wide variety of distribution forms and instead focussed on determining the effect of local non-uniformity of the stimulus distribution on the mSSI of neurons with both unimodal and monotonic tuning curves. To this end, we modelled a broadly peaked stimulus distribution $p(s)$ and considered three neurons of each tuning curve type: one with its characteristic stimulus at the peak of the stimulus distribution and one on either flank (see Figure 5.8A and C). The mSSI of each of the neurons of interest was computed for both the peaked stimulus

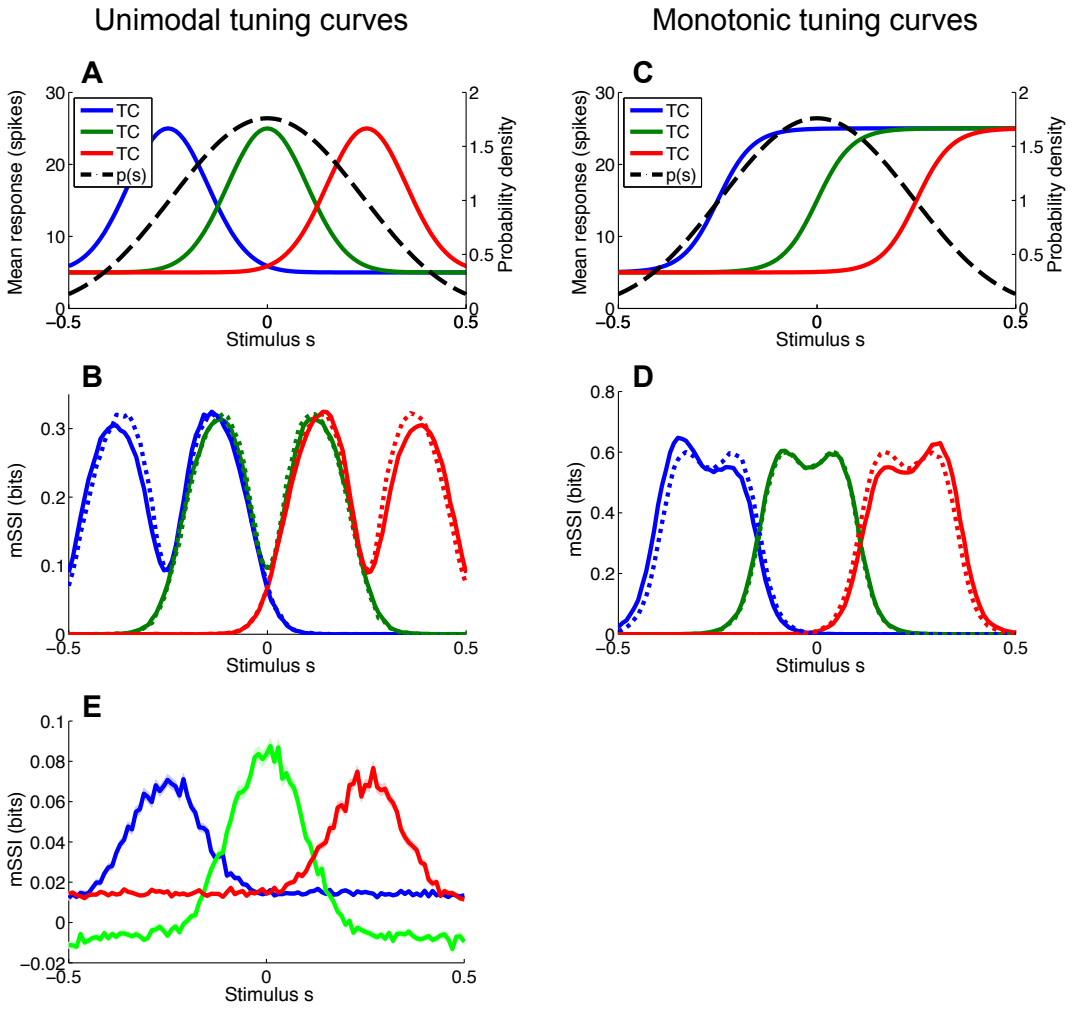

Figure 5.8: Effect of stimulus distribution non-uniformity on the marginal SSI. Non-uniformity of the stimulus distribution can affect the marginal SSI. Here we consider a stimulus distribution that is peaked around zero and compute the mSSI for three neurons, one at the peak of the stimulus distribution and one on either flank. (**A**) Stimulus distribution $p(s)$ and tuning curves (TC) for unimodal neurons of interest. (**B**) The mSSI for the tuning curves and stimulus distribution shown in A (solid lines), together with the mSSI for identical tuning curves and variability, but a uniform stimulus distribution (broken lines). (**C**)–(**D**) As (**A**) and (**B**), but for monotonic tuning curves. (**E**) The mSSI for the same neurons illustrated in (**A**), but with a very short integration time ($\tau = 5\,\mathrm{ms}$). Notice the negative mSSI values coinciding with the tails of the central (green) tuning curve. Uncertainty in the mSSI is indicated by shaded regions of ± 1 StdErr where it is greater than the width of the line. Parameters: $N = 17$, $\tau = 500\,\mathrm{ms}$, $f_{mod} = 40\,\mathrm{spikes/s}$, $f_0 = 10\,\mathrm{spikes/s}$, $\omega = 0.1$ (unimodal), $\omega = 0.044$ (monotonic), stimulus variable is continuous and distributed as shown, or uniformly across the interval $[-1, 1]$.

distribution and a uniform distribution extending well beyond the tuned regions of the neurons (Figure 5.8B and D). The simulations were repeated for a range of population sizes and integration times.

As previously discussed, the mSSI peaks converge towards those of the Fisher information as the population size goes to infinity (see Chapter 4). This holds for non-uniform stimulus distributions, so the stimulus distribution (provided that it is non-zero) has no effect upon the best-encoded stimulus predicted by the mSSI in the limit of large populations (this was confirmed by numerical simulation; results not shown). As we shall see, away from the asymptotic regime the stimulus distribution can strongly affect the best-encoded stimulus, particularly for neurons with monotonic tuning curves.

5.2.3.1 Unimodal tuning curves

The mSSI for unimodal tuning curves is relatively lightly affected by a local gradient in the stimulus probability. For small populations and high variability (i.e. in the peak coding regime) the tuning curve peak defines the best-encoded stimulus, and the stimulus distribution has little or no effect. Therefore, non-uniformity of the stimulus distribution affects the best-encoded stimulus primarily in the transition between the peak and flank coding regimes, where the mSSI is bimodal. The effect of non-uniformity is to skew the mSSI towards greater stimulus probability: the mSSI peak coinciding with the greater stimulus probability is amplified, while the other is suppressed (Figure 5.8B). Thus the best-encoded stimulus is shifted towards the peak of the stimulus distribution.

5.2.3.2 Monotonic tuning curves

For monotonic tuning curves, the effect of non-uniformity in the stimulus distribution is that the mSSI peak with lower stimulus probability is amplified and the peak with higher stimulus probability is attenuated (Figure 5.8D); thus the best-encoded stimulus is shifted away from the most probable stimulus. This is the opposite of the effect that occurs with peaked tuning curves. As is the case for the singleton SSI, it is difficult to predict the shape of the mSSI in small populations with high variability without actually computing it, as it is sensitive to the tuning curve shape, the position of the characteristic stimulus within the stimulus range, as well as to the stimulus distribution.

5.2.3.3 Unexpected stimuli

The presence of unexpected stimuli—those with a low probability of occurrence—in a stimulus ensemble can lead to counter-intuitive SSI values. If much of the

probability mass in the stimulus distribution is concentrated around a subset of common stimuli, then evidence of an unusual stimulus causes an initial increase in the posterior entropy; this can be seen as a negative SSI or mSSI (Figure 5.8E shows an example of this: the mSSI at the tails of the central (green) tuning curve are slightly negative). Although the SSI is transiently negative, the weak evidence provided by the early response does make the posterior more 'correct' i.e. it increases the posterior probability of the true stimulus relative to other stimuli. This odd property of the SSI serves as a reminder that care is sometimes required when interpreting information theoretic measures. It is important to note that a locally negative SSI does not violate the non-negativity of Shannon information, as the mutual information—the expected value of the SSI—remains non-negative.

5.2.3.4 Summary

The stimulus distribution itself can affect the shape of the mSSI and the best-encoded stimulus in the pre-asymptotic regime. When there is a gradient in the stimulus distribution $p(s)$ around the characteristic stimulus of a neuron (peak of unimodal, or flank of monotonic tuning curve), the best-encoded stimulus is shifted in the direction of greater $p(s)$ in the case of unimodal tuning curves and in the opposite direction in the case of monotonic tuning curves. In small, noisy populations with monotonic tuning curves the stimulus distribution can strongly effect the best-encoded stimulus; such cases should be analysed individually.

5.2.4 Best-encoded stimuli in forced choice tasks

The preceding sections described how the best-encoded stimulus of a neuron depends upon trial-to-trial variability and population size when there are many possible stimuli. Large changes in the best-encoded stimulus, such as the peak-flank transition for unimodal tuning curves, are caused by changes in the relative amounts of information associated with fine versus coarse discrimination (Butts and Goldman, 2006). These effects can only be observed when the stimulus ensemble is sufficiently rich, in the sense that it must include both narrowly and widely separated stimuli so that both fine and coarse discrimination are relevant. When applied to rich stimulus ensembles, information theoretic measures, such as the SSI, quantify information in a way that is not specific to any particular behavioural task. Sometimes, however, it is useful to predict what the best-encoded stimulus is for a specific behavioural task, particularly when comparing precision at the neural and behavioural levels, as behavioural experiments typically involve simplified tasks (e.g. two-alternative forced choice; 2AFC). This type of task involves making a decision between predefined choices, and the SSI can be used to analyse these tasks by tailoring the stimulus ensemble to match the choices

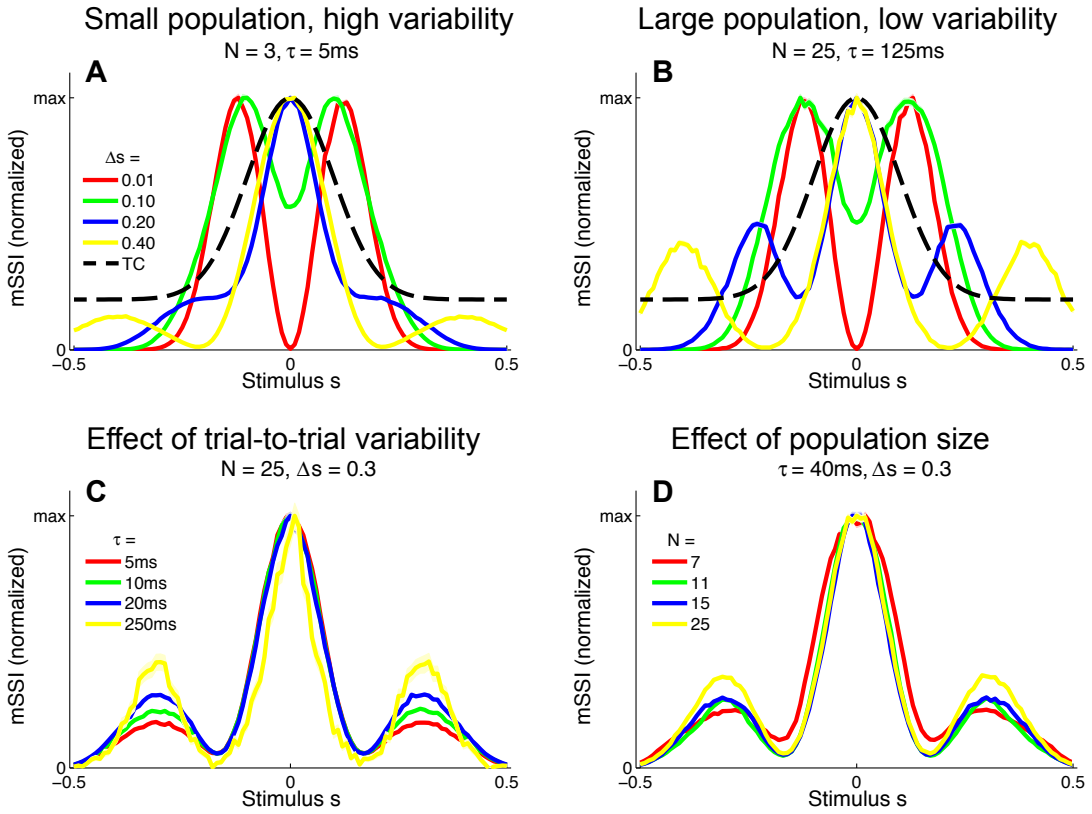

Figure 5.9: Best-encoded stimuli for unimodal tuning curves in a two-alternative discrimination task. (**A**) The mSSI of a neuron in a very small population with high trial-to-trial variability, for two-alternative discrimination tasks with varying normalised stimulus interval Δs. (**B**) As (**A**), but for a slightly larger population and lower variability. (**C**)–(**D**) Typical effects of changing the variability and population size respectively. Uncertainty in the mSSI is indicated by shaded regions of ± 1 StdErr where it is greater than the width of the line. Parameters: $f_{mod} = 40\,\text{spikes/s}$, $f_0 = 10\,\text{spikes/s}$, $\omega = 0.1$.

in the task: for instance, a 2AFC task corresponds to an ensemble with two stimuli. Modelling the task in this way allows us to identify the stimulus values for which a neuron conveys the most decision-relevant information, where the decision to be made is defined by the task.

5.2.4.1 Unimodal tuning curves

In single neurons with unimodal tuning curves, when the stimulus ensemble is simple (contains few stimuli) it is the ensemble itself that determines the best-encoded stimulus (Butts and Goldman, 2006). For example, in a 2AFC task involving two closely-spaced stimuli (i.e. when Δs is much less than the width of the tuning curve flank), coarse discrimination is irrelevant as there are no widely-spaced stimuli in the ensemble, so information about the stimulus is maximised when the stimuli fall on the steepest part of the tuning curve flank, where Fisher

information is maximal. Thus, for closely-spaced stimuli the best-encoded stimuli are on the flanks of the tuning curve, meaning that a neuron conveys the most information in support of a decision between the two stimuli when they fall on a flank of the tuning curve. Conversely, for widely spaced pairs of stimuli (such that when one lies at the tuning curve peak the other lies on the non-tuned baseline region) fine discrimination is irrelevant and the best-encoded stimulus is at the peak of the tuning curve. Because the coding regime is determined entirely by the stimulus ensemble, the best-encoded stimulus is not affected by the level of trial-to-trial variability as they are when the stimulus ensemble is rich.

To investigate the effect of the task (stimulus ensemble) on the best-encoded stimuli in the context of a population code, we computed the mSSI for unimodal tuning curves in 2AFC tasks with stimulus spacings in the range $\Delta s = [0.01, 0.6]$. The results show that the best-encoded stimuli indicated by the mSSI in 2AFC tasks is determined by the stimulus spacing in the same way that the singleton SSI is, i.e. it is at the tuning curve peak for coarse discrimination and at the flanks for fine discrimination (see Figure 5.9A and B; the transition between coarse and fine discrimination occurs at approximately $\Delta s = 0.15$). With unimodal tuning curves, the trial-to-trial variability and the size of the population have little effect upon the best-encoded stimulus. In the coarse discrimination regime, the only effect of changing the level of variability (Figure 5.9C) or the number of neurons in the population (Figure 5.9D) is to change the height of the secondary mSSI peaks relative to the central maximum. In the fine discrimination regime, neither the level of variability or the population size have any effect on best-encoded stimulus—it is always the same as that predicted by the Fisher information, but for intermediate, transitional stimulus spacings where the mSSI is bimodal, changes in τ can cause small shifts in the best-encoded stimuli (results not shown).

5.2.4.2 Monotonic tuning curves

We extended the analysis to populations of monotonic tuning curves by again computing the mSSI for 2AFC tasks with stimulus spacings in the range $\Delta s = [0.01, 0.6]$. For monotonic tuning curves, the best-encoded stimulus for fine discrimination tasks is again at the point of maximum Fisher information i.e. the flank of the tuning curve (Figure 5.10). However, for coarse discrimination tasks, predicting the best-encoded stimulus is less straightforward than is the case for peaked tuning curves, as the shape of the mSSI is somewhat dependent on trial-to-trial variability and population size. As the stimulus interval is increased, the best encoded stimulus shifts towards the upper end of the tuning curve flank (Figure 5.10A). For coarse discrimination a secondary mSSI peak emerges on the low-responding side of the flank and grows as the variability is reduced or the population size increased (Figure 5.10B). The two peaks are due to the two

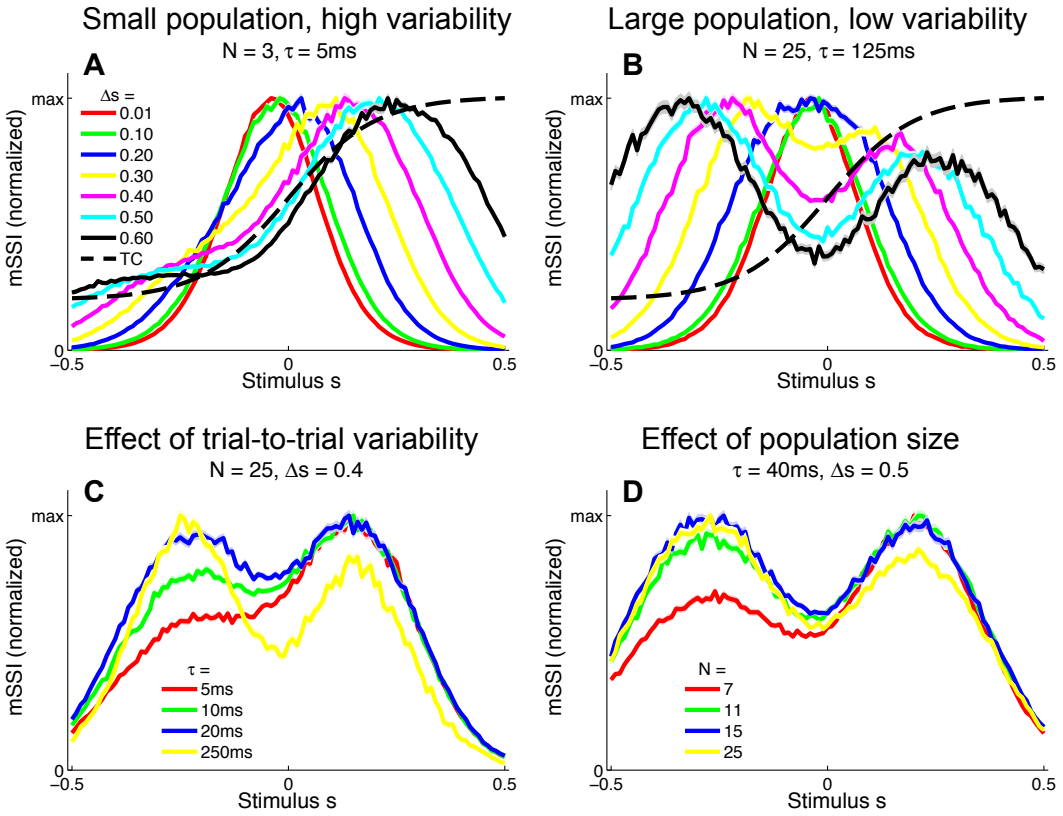

Figure 5.10: Best-encoded stimuli for monotonic tuning curves in a two-alternative discrimination task. Panel (**A**): mSSI of a neuron in a very small population with high trial-to-trial variability, for two-alternative discrimination tasks with varying normalised stimulus interval Δs. Panel (**B**): as (**A**), but for a slightly larger population and lower variability. Panels (**C**) and (**D**) show typical effects of changing the variability and population size respectively. Uncertainty in the mSSI is indicated by shaded regions of ± 1 StdErr where it is greater than the width of the line. Parameters: $f_{mod} = 40\,\text{spikes/s}$, $f_0 = 10\,\text{spikes/s}$, $\omega = 0.1$.

components of the SSI in a two-alternative task (see section 5.1.6), each of which is bell-shaped. The distance between the peaks of the two components is determined by the stimulus interval, until the stimulus interval is greater than the width of the tuning curve flank. The effect of decreasing the variability (Figure 5.10C) or increasing the number of neurons in the population (Figure 5.10D) is to increase the magnitude of the mSSI peak on the low-responding side of the tuning curve; eventually this peak exceeds the one at the plateau end of the flank, and the best-encoded stimulus shifts to the baseline end of the flank. In general, the best-encoded stimulus is at the baseline end of the tuning curve flank for large populations, low variability or widely-spaced stimuli.

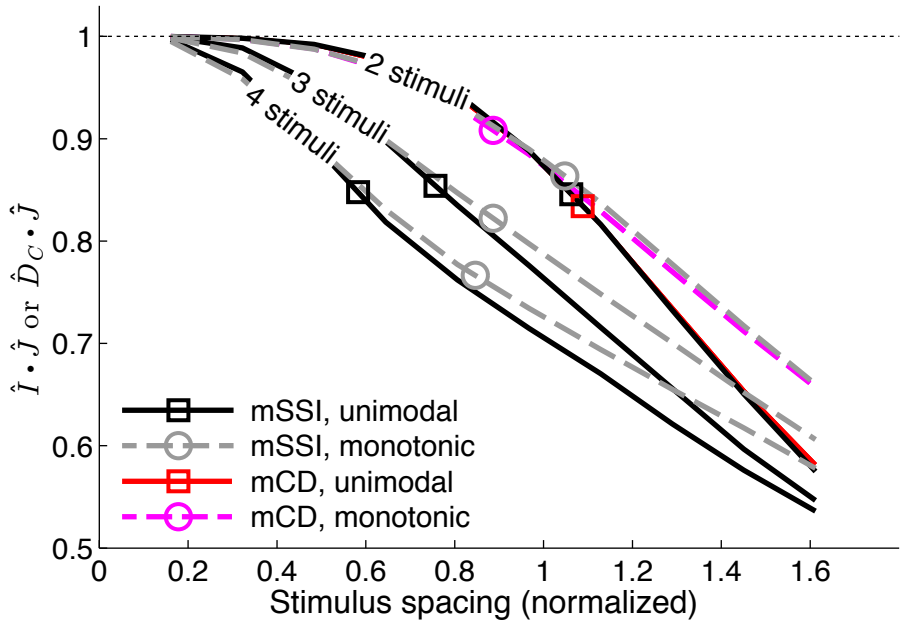

Figure 5.11: Best-encoded stimulus depends on stimulus spacing in a forced-choice task. This figure summarises how the best-encoded stimulus predicted by the marginal SSI and marginal Chernoff distance diverges from that predicted by the Fisher information as the distance between stimuli is increased in a forced-choice discrimination task. The normalised dot product (see section 5.1.5) of the mSSI and the Fisher information (for 2, 3 and 4-alternative tasks), together with the normalised dot product of the marginal Chernoff distance and Fisher information (for 2-alternative tasks), is plotted against the normalised stimulus spacing (δs; see section 5.1.6) for both unimodal and monotonic tuning curves. Markers indicate the approximate transition point between the fine and coarse discrimination regimes; these correspond to the δs values indicated in Figure 5.12. Parameters: $N = 81$, $\tau = 20\,\text{ms}$, $f_{mod} = 40\,\text{spikes/s}$, $f_0 = 10\,\text{spikes/s}$, $\omega = 0.1$ (unimodal), $\omega = 0.044$ (monotonic).

5.2.4.3 What constitutes coarse discrimination?

In contrast to the SSI, the Fisher information is specific to tasks where information on fine-grained stimulus differences dominates—fine discrimination, estimation, reconstruction—and does not quantify information that contributes to coarse discrimination. The Fisher information always predicts that the best-encoded stimuli coincide with the flanks of peaked tuning curves, and when the SSI is used to analyse equivalent fine-grained tasks it yields similar predictions. So far, we have described the stimulus spacing in 2AFC tasks rather loosely as being coarse or fine, but what actually constitutes a 'fine' or 'coarse' discrimination task? Where is the boundary between fine and course discrimination, and what are the limits of applicability of Fisher information in terms of stimulus spacing; when does Fisher information correctly identify the best-encoded stimulus? What if there are more than two alternatives in a forced choice task?

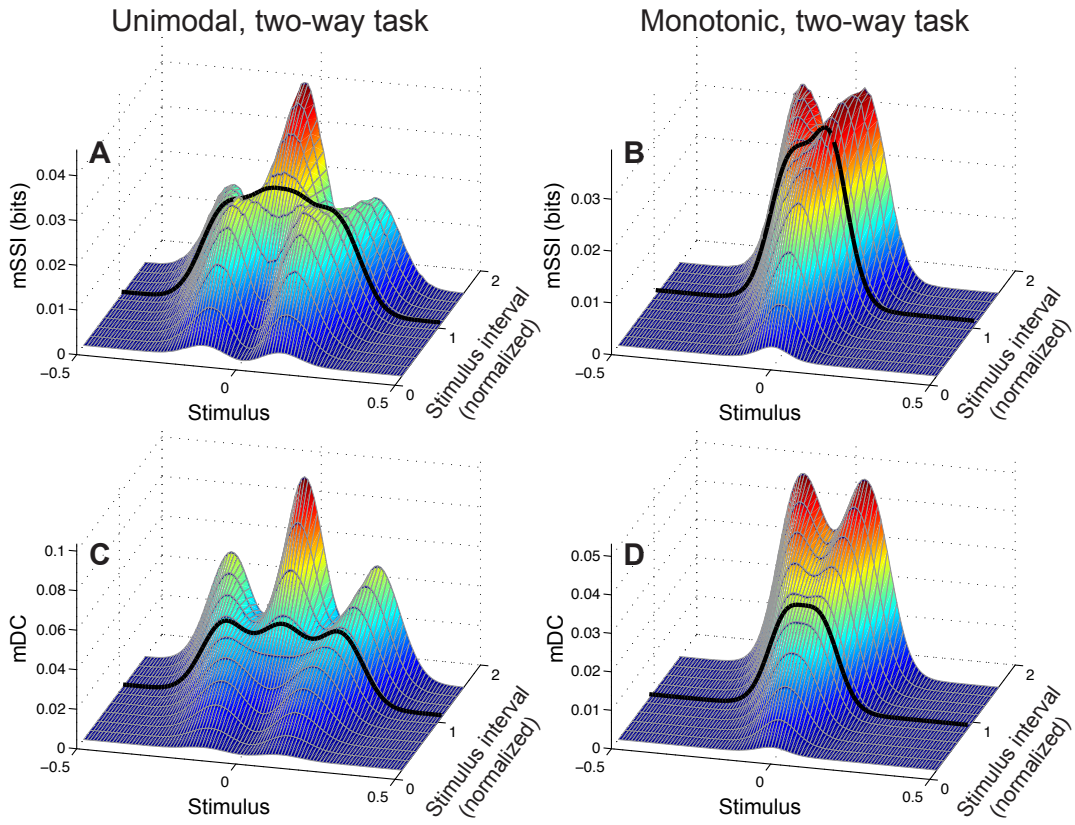

Figure 5.12: Examples of SSI and Chernoff distance for forced-choice tasks. Each panel shows the marginal SSI (**A** and **B**) or marginal Chernoff distance (**C** and **D**) for 2-alternative forced choice tasks with normalised stimulus intervals ranging from $\delta s = 0.16$ to $\delta s = 1.6$. Data for both unimodal (**A** and **C**) and monotonic (**B** and **D**) tuning curves are shown. The model parameters are the same as in Figure 5.11 and the heavy black lines on each panel correspond to the transition points indicated by markers in Figure 5.11. Parameters: $N = 81$, $\tau = 20\,\text{ms}$, $f_{mod} = 40\,\text{spikes/s}$, $f_0 = 10\,\text{spikes/s}$, $\omega = 0.1$ (unimodal), $\omega = 0.044$ (monotonic).

To address these questions we computed the mSSI for two, three and four-alternative forced choice tasks with a range of stimulus spacings, for populations of 81 neurons with either unimodal or monotonic tuning curves (Figure 5.11). We modelled normalised stimulus spacings in the range $\delta s = [0.16, 1.6]$, which covers the transition between the fine and coarse discrimination regimes. For both types of tuning curve, the transition between fine discrimination and coarse discrimination, according to the mSSI, occurs at approximately $\delta s = 1$ (Figure 5.11A), i.e. where the distance between the stimuli is roughly the same as the width of the tuning curve flank (Figure 5.11B). Increasing the number of alternatives in the task introduces elements of coarser discrimination; for example, a three-alternative task involves distinguishing between non-adjacent stimuli separated by $2\Delta s$, as well as adjacent stimuli with an interval of Δs. The presence of more widely spaced stimulus pairs within the ensemble can drive the shape of the mSSI towards the coarse discrimination regime i.e. a shape less like that of the Fisher information, as

indicated by lower $\hat{I}_m \bullet \hat{J}$ values (Figure 5.11A; see section 5.1.5 for more details of how similarity of shape was quantified). Figure 5.11C shows how the mSSI varies with δs for a two-alternative task and unimodal tuning curves; the transition between the Fisher-like bimodal shape of the fine discrimination regime and the strong central peak of the coarse discrimination regime can be clearly seen. Similarly, for monotonic tuning curves, the transition is between unimodal mSSI for fine discrimination (best-encoded stimulus at the flank of the tuning curve) and bimodal mSSI for coarse discrimination (best-encoded stimulus near one end of the flank), as can be seen in Figure 5.11D.

5.2.4.4 Chernoff distance: an efficient measure for 2AFC tasks

For two-alternative tasks, the Chernoff distance can be used to measure the discriminability of the two stimuli. To compare the best-encoded stimuli predicted by the SSI and Chernoff distance, we computed the marginal Chernoff distance (mDC) for the same population models and tasks described in the preceding paragraph (Figure 5.11E–F). For unimodal tuning curves, the difference in shape between the mSSI and mDC is primarily in the outer information peaks located at approximately $\pm\Delta s$ (Figure 5.11C and E), and this has little effect on the best-encoded stimulus. The asymmetry of the monotonic tuning curve, however, reveals an important difference between the mSSI and mDC: because the Chernoff distance is commutative ($D(s, s+\Delta s) = D(s+\Delta s, s)$) both components of the mDC are identical, but shifted by Δs. This means that the mDC for the monotonic tuning curve is symmetrical, with two information peaks of equal height and thus two best-encoded stimuli. Although the shapes of the mSSI and mDC are not exactly the same, the best-encoded stimuli predicted by the mDC are qualitatively consistent with those predicted by the mSSI and, as for the mSSI, the transition between fine and coarse discrimination regimes occurs at approximately $\delta s = 1$. The shape of the mDC is also very close to the shape of the Fisher information for fine discrimination: $\hat{D}_m \bullet \hat{J}$ approaches 1 as $\delta s \to 0$ (Figure 5.11A). The best-encoded stimulus predictions of the Chernoff distance agree with those of the SSI, and the mDC is therefore a computationally efficient method of estimating best-encoded stimuli for two-alternative tasks. Additionally, the Chernoff distance is closely related to the probability of making an incorrect choice in a 2AFC task (Cover and Thomas, 2006; Kang and Sompolinsky, 2001), allowing comparisons between information at the neural level and performance in behavioural tasks.

5.2.4.5 Summary

When the task in hand limits the possible stimulus choices, such as in a 2AFC protocol, the spacing between the alternatives determines the best-encoded

stimulus. If the stimulus alternatives are separated by less than the width of the tuning curve flank, a neuron is most informative when the stimuli fall on the steepest parts of the tuning curve flanks, as predicted by the Fisher information. For stimulus spacings greater than the width of the tuning curve flank, the best-encoded stimuli are at the tuning curve peak for unimodal tuning curve, and at either end of the flank for monotonic tuning curves. The Chernoff distance can be used to assess best-encoded stimulus for 2AFC tasks, and is an efficient alternative to the SSI while yielding qualitatively equivalent predictions.

5.3 Conclusion

We have shown how the best-encoded stimulus for neurons with sigmoidal monotonic tuning curves depends on the number of neurons in the population, the level of trial-to-trial variability and on the location of the tuning curve flank within the stimulus space. This builds upon the work of Butts and Goldman (2006) and Chapter 4 that showed how best-encoded stimuli for unimodal tuning curves depend on the level of trial-to-trial variability and the population size. For large populations (of the order of hundreds of neurons), we confirmed by numerical simulation that the best-encoded stimulus predicted by the SSI for monotonically tuned neurons agrees with that predicted by the Fisher information: i.e. the best-encoded stimulus is on the sloping flank of the tuning curve. This is in agreement with the earlier studies addressing unimodal tuning curves and also with the proven equivalence of Fisher information and mutual information in the limit as the population size $N \to \infty$ (Brunel and Nadal, 1998). Away from the asymptotic regime, in smaller populations where the shapes of the marginal SSI and the Fisher information differ, the best-encoded stimulus is harder to predict as it can be strongly affected by the population size, variability and stimulus distribution as well as the tuning curve; this is in contrast to peaked tuning curves, where the SSI is greatest at the peak of the tuning curve under similar conditions. This difference is due to the fact that strong responses by neurons with saturating monotonic tuning curves are often triggered by a wider range of stimuli, which makes them less informative. Far from the asymptotic regime, the best-encoded stimuli can be at either end of the tuning curve flank, and may extend (i.e. the SSI or mSSI may be flat) across either the plateau or baseline regions of the tuning curve. Where the variability is very high, for instance in the first few milliseconds post-stimulus in the case of Poisson-like noise, the best-encoded stimulus is determined by the characteristic stimulus (cut-off point) of the tuning curve, which paradoxically can mean that an absence of activity conveys the most information in terms of estimation; this is not the case for peaked tuning curves.

We next examined how the behavioural task can affect the best-encoded stimulus. One of the strengths of the SSI is its flexibility: it can be used to analyse an arbitrary

task by manipulating the stimulus distribution. Butts and Goldman (Butts and Goldman, 2006) showed that when the SSI is used to quantify information in the context of a specific task, the task itself can determine what the best-encoded stimulus is, with population size and variability having little effect. The results support this and show that it also holds for monotonic tuning curves. In general, if tasks are thought of as lying on a continuum between the two extremes of stimulus distribution richness (i.e. two discrete stimuli and a continuous stimulus distribution), the simpler the task, the more influence it will have on the best-encoded stimulus. Where the best encoded stimulus is determined mainly by the task, population size and variability have little effect and no distinct asymptotic and non-asymptotic regimes exist (such as the peak and flank coding regimes for unimodal tuning curves). For fine discrimination tasks, the best-encoded stimuli are as predicted by the Fisher information: at the steepest parts of the tuning curve. As a rule of thumb, fine discrimination tasks are those where the stimulus spacing is less than the width of the tuning curve flank or flanks. For greater stimulus spacings the best-encoded stimuli shift from the flanks to the peak, or the ends of the flank for monotonic tuning curves. For tasks involving more than two distinct stimuli, the distance between the closest pair of stimuli in an ensemble is important, as it determines whether the fine discrimination regime is relevant. In addition to this, the closest stimuli are likely to have the most similar response distributions, which in turn determines the rate at which the mutual information increases with population size (Kang and Sompolinsky, 2001).

Although the SSI can be used to analyse many different tasks, using it to determine best-encoded stimuli in a two-alternative task is not computationally efficient. This type of analysis boils down to measuring the difference between two response distributions, and if these distributions are known—as in a modelling study—then a measure such as the Chernoff distance is more efficient. The results confirm that the SSI and Chernoff distance are in close agreement as to the best-encoded stimuli in two-alternative tasks where each stimuli is equally likely. This is related to the work of Kang et al. (2004), who used Chernoff distance to analyse the relative amounts of information conveyed by neurons in discrimination tasks of varying coarseness. Whereas they constructed curves of Chernoff distance versus stimulus interval for whole populations, we used it to identify which stimuli were best encoded by a given neuron. Just as the SSI and the Chernoff distance are consistent with one another where both are valid (two-alternative tasks), the SSI and the Fisher information are consistent where the Fisher information is valid, i.e. in the asymptotic regime and for fine discrimination tasks. The Fisher information gives insight into fine discrimination, the Chernoff distance into two-alternative discrimination of arbitrary coarseness, while the SSI is flexible and can give a generalised picture of the informational tuning curve. These measures do not contradict each other, they simply have limited validity and may not yield meaningful predictions outside their respective domains of validity.

Like any study, this one has a few limitations. Firstly, the model only considers information transmitted in firing rates, and assumes uncorrelated Poisson trial-to-trial variability. These are clearly simplifications that discount any information conveyed by spike times, but analysis of rate coding remains important as tuning curves are still widely used by experimentalists to describe the information-bearing activity of neurons. Also, correlations in trial-to-trial variability have been shown to have a relatively small effect on best-encoded stimuli (see Chapter 4).

To conclude, these results should serve as a reminder that it is not safe to assume that strong neuronal responses are informative; it is perhaps more often the case that moderate responses are most informative, as these occur in response to stimuli that lie on the flanks of the tuning curve. The best approach, however, may be to analyse observed population codes on a case by case basis, as information tuning curves and best-encoded stimuli can be easily estimated from experimentally measured tuning curves using the measures discussed in this chapter.

Chapter 6

Discussion

This thesis has addressed the quantification of several aspects of rate population codes: information, best-encoded stimuli and topography. More specifically, we have investigated the relationship between Fisher information and information theoretic measures in finite-size populations, both in terms of overall precision and quantifying the information associated with specific stimuli. We have also examined how tuning curves, variability, noise correlations and task affect the information that neurons transmit about specific stimuli, in particular which stimuli are best encoded, and how best to estimate this from a model of tuning curves and variability. Finally, we proposed an objective and quantitative method for analysing topography in neural maps. In this chapter we recapitulate the results relevant to each of these three aspects of population coding, and discuss the wider significance of the findings.

6.1 Fisher and Shannon information

Information theory provides a powerful and general set of tools for assessing the precision of neural codes, but information theoretic measures are difficult to compute for experimentally-characterised populations due to the large number of observations required. Fisher information is an alternative statistical measure of precision that is generally easier to compute in modelling studies, but specifies an upper bound on coding precision that is only achieved in infinite populations. Brunel and Nadal (1998) showed that I_{Fisher}, an information theoretic measure derived from Fisher information, could provide an estimate of the MI in infinite populations (given certain conditions).

Chapter 4 examined the relationship between Fisher information and information theoretic measures in finite size populations with unimodal tuning curves, taking into account the effects of trial-to-trial variability level and population size.

Numerical simulations showed that Fisher information can accurately predict both the mutual information and the stimulus-specific information for population sizes of the order of 100 neurons, even with supra-Poisson variability and integration times on the order of 10 ms. The MI is well approximated by I_{Fisher} (3.5% error) in populations with more than approximately 50 neurons, even with high variability and small time windows (e.g. $F = 3$ and $\tau = 30\,\text{ms}$). For populations with fewer neurons, I_{Fisher} tends to overestimate the MI and this disparity is greater for smaller populations. Increasing the amount of trial-to-trial variability (noise), or reducing the time window over which spikes are counted, increases the difference between the MI and I_{Fisher}, but does not change the rate at which the two measures converge as a function of population size. However, the predictions of the Fisher information can be close to those of information theory even in much smaller populations, particularly if the variability is sub-Poisson. This means that for population codes involving large numbers of neurons, such as those thought to operate in the cortex, Fisher information provides effectively the same insights as the mutual information and SSI for tasks where discrimination between closely spaced stimuli are important; this includes fine discrimination and estimation tasks.

Chapter 4 also investigated the effect of noise correlations with both uniform pairwise correlations and a localised, spatially decaying structure. Correlations in the biologically realistic range (pairwise correlation coefficients up to 0.3 were modelled) slightly increase the difference between the two measures (i.e. cause I_{Fisher} to overestimate the MI) and delay their convergence, but these effects are small in comparison to those of population size or noise level. The effect of the correlation range or spatial decay rate of localised correlations was also investigated. Exponentially decaying correlations have the greatest effect on the MI when the spatial range of correlations matches the tuning curve width i.e. when noise and signal correlations have the same extent. This is in agreement with the findings of Sompolinsky et al. (2001), who reported that Fisher information was at a minimum when the correlation decay was closely matched to the widths of the tuning curves.

As with any study, this analysis has a number of limitations. Perhaps most importantly, the results are specific to tasks involving fine discrimination, as Fisher information is defined as a very local measure of precision around a particular stimulus value. This limitation also applies to the information theoretic measures, since the stimulus ensembles used in Chapter 4 were constructed in such a way as to simulate a reconstruction or fine discrimination task in order for the SSI and the MI to be comparable with Fisher information and I_{Fisher}. For detection tasks, and probably also for coarse discrimination tasks, neurons best encode stimuli at the peak of their tuning curves. Fisher information is not applicable to these tasks, but other measures, such as Chernoff distance (Kang et al., 2004), can be used to estimate the MI as a function of the discrimination 'coarseness', the spacing

between stimuli. We also assume that information is carried by a rate code; in cases where this assumption does not hold, the tuning curves and rate variability do not necessarily determine the best-encoded stimuli, as additional information about other stimuli may be conveyed by spike timing. In addition, all models were based on broadly-tuned neurons; we did not investigate how tuning curve width contributes towards determining the coding regime. All of the populations modelled in chapters 4 and 5 consisted of arrays of identical, uniformly spaced tuning curves. Heterogeneity of tuning curve widths (Wilke and Eurich, 2002) or amplitudes (Ecker et al., 2011) is known to have the potential to increase the precision of the code, so it is possible that diversity of tuning could also affect the relationship between e.g. I_{Fisher} and the MI.

Some other studies addressing the validity of Fisher information have asked: what are the properties of population codes that are optimal in terms of Fisher information? These have suggested that tuning curves optimised to give maximal Fisher information would not resemble those observed experimentally (Bethge et al., 2002). If the tuning curves are constrained to be bell-shaped, maximising the Fisher information of a population means that tuning curve width is dependent upon population size, and is narrow in large populations. This is due to the fact that Fisher information increases as the tuning function width is decreased, up to the point where the overlap of neighbouring tuning curves is insufficient to give full coverage of the stimulus space (Berens et al., 2011). These studies also found that Fisher-optimal population codes are often sub-optimal in terms of other measures. This thesis does not contradict those findings, but addresses a separate question: when can Fisher information be used to assess the precision of population codes that have been characterised experimentally? The model neurons considered here are broadly-tuned, in line with experimental findings (see e.g. Clifford, 2002), and the width does not vary with population size. Whilst Fisher information appears to be a poor tool for assessing the optimality of population codes, the results described in this chapter suggest that it is a valid measure of discrimination precision, albeit with limitations.

There has been much interest in calculating Fisher information from experimental data, and there are several possible approaches to estimating it, depending on the data available. The simplest method of obtaining Fisher information is to compute it directly from the tabular conditional response distribution $p(\boldsymbol{r}|\theta)$ by numerically evaluating Equation 2.6 (as in e.g. Dean et al., 2005). Measuring $p(\boldsymbol{r}|\theta)$ directly is only feasible for single neurons or very small populations, so the population Fisher information can only be obtained by this method in the case of uncorrelated noise. Alternatively, it is possible to use experimental data to construct a model of tuning curves and variability, and then to compute Fisher information from the model (as in e.g. Durant et al., 2007). Independent noise is typically modelled as a Poisson or univariate Gaussian distribution and correlated

noise by a multivariate Gaussian distribution. While the best-encoded stimuli in large populations can be identified by computing the singleton Fisher information, computing the population Fisher information under a Gaussian variability model requires knowledge of the stimulus-dependent covariance matrix $Q(\theta)$. The measurement of $Q(\theta)$ represents the most challenging obstacle to computing the population Fisher information, as this requires many trials and simultaneous recording of multiple neurons. In addition, any inaccuracies will be amplified when $Q(\theta)$ is inverted to obtain $Q^{-1}(\theta)$ (see Equation 2.7). It is not yet clear what the best method of determining the covariance matrix is, or how many trials are required to measure $Q(\theta)$ with sufficient accuracy to obtain a reasonable estimate of the Fisher information; more work is required to establish the answers to these open questions. Additionally, the level of noise correlations present in the brain is a matter of active debate (Ecker et al., 2010); in cases where trial-to-trial variability is effectively uncorrelated, the process of calculating the population Fisher information is greatly simplified.

The problem of determining the covariance matrix can be avoided by using a decoding approach. This involves constructing a function that estimates the stimulus given single-trial response spike counts for each neuron in the population. The variance of this estimator $\hat{\theta}(\boldsymbol{r})$ over many trials can then be used to determine a lower bound on the Fisher information:

$$J(\theta) \geq \frac{1}{\mathrm{Var}(\hat{\theta}(\boldsymbol{r}))} \tag{6.1}$$

This approach has been used in theoretical studies (e.g. Seriès et al., 2004; Beck et al., 2008; Chelaru and Dragoi, 2008). With this method, the most difficult part of the analysis is constructing an efficient estimator; this can be done via a number of machine learning techniques and the quantity of data required to train the estimator will depend upon the method used.

Our findings have two main implications for the experimental characterisation of neurons. Firstly, Fisher information can be used to obtain approximations of both the MI and the mSSI for neurons within large populations. As such, it is a reliable indicator of both coding precision and best-encoded stimuli for discrimination or reconstruction. In cases where it appears that the population size is of the order of 100 neurons, Fisher information can be safely used, given the limitations discussed above. Secondly, for smaller populations where the number of neurons is known or can be accurately estimated, it is feasible to compute the SSI and marginal SSI if the tuning curves, trial-to-trial variability, and correlations if necessary, can be modelled.

6.2 Interpreting tuning curves

In chapters 4 and 5, we examined how tuning curve type, variability, population size and behavioural task affect the information that rate coding neurons transmit about specific stimuli. To quantify this we used the SSI, which has a number of advantageous properties. In particular, it is not limited to analysing any one type of task in the way that the Fisher information is specific to fine discrimination and the Chernoff distance is specific to two-alternative tasks. By manipulating the stimulus distribution, the SSI can be used to analyse any specific task or to give an indication of best-encoded stimulus that takes all possible tasks into account. The principle drawback of the SSI is its computational complexity, which until recently meant that it had only been used to analyse very small populations (Butts and Goldman, 2006). Even when the computational cost of obtaining the SSI is reduced by the use of Monte Carlo integration, it is still more difficult to compute than the Fisher information or Chernoff distance. As discussed in the preceding section, Fisher information is a useful alternative to the SSI for large populations and tasks where most of the information conveyed by the code relates to fine discrimination. Under these conditions it provides the same insights as the SSI. The Chernoff distance is similarly useful for quantifying the discriminability of two specific stimuli, as in a two-alternative forced choice task.

Computing the marginal SSI, or an equivalent measure, allows us to answer the question: which stimuli are best encoded by a neuron operating within a population? Which is most informative: the highest spike counts, or responses that are most sensitive to small changes in the stimulus, where the tuning curve gradient is greatest (as predicted by the Fisher information)? Butts and Goldman (2006) found that in very small populations with unimodal tuning curves the best-encoded stimulus indicated by the marginal SSI was dependent on the level of variability, and conflicted with the predictions of Fisher information when the variability was high. In the high variability, pre-asymptotic regime the best-encoded stimuli coincided with the peaks of the tuning curves. Single neurons operating in this peak coding regime have also been found when using the SSI to analyse empirical data (Montgomery and Wehr, 2010). Chapter 4 examined the effect of population size on the best-encoded stimulus, extending the analysis of Butts and Goldman to populations of up to 256 neurons. In Chapter 5, the analysis was further extended to monotonic tuning curves. We found that the shape of the mSSI converges towards that of the Fisher information as the population size increases and, consequently, both measures predict the same best-encoded stimulus in large populations. Discrepancies occur only within the pre-asymptotic regime: a restricted domain of small populations (approximately $N < 50$) combined with high Fano factors or short integration times. Under these conditions, the mSSI indicates peak coding in unimodal tuning curves, whereas Fisher information, as

in all cases, indicates flank coding. It seems likely that this regime coincides with that in which the error of efficient unbiased estimators does not saturate the Cramér-Rao bound and the conditional distribution of these estimators is non-Gaussian, but further work would be required to establish if this is the case. For monotonic tuning curves, the pre-asymptotic best-encoded stimulus depends on additional factors, principally the relative extents of the baseline and plateau regions of the tuning curve. Outside this limited domain, both measures indicate that the best-encoded stimuli fall at the flank(s) of the tuning curve, regardless of whether it is unimodal or monotonic. Decreasing variability, increasing integration time, and uniform noise correlations drive the system towards the asymptotic regime, while localised correlations have the opposite effect. This dependence upon integration time means that what stimulus is best encoded by individual neurons is a dynamical process: neurons will operate in the peak coding regime immediately following stimulus presentation and transition to flank coding as time progresses. This time dependence of best-encoded stimulus raises some interesting questions about the shape of optimal tuning curves. For instance, should tuning curves change shape dynamically following stimulus presentation in order to maximise information? While this thesis has focussed on tools for analysing known tuning curves, investigating the consequences of these results for optimal tuning curves may be an interesting topic for future work.

Estimating the best-encoded stimulus of a neuron is not an empty theoretical exercise; there is a growing body of evidence that the neurons with the most informative activity contribute strongly to decision-making. For any given task, determining the best-encoded stimulus of a given neuron, and identifying the neurons in a population that are most informative, are closely related problems (in fact they are exactly equivalent when all tuning curves are identical, shifted copies and the stimulus distribution is uniform). Testing the hypothesis that the most informative neurons contribute most to decision-making is relatively tractable for two-alternative tasks where the best-encoded stimulus is controlled by the task itself and behavioural performance in the task is easily measured. In a theoretical study of maximum-likelihood decoding of population codes, Jazayeri and Movshon (2006) found that the spacing between stimuli in a discrimination task was an important determinant of which neurons contribute to decision-making, with neurons tuned to the task stimuli contributing most to coarse discrimination and flanking neurons (i.e. those whose tuning curve flanks span the stimuli to be discriminated) contributing most to fine discrimination. The question of whether subsequent neural processing makes optimal use of the information propagated by a population code is an important one, as the information content of the population activity itself is less relevant if it is not fully utilised. Some experimental studies have reported evidence of population codes that are utilised optimally, in that the most informative neurons appear to have the greatest causal effect on behaviour in both fine and coarse discrimination tasks. Evidence

that flanking, 'off-channel' neurons are most important in fine discrimination tasks has been found in psychophysical studies (e.g. Hol and Treue, 2001) and studies involving direct measurement of single neuron activity (Purushothaman and Bradley, 2005) and BOLD response in fMRI voxels (Scolari and Serences, 2010). Recent theoretical progress in the interpretation of choice probabilities (Haefner et al., 2013) opens the door to more robust estimation of neuronal contributions to decision making. Given these advances, we might expect to see new experimental evidence indicating that highly informative, flanking neurons contribute strongly to decision-making in fine discrimination tasks, and unimodally-tuned neurons selective for the stimuli in coarse discrimination tasks—if firing rate information is indeed optimally utilised. Similarly, the results of the simulations suggest that neurons whose activity is just above baseline, or just reaching its saturated plateau level, at the stimulus of interest will have high choice probabilities in coarse discrimination tasks involving monotonically-tuned populations.

Comparing precision at the neuronal and behavioural levels is much more difficult when the stimulus ensemble is richer than that of a simple two-alternative discrimination task, for instance in the case of estimation. This makes it difficult to test whether the pre-asymptotic regime, where the best encoded stimuli do not coincide with the flanks of the tuning curves, is biologically relevant. The results confirm that the pre-asymptotic regime is restricted to very high levels of trial-to-trial variability or very short integration times when the population size is of the order of hundreds of neurons, as it likely to be the case in the cortex. Do neural systems ever operate in the pre-asymptotic regime? If they do, is subsequent information processing adapted to make use of the fact that different neurons may be most informative at short versus long integration times? These remain open questions. It may be important to consider stimulus detection when addressing these questions, as the time required to accumulate evidence for detection may impose a lower limit on the range of integration times that are relevant for estimation.

Although we have focussed upon the SSI, the specific surprise also gives similar predictions as to the best-encoded stimuli (Chapter 4). Other stimulus-specific decompositions of the MI are possible, in particular the local information or stimulus information density proposed by Bezzi et al. (2002). Under the uniform stimulus distributions modelled in Chapter 4 and most of Chapter 5, the latter measure approaches the specific surprise (up to a multiplicative constant), so its predictions in the cases examined here are likely to be very close to those of the specific surprise and SSI.

6.3 Defining maps and assessing topography

Chapter 3 compared seven measures of topography on the basis of their statistical power when used to detect significant topographic organisation in simulated neural maps. The measures were then used to quantify the topography in maps of source azimuth and IID measured in the auditory cortex of six pallid bats. This analysis showed that there was significant topography in the layout of these spatial tuning properties in most of the bats individually, and in the population as a whole when data from all six bats was combined.

One of the criteria used to select the measures was that they should be flexible in terms of the dimensionality of feature and map spaces. It is therefore a limitation of this study that only 1-D feature spaces and 2-D map spaces were investigated. The dimensionality of feature spaces in particular can vary greatly depending on how the stimulus space is decomposed into features, and how many of these features are taken into account in an analysis. Another limitation was the way that we modelled degradation of the map by adding random noise to the feature space positions. In reality, natural variability of maps is likely to be much more complex and could involve processes very different from independent random noise, for example warping or fracturing of the map. Additionally, it was a limitation of the analysis in Chapter 3 only SNRs ≤ 10 were analysed. In most cases an SNR of 10 was qualitatively similar to zero noise, but this was not always the case (see e.g. the topographic product measure in Figure 3.4D); computing the power of the measures for the zero-noise case would have been a useful addition to the study.

Although Chapter 3 focussed on the use of map measures in statistical tests for map detection, the same measures have other potential applications. They could be used, for example, to assess differences in map orderliness between anatomical regions, developmental stages or experimental groups, or changes in maps as a result of ageing or changes in properties of the environment.

The results of this analysis of pallid bat mapping data confirm that topographic maps of source azimuth and IID exist within binaural clusters in pallid bat primary auditory cortex. As auditory spatial tuning properties within binaural clusters have not been mapped in any other species with sufficient resolution to identify equivalent maps, these findings suggest that high-resolution mapping of spatial tuning properties in the auditory cortex is likely to be important in understanding how auditory space is represented in the cortex.

It is important to note that the systematic arrangements of IID and azimuth selectivity in the pallid bat are confined to clusters of neurons with similar patterns of binaural selectivity. Binaural clusters are a ubiquitous organisational feature of the auditory cortex across species. It remains unclear if topographic maps of source azimuth are present in the intrinsic organisation of binaural clusters in

other species because such studies have not yet been conducted. In the pallid bat these binaural clusters are of the order of 1 mm across, so measuring maps within the clusters requires a technique with spatial resolution of the order of 0.1 mm. In addition, the focus needs to be on mapping within binaural clusters across isofrequency contours as most previous studies have concentrated on mapping along isofrequency contours, potentially spanning multiple binaural clusters. The characterisation of internal organisation of binaural clusters in other species will have significant consequences for our understanding of how auditory spatial information is represented in the cortex and is an important area for future research, particularly given the availability of high-resolution techniques such as multi-photon calcium imaging and multi-electrode arrays.

For the purposes of this thesis, we define a topographic map as any systematic relationship between a given tuning property and the physical location of a neuron. This is perhaps a broader definition than is typically used, because it is not limited to place maps where the characteristic stimuli—the positions of the neurons in feature space—correspond to firing rate maxima. In the pallid bat, the arrangements of azimuth and IID tuning in the EI cluster are examples of maps that are not 'place' maps; in these cases the tuning curves have no well-defined peak and the characteristic stimulus labels on the slopes of the tuning curves are used. The question of what stimulus value to use to characterise or represent a neuron is a nuanced one. Traditionally, those that elicit the maximum response (i.e. the location of the tuning curve peak) have been used, but this does not make sense when the tuning curve is monotonic and has no distinct maximum. One approach would be to use the stimulus value that the neuron conveys the most information about, but this is not easy to determine; maximum information depends on a number of factors and can coincide with steeply sloping regions of the tuning curve, or the peak, or somewhere in between (Butts and Goldman, 2006, and Chapter 4). Using one characteristic stimulus to represent the entire receptive field is clearly a simplification, albeit one that is widely accepted. If this simplification was to prove problematic in the future, it would be possible to adopt new map measures using the same basic form given in Equation 3.1 (i.e. a sum over the products of pairwise distances in map and feature spaces), but using distances in some higher-dimensional and potentially non-parametric space of tuning functions, rather than distances between one-dimensional characteristic stimuli.

The measures used in the analysis were chosen for their flexibility and can be applied to a wide variety of datasets with different dimensionalities of map and feature spaces. Here we have only addressed one-dimensional feature spaces and two-dimensional map spaces, so further work is required to investigate the properties of the measures in spaces with other dimensionality. Feature spaces with more than one dimension will be of particular interest. These methods could also be applied to grid-like mapping data, for example fMRI data. With this type

of data, the regular spatial sampling and greater number of measurements may affect the relative statistical power of the measures, so further investigation of the properties of these map measures using simulated gridded data would be valuable.

In this thesis, we have explored statistical tools that can be used to analyse both topography and the stimulus-specific information contributions of population coding neurons. To conclude, it is useful to discuss what best-encoded stimuli might be able to tell us about topography and, conversely, what topography might be able to tell us about the information contributions of neurons. Currently, neural maps are most often defined in terms of response strength, so the positions of the neurons in feature space correspond to the peaks of tuning curves or receptive fields. An alternative approach would be to label neurons and assign their positions in feature space on the basis of their best-encoded stimuli. For unimodal tuning curves this could be problematic, as there are likely to be two information peaks per neurons, one on each flank, so assigning an unambiguous label could be difficult. Labelling by best-encoded stimulus may be more useful in populations consisting of monotonic tuning curves or tuning curves with diverse shapes, particularly where the neurons are broadly tuned. This is effectively how the source azimuth and IID maps in Chapter 5 were defined, as the 50% activity points are generally close to the Fisher information maxima and the analysis yields similar results if the Fisher information maxima are used (results not shown). It is also possible that the topographic organisation on maps could to explain which stimuli a neuron is informative about. Topography could be quantified on the basis of different labelling schemes, for example maximum Fisher information (mid flank) versus the activity saturation point (e.g. 95% of maximum response) for monotonic tuning curves. If the measured topography for one labelling scheme was significantly stronger, it could imply that the tuning curve regions picked out by that scheme are particularly important for the code. Testing the utility of these approaches remains as a topic for future research.

Appendix A

Mathematical Supplement

A.1 Justification for use of F/τ

The response spike counts are distributed as a multivariate Gaussian with mean $\tau \boldsymbol{f}(\theta)$ and covariance $Q(\theta)$:

$$\boldsymbol{r}(\theta) \sim \mathcal{N}[\tau \boldsymbol{f}(\theta), Q(\theta)]$$

Where the covariance matrix is defined as a function of the tuning curves $f(\theta)$ and the correlation matrix R as:

$$\begin{aligned} Q_{i,j}(\theta) &= F\,[\tau f_i(\theta)]^{\alpha} R_{i,j}\,[\tau f_j(\theta)]^{\alpha} \\ &= F\tau^{2\alpha} f_i(\theta)^{\alpha} R_{i,j}\, f_j(\theta)^{\alpha} \end{aligned} \tag{A.1}$$

Collecting the non-scalar terms as $P(\theta)$:

$$P_{i,j}(\theta) = f_i(\theta)^{\alpha} R_{i,j}\, f_j(\theta)^{\alpha}$$

gives the following expressions for $Q(\theta)$, its inverse, and its derivative with respect to θ:

$$\begin{aligned} Q(\theta) &= F\tau^{2\alpha} P(\theta) \\ Q(\theta)^{-1} &= \frac{1}{F\tau^{2\alpha}} P(\theta)^{-1} \\ Q'(\theta) &= F\tau^{2\alpha} P'(\theta) \end{aligned}$$

The Fisher information is given by (this is the same as Equation 2.7, but the integration time τ is stated explicitly rather than being included in the mean response term):

$$J(\theta) = \tau f'(\theta)^T Q(\theta)^{-1} \tau f'(\theta) + \frac{1}{2} \mathrm{Tr}\left[Q(\theta)^{-1}\, Q'(\theta)\, Q(\theta)^{-1}\, Q'(\theta)\right]$$

Separating out the scalar terms as above, we have:

$$J(\theta) = \tau f'(\theta)^T \frac{1}{F\tau^{2\alpha}} P(\theta)^{-1} \tau f'(\theta) + \frac{1}{2}\mathrm{Tr}\left[\frac{1}{F\tau^{2\alpha}} P(\theta)^{-1} F\tau^{2\alpha} P'(\theta) \frac{1}{F\tau^{2\alpha}} P(\theta)^{-1} F\tau^{2\alpha} P'(\theta)\right]$$
$$= \frac{\tau^{2-2\alpha}}{F} f'(\theta)^T P(\theta)^{-1} f'(\theta) + \frac{1}{2}\mathrm{Tr}\left[P(\theta)^{-1} P'(\theta) P(\theta)^{-1} P'(\theta)\right]$$

Thus for Fano factor variability (i.e. when $\alpha = 0.5$), F and τ appear in the expression for Fisher information only in the ratio τ/F:

$$J(\theta) = \frac{\tau}{F} f'(\theta)^T P(\theta)^{-1} f'(\theta) + \frac{1}{2}\mathrm{Tr}\left[P(\theta)^{-1} P'(\theta) P(\theta)^{-1} P'(\theta)\right] \tag{A.2}$$

Note that Equation A.2 suggests that the Fisher information tends towards a non-zero value as $\tau \to 0$. This is due to a breakdown in the validity of the covariance matrix formulation (Equation A.1) when the mean responses become very small (Panzeri et al., 1999; Wilke and Eurich, 2002).

A.2 Stimulus-specific I_{Fisher}

The stimulus-specific I_{Fisher} (SSI_{Fisher}) is the SSI of an optimal Gaussian estimator $\hat{\theta}_{opt}(\boldsymbol{r})$ with variance equal to the Cramér-Rao bound:

$$I_{ssiF}(\theta) = \sum_{\hat{\theta}_{opt}(\boldsymbol{r})\in\Theta} p(\hat{\theta}_{opt}(\boldsymbol{r})|\theta)\left[\sum_{\theta\in\Theta} p(\theta|\hat{\theta}_{opt}(\boldsymbol{r}))\log p(\theta|\hat{\theta}_{opt}(\boldsymbol{r})) - p(\theta)\log p(\theta)\right]$$

$$\text{where} \quad p(\theta|\hat{\theta}_{opt}(\boldsymbol{r})) = \frac{p(\hat{\theta}_{opt}(\boldsymbol{r})|\theta)p(\theta)}{\sum_{\theta\in\Theta} p(\hat{\theta}_{opt}(\boldsymbol{r})|\theta)p(\theta)}$$

and (due to the Cramér-Rao bound) $\quad p(\hat{\theta}_{opt}(\boldsymbol{r})|\theta) = \mathcal{N}(\theta, J(\theta)^{-1})$

A.3 Computing information theoretic quantities using Monte Carlo methods

A.3.1 Differential entropy and continuous MI

$$H(X) = -\sum_{x\in X} p(x)\log p(x) \tag{A.3}$$

$$h(X) = -\int_X p(x)\log p(x)\,dx \tag{A.4}$$

Shannon entropy (Equation A.3) can only be calculated for discrete variables. Differential entropy (Equation A.4) is a generalisation of Shannon entropy to

continuous-valued random variables, but unfortunately does not retain all of the useful properties of Shannon entropy. In particular, differential entropy is not invariant under a change of variables, such as a change in the units used to measure the stimulus. Also, while Shannon entropy is always positive, differential entropy can take negative values. However, mutual information computed using differential entropies (continuous mutual information) does not suffer from these problems and retains the properties of its discrete counterpart (Cover and Thomas, 2006). Since both the stimulus and response variables in our model are continuous, all the entropies discussed here in relation to our model are differential entropies. How these entropies were calculated in order to find the SSI is described in the following section.

A.3.2 Calculating the MI, SSI and I_{sur}

In all simulations, SSI and specific surprise were calculated simultaneously via Monte Carlo integration. The method for computing the SSI is given here as an example; the MI and specific surprise are evaluated similarly. Referring to Equation 2.5, it can be seen that the SSI is an average over the entire N-dimensional response ensemble (the outer summation). Since the complexity of computing the average over the response ensemble grows exponentially with N, the calculation quickly becomes intractable as the population size increases. Monte Carlo integration enables us to avoid this problem by sampling at random from the response distribution, computing the value of the measure based on this sample, and averaging across all samples to find the final value. This process is repeated until the desired level of precision is reached.

The SSI is defined as:

$$I_{SSI}(\theta) = \sum_{r \in R} p(r|\theta) \left[\sum_{\theta' \in \Theta} p(\theta'|r) \log p(\theta'|r) - p(\theta') \log p(\theta') \right] \tag{A.5}$$

To calculate the SSI for a given stimulus θ, we first sample a vector of neuronal responses $\boldsymbol{r}^k$ (where the superscript k is an index over Monte Carlo samples) from the conditional distribution.

$$\boldsymbol{r}^k \sim p(\boldsymbol{r}|\theta) = \mathcal{N}[\tau \boldsymbol{f}(\theta), Q(\theta)] \tag{A.6}$$

We then calculate $p(\boldsymbol{r}^k|\theta')$ for many values of θ' regularly spaced across the entire stimulus space Θ; this is trivial since the distribution $p(\boldsymbol{r}|\theta)$ is known. We can then apply Bayes' theorem to find $p(\theta'|\boldsymbol{r}^k)$:

$$p(\theta'|\boldsymbol{r}^k) = \frac{p(\boldsymbol{r}^k|\theta')p(\theta')}{\int_\Theta p(\boldsymbol{r}^k|\theta')p(\theta')\, d\theta'} \tag{A.7}$$

Where the integral $\int_\Theta p(\boldsymbol{r}^k|\theta')p(\theta')d\theta'$ is evaluated by adaptive numerical quadrature. We then calculate the specific information sample:

$$I_{SI}^k(\theta) = \int_\Theta p(\theta'|\boldsymbol{r}^k) \log p(\theta'|\boldsymbol{r}^k) - p(\theta') \log p(\theta')\, d\theta' \tag{A.8}$$

This sampling process is repeated many times, and the SSI is found by averaging over the samples:

$$I_{SSI}(\theta) = \frac{1}{n}\sum_{k=1}^{n} I_{SI}^{k}(\theta) \tag{A.9}$$

Where n is the number of Monte Carlo samples. The estimate of the SSI is guaranteed to converge towards the true value as $n \to \infty$. The precision of the estimate was monitored by computing the standard deviation $s_{SSI}(\theta)$ of the Monte Carlo samples. This allowed the standard error of the SSI estimate to be found using the equation for the standard error of the mean:

$$SE_{SSI}(\theta) = \frac{s_{SSI}(\theta)}{\sqrt{n}} \tag{A.10}$$

The standard error decreases as the number of samples increases and the sampling process was halted when the standard error reached a predetermined threshold, or when n reached a predetermined limit.

A.4 SSI and Bayesian specific information

In Chapter 5 we saw how the conditional posterior distribution $P(Z|S)$ could be used to visualise the precision of a population code and its potential for ambiguity or confusion between stimuli that trigger similar responses. The SI, and hence the SSI, are based upon a difference in entropies (Equation 5.12) and a similar stimulus-specific measure, the Bayesian specific information (BSI), could be derived from the posterior conditional distribution:

$$I_{BSI}(s) = H(S) - H(Z|S = s) \tag{A.11}$$

These measures are not the same: the SSI is based on an expected entropy, whereas the BSI is based on the entropy of an expected distribution; the difference is in the order in which the average is taken and the entropy calculated (see Figure A.1). Preliminary investigations indicate that the BSI is less than the SSI for the types of population models studied in Chapter 5 and that, similarly to the SSI, the BSI for peaked tuning curves undergoes a peak-flank transition. A more detailed investigation of the properties of the BSI measure was outwith the scope of this thesis, and might be an interesting topic for future study.

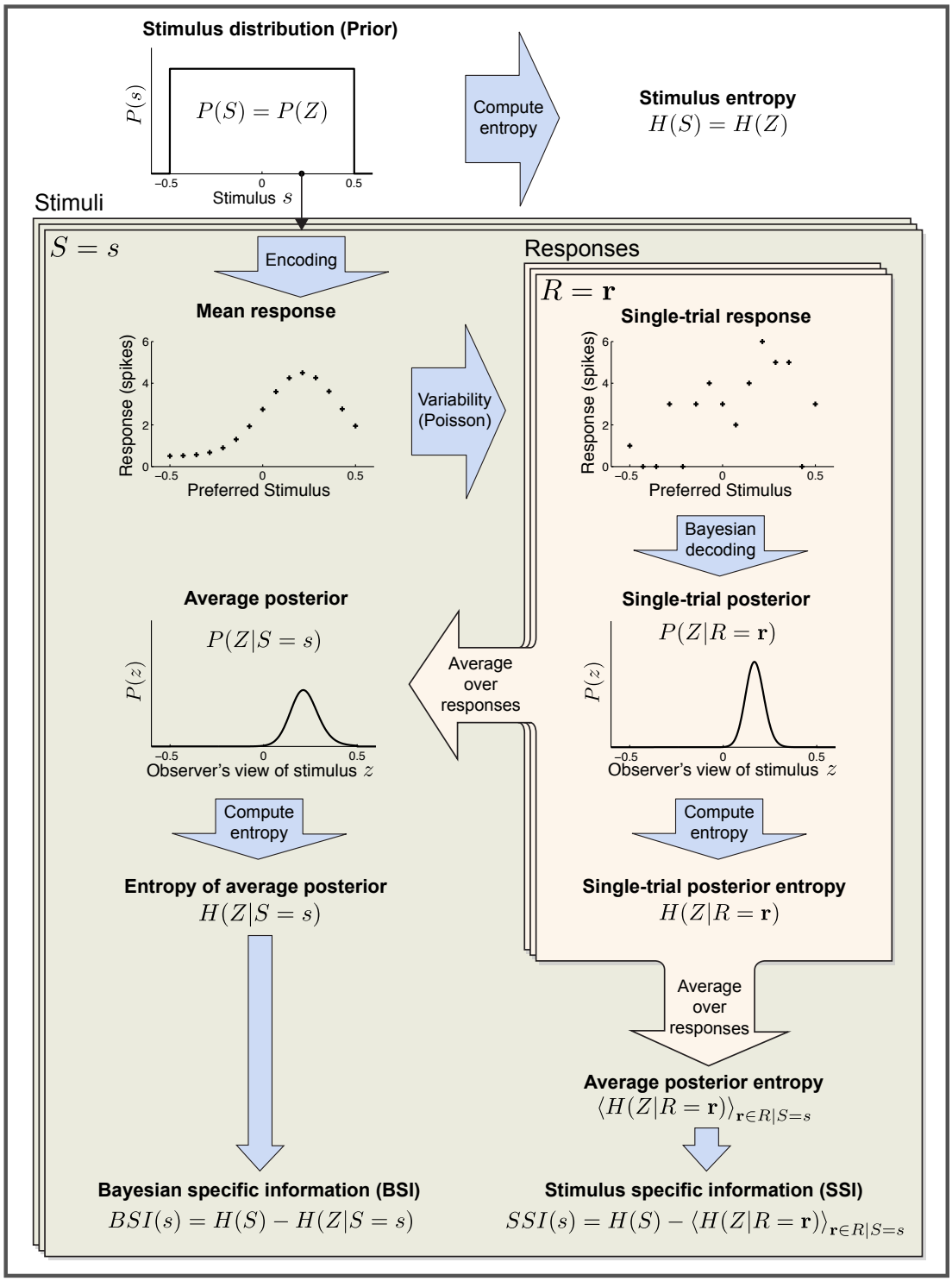

Figure A.1: Overview of the population coding model and definition of the SSI and BSI.

References

Abbott LF, Dayan P (1999). The effect of correlated variability on the accuracy of a population code. *Neural Comput* 11:91–101.

Adrian ED, Zotterman Y (1926a). The impulses produced by sensory nerve endings: Part 2. The response of a single end-organ. *J Physiol* 61:151–71.

Adrian ED, Zotterman Y (1926b). The impulses produced by sensory nerve endings: Part 3. Impulses set up by touch and pressure. *J Physiol* 61:465–83.

Alvarez SA, Levitan S, Reggia JA (1998). Metrics for cortical map organization and lateralization. *Bull Math Biol* 60:27–47.

Averbeck BB, Latham PE, Pouget A (2006). Neural correlations, population coding and computation. *Nat Rev Neurosci* 7:358–366.

Bandyopadhyay S, Shamma SA, Kanold PO (2010). Dichotomy of functional organization in the mouse auditory cortex. *Nat Neurosci* 13:361–8.

Bauer HU, Herrmann M, Villmann T (1999). Neural maps and topographic vector quantization. *Neural Netw* 12:659–676.

Bauer HU, Pawelzik KR (1992). Quantifying the neighborhood preservation of self-organizing feature maps. *IEEE Trans Neural Netw* 3:570–9.

Beck JM, Ma WJ, Kiani R, Hanks T, Churchland AK, Roitman J, Shadlen MN, Latham PE, Pouget A (2008). Probabilistic population codes for Bayesian decision making. *Neuron* 60:1142–52.

Bell G (1982). Behavioral and ecological aspects of gleaning by a desert insectivorous bat *Antrozous pallidus* (Chiroptera: Vespertilionidae). *Behav Ecol Sociobiol* 10:217–223.

Benjamini Y, Yekutieli D (2001). The control of the false discovery rate in multiple testing under dependency. *Ann Stat* 29:1165–1188.

Berens P (2009). CircStat: A MATLAB toolbox for circular statistics. *J Stat Softw* 31:1–21.

Berens P, Ecker AS, Cotton RJ, Ma WJ, Bethge M, Tolias AS (2012). A fast and simple population code for orientation in primate v1. *J Neurosci* 32:10618–26.

Berens P, Ecker AS, Gerwinn S, Tolias AS, Bethge M (2011). Reassessing optimal neural population codes with neurometric functions. *Proc Natl Acad Sci USA* 108:4423–8.

Bethge M, Rotermund D, Pawelzik K (2002). Optimal short-term population coding: when Fisher information fails. *Neural Comput* 14:2317–2351.

Bezdek JC, Pal NR (1995). An index of topological preservation for feature extraction. *Pattern Recogn* 28:381–391.

Bezzi M, Samengo I, Leutgeb S, Mizumori SJ (2002). Measuring information spatial densities. *Neural Comput* 14:405–420.

Binns KE, Grant S, Withington DJ, Keating MJ (1992). A topographic representation of auditory space in the external nucleus of the inferior colliculus of the guinea-pig. *Brain Res* 589:231–42.

Bonnasse-Gahot L, Nadal J (2008). Neural coding of categories: information efficiency and optimal population codes. *J Comput Neurosci* 25:169–187.

Borst A, Theunissen FE (1999). Information theory and neural coding. *Nat Neurosci* 2:947–957.

Britten KH, Shadlen MN, Newsome WT, Movshon JA (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J Neurosci* 12:4745–65.

Brunel N, Nadal J (1998). Mutual information, Fisher information, and population coding. *Neural Comput* 10:1731–1757.

Butts DA, Goldman MS (2006). Tuning curves, neuronal variability, and sensory coding. *PLOS Biol* 4:639–646.

Butts DA (2003). How much information is associated with a particular stimulus? *Network* 14:177–187.

Challis E (2007). Measuring the accuracy of the neural code. Master's thesis, University of Edinburgh.

Chelaru MI, Dragoi V (2008). Efficient coding in heterogeneous neuronal populations. *Proc Natl Acad Sci U S A* 105:16344–9.

Chernoff H (1952). Measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann Math Stat* 23:493–507.

Clifford CW (2002). Perceptual adaptation: motion parallels orientation. *Trends Cogn Sci* 6:136–143.

Cohen YE, Knudsen EI (1999). Maps versus clusters: different representations of auditory space in the midbrain and forebrain. *Trends Neurosci* 22:128–35.

Coppola DM, Purves HR, McCoy AN, Purves D (1998). The distribution of oriented contours in the real world. *Proc Natl Acad Sci USA* 95:4002–6.

Cover TM, Thomas JA (2006). *Elements of information theory*. Wiley-Interscience, Hoboken, NJ, 2nd edition.

Dean I, Harper NS, McAlpine D (2005). Neural population coding of sound level adapts to stimulus statistics. *Nat Neurosci* 8:1684–1689.

Deneve S, Latham PE, Pouget A (1999). Reading population codes: a neural implementation of ideal observers. *Nat Neurosci* 2:740–745.

DeWeese MR, Meister M (1999). How to measure the information gained from one symbol. *Network* 10:325–340.

Doherty KAJ, Adams RG, Davey N (2006). Topological correlation. In *ESANN2006 proceedings - European Symposium on Artificial Neural Networks*.

Dubner R, Zeki SM (1971). Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey. *Brain Res* 35:528–32.

Durant S, Clifford CWG, Crowder NA, Price NSC, Ibbotson MR (2007). Characterizing contrast adaptation in a population of cat primary visual cortical neurons using Fisher information. *J Opt Soc Am A Opt Image Sci Vis* 24:1529–37.

Ecker AS, Berens P, Keliris GA, Bethge M, Logothetis NK, Tolias AS (2010). Decorrelated neuronal firing in cortical microcircuits. *Science* 327:584–7.

Ecker AS, Berens P, Tolias AS, Bethge M (2011). The effect of noise correlations in populations of diversely tuned neurons. *J Neurosci* 31:14272–83.

Engel SA, Rumelhart DE, Wandell BA, Lee AT, Glover GH, Chichilnisky EJ, Shadlen MN (1994). fMRI of human visual cortex. *Nature* 369:525.

Eurich CW, Wilke SD (2000). Multidimensional encoding strategy of spiking neurons. *Neural Comput* 12:1519–29.

Faisal AA, Selen LPJ, Wolpert DM (2008). Noise in the nervous system. *Nat Rev Neurosci* 9:292–303.

Ferrier D (1874). Experiments on the brain of monkeys.—No. I. *Proc. R. Soc. Lond.* 23:409–430.

Fisher RA (1925). Theory of statistical estimation. *Proc Camb Philos Soc* 22:700–725.

Froudarakis E, Berens P, Ecker AS, Cotton RJ, Sinz FH, Yatsenko D, Saggau P, Bethge M, Tolias AS (2014). Population code in mouse v1 facilitates readout of natural scenes through increased sparseness. *Nat Neurosci* 17:851–7.

Galambos R, Davis H (1943). The response of single auditory-nerve fibers to acoustic stimulation. *J Neurophysiol* 6:39–57.

Georgopoulos AP, Schwartz AB, Kettner RE (1986). Neuronal population coding of movement direction. *Science* 233:1416–9.

Goodhill GJ, Sejnowski TJ (1996). Quantifying neighbourhood preservation in topographic mappings. In *Proceedings of the 3rd Joint Symposium on Neural Computation*, pp. 61–82.

Goodhill G, Sejnowski T (1997). A unifying objective function for topographic mappings. *Neural Comput* 9:1291–1303.

Gordon N, Shackleton TM, Palmer AR, Nelken I (2008). Responses of neurons in the inferior colliculus to binaural disparities: insights from the use of fisher information and mutual information. *J Neurosci Methods* 169:391–404.

Goris RLT, Movshon JA, Simoncelli EP (2014). Partitioning neuronal variability. *Nat Neurosci* 17:858–65.

Grinvald A, Lieke E, Frostig RD, Gilbert CD, Wiesel TN (1986). Functional architecture of cortex revealed by optical imaging of intrinsic signals. *Nature* 324:361–4.

Guigon E (2003). Computing with populations of monotonically tuned neurons. *Neural Comput* 15:2115–27.

Guo W, Chambers AR, Darrow KN, Hancock KE, Shinn-Cunningham BG, Polley DB (2012). Robustness of cortical topography across fields, laminae, anesthetic states, and neurophysiological signal types. *J Neurosci* 32:9159–72.

Gutnisky DA, Dragoi V (2008). Adaptive coding of visual information in neural populations. *Nature* 452:220–224.

Haefner RM, Gerwinn S, Macke JH, Bethge M (2013). Inferring decoding strategies from choice probabilities in the presence of correlated variability. *Nat Neurosci* 16:235–42.

Halton JH (1964). Algorithm 247: Radical-inverse quasi-random point sequence. *Commun. ACM* 7:701–702.

Harper NS, McAlpine D (2004). Optimal neural population coding of an auditory spatial cue. *Nature* 430:682–686.

Heller J, Hertz JA, Kjaer TW, Richmond BJ (1995). Information flow and temporal coding in primate pattern vision. *J Comput Neurosci* 2:175–93.

Hietanen MA, Crowder NA, Ibbotson MR (2007). Contrast gain control is drift-rate dependent: an informational analysis. *J Neurophysiol* 97:1078–87.

Hol K, Treue S (2001). Different populations of neurons contribute to the detection and discrimination of visual motion. *Vision Res* 41:685–9.

Hubel DH, Wiesel TN (1959). Receptive fields of single neurones in the cat's striate cortex. *J Physiol* 148:574–91.

Hubel DH, Wiesel TN (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol* 160:106–54.

Hübener M, Bonhoeffer T (2005). Visual cortex: two-photon excitement. *Curr Biol* 15:R205–8.

Jazayeri M, Movshon JA (2006). Optimal representation of sensory information by neural populations. *Nat Neurosci* 9:690–6.

Jenison RL (2000). Correlated cortical populations can enhance sound localization performance. *J Acoust Soc Am* 107:414–21.

Jenison RL, Reale RA (2003). Likelihood approaches to sensory coding in auditory cortex. *Network* 14:83–102.

Johnson DH, Ray W (2004). Optimal stimulus coding by neural populations using rate codes. *J Comput Neurosci* 16:129–38.

Johnson DH, Sinanović S (2001). Symmetrizing the Kullback-Leibler distance. Technical report, Rice University.

Josić K, Shea-Brown E, Doiron B, de la Rocha J (2009). Stimulus-dependent correlations and population codes. *Neural Comput* 21:2774–804.

Kang K, Sompolinsky H (2001). Mutual information of population codes and distance measures in probability space. *Phys Rev Lett* 86:4958–4961.

Kang K, Shapley R, Sompolinsky H (2004). Information tuning of populations of neurons in primary visual cortex. *J Neurosci* 24:3726–3735.

Kaschube M, Schnabel M, Löwel S, Coppola DM, White LE, Wolf F (2010). Universality in the evolution of orientation columns in the visual cortex. *Science* 330:1113–6.

Kaschube M, Wolf F, Puhlmann M, Rathjen S, Schmidt KF, Geisel T, Löwel S (2003). The pattern of ocular dominance columns in cat primary visual cortex: intra- and interindividual variability of column spacing and its dependence on genetic background. *Eur J Neurosci* 18:3251–66.

Kaski S, Lagus K (1996). Comparing self-organizing maps. In von der Malsburg C, von Seelen W, Vorbruggen J, Sendhoff B, editors, *Artificial Neural Networks – ICANN 1996 (Proceedings)*, Vol. 1112 of *Lecture Notes in Computer Science*, pp. 809–814. Springer Berlin / Heidelberg.

King AJ, Bajo VM, Bizley JK, Campbell RAA, Nodal FR, Schulz AL, Schnupp JWH (2007). Physiological and behavioral studies of spatial coding in the auditory cortex. *Hear Res* 229:106–15.

Krüger J, Bach M (1981). Simultaneous recording with 30 microelectrodes in monkey visual cortex. *Exp Brain Res* 41:191–4.

Latham PE, Nirenberg S (2005). Synergy, redundancy, and independence in population codes, revisited. *J Neurosci* 25:5195–5206.

Lee DT, Schachter BJ (1980). Two algorithms for constructing a Delaunay triangulation. *Int J Parallel Prog* 9:219–242.

Li Y, Van Hooser SD, Mazurek M, White LE, Fitzpatrick D (2008). Experience with moving visual stimuli drives the early development of cortical direction selectivity. *Nature* 456:952–6.

Machens CK, Gollisch T, Kolesnikova O, Herz AVM (2005). Testing the efficiency of sensory coding with optimal stimulus ensembles. *Neuron* 47:447–56.

Macke JH, Gerwinn S, White LE, Kaschube M, Bethge M (2010). Gaussian process methods for estimating cortical maps. *Neuroimage* 56:570–81.

Maler L (2009). Receptive field organization across multiple electrosensory maps. ii. computational analysis of the effects of receptive field size on prey localization. *J Comp Neurol* 516:394–422.

McAlpine D, Grothe B (2003). Sound localization and delay lines—do mammals fit the model? *Trends Neurosci* 26:347–50.

McDonnell MD, Stocks NG (2008). Maximally informative stimuli and tuning curves for sigmoidal rate-coding neurons and populations. *Phys Rev Lett* 101:058103.

Metropolis N, Ulam S (1949). The Monte Carlo method. *J Am Stat Assoc* 44:335–41.

Montemurro MA, Panzeri S (2006). Optimal tuning widths in population coding of periodic variables. *Neural Comput* 18:1555–76.

Montemurro MA, Panzeri S, Maravall M, Alenda A, Bale MR, Brambilla M, Petersen RS (2007). Role of precise spike timing in coding of dynamic vibrissa stimuli in somatosensory thalamus. *J Neurophysiol* 98:1871–82.

Montgomery N, Wehr M (2010). Auditory cortical neurons convey maximal stimulus-specific information at their best frequency. *J Neurosci* 30:13362–6.

Nakahara H, Amari SI (2002). Attention modulation of neural tuning through peak and base rate in correlated firing. *Neural Netw* 15:41–55.

Nakahara H, Wu S, Amari S (2001). Attention modulation of neural tuning through peak and base rate. *Neural Comput* 13:2031–47.

Nauhaus I, Benucci A, Carandini M, Ringach DL (2008). Neuronal selectivity and local map structure in visual cortex. *Neuron* 57:673–9.

Nelken I, Chechik G (2007). Information theory in auditory research. *Hear Res* 229:94–105.

Nielsen F (2013). An information-geometric characterization of Chernoff information. *IEEE Signal Proc Let* 20:269–272.

Nikitin AP, Stocks NG, Morse RP, McDonnell MD (2009). Neural population coding is optimized by discrete tuning curves. *Phys Rev Lett* 103:138101.

O'Donnell C, Nolan MF (2011). Tuning of synaptic responses: an organizing principle for optimization of neural circuits. *Trends Neurosci* 34:51–60.

O'Keefe J, Dostrovsky J (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res* 34:171–5.

O'Keefe J, Recce ML (1993). Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus* 3:317–30.

Oram MW, Földiák P, Perrett DI, Sengpiel F (1998). The 'Ideal Homunculus': decoding neural population signals. *Trends Neurosci* 21:259–65.

Panzeri S, Petersen RS, Schultz SR, Lebedev M, Diamond ME (2001). The role of spike timing in the coding of stimulus location in rat somatosensory cortex. *Neuron* 29:769–77.

Panzeri S, Schultz SR, Treves A, Rolls ET (1999). Correlations and the encoding of information in the nervous system. *Proc R Soc Lond B* 266:1001–1012.

Paradiso MA (1988). A theory for the use of visual orientation information which exploits the columnar structure of striate cortex. *Biol Cybern* 58:35–49.

Penfield W, Boldrey E (1937). Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain* 60:389–443.

Polley DB, Steinberg EE, Merzenich MM (2006). Perceptual learning directs auditory cortical map reorganization through top-down influences. *J Neurosci* 26:4970–82.

Pouget A, Deneve S, Ducom JC, Latham PE (1999). Narrow versus wide tuning curves: What's best for a population code? *Neural Comput* 11:85–90.

Purushothaman G, Bradley DC (2005). Neural population code for fine perceptual decisions in area MT. *Nat Neurosci* 8:99–106.

Quian Quiroga R, Panzeri S (2009). Extracting information from neuronal populations: information theory and decoding approaches. *Nat Rev Neurosci* 10:173–85.

Razak KA, Fuzessery ZM (2000). A systematic representation of interaural intensity differences in the auditory cortex of the pallid bat. *Neuroreport* 11:2919–24.

Razak KA (2011). Systematic representation of sound locations in the primary auditory cortex. *J Neurosci* 31:13848–59.

Razak KA, Fuzessery ZM (2002). Functional organization of the pallid bat auditory cortex: emphasis on binaural organization. *J Neurophysiol* 87:72–86.

Remedios R, Logothetis NK, Kayser C (2009). An auditory region in the primate insular cortex responding preferentially to vocal communication sounds. *J Neurosci* 29:1034–45.

Rojer AS, Schwartz EL (1990). Cat and monkey cortical columnar patterns modeled by bandpass-filtered 2D white noise. *Biol Cybern* 62:381–91.

Sachs MB, Abbas PJ (1974). Rate versus level functions for auditory-nerve fibers in cats: tone-burst stimuli. *J Acoust Soc Am* 56:1835–47.

Sakmann B, Creutzfeldt OD (1969). Scotopic and mesopic light adaptation in the cat's retina. *Pflugers Arch* 313:168–85.

Salinas E (2006). How behavioral constraints may determine optimal sensory representations. *PLOS Biol* 4:e387.

Sanger TD (2003). Neural population codes. *Curr Opin Neurobiol* 13:238–249.

Sawtell NB, Williams A (2008). Transformations of electrosensory encoding associated with an adaptive filter. *J Neurosci* 28:1598–612.

Scolari M, Serences JT (2010). Basing perceptual decisions on the most informative sensory neurons. *J Neurophysiol* 104:2266–73.

Seriès P, Latham PE, Pouget A (2004). Tuning curve sharpening for orientation selectivity: coding efficiency and the impact of correlations. *Nat Neurosci* 7:1129–1135.

Seriès P, Stocker AA, Simoncelli EP (2009). Is the homunculus "aware" of sensory adaptation? *Neural Comput* 21:3271–304.

Seung HS, Sompolinsky H (1993). Simple models for reading neuronal population codes. *Proc Natl Acad Sci USA* 90:10749–53.

Shamir M, Sompolinsky H (2004). Nonlinear population codes. *Neural Comput* 16:1105–1136.

Shannon C (1948). A mathematical theory of communication. *The Bell System Technical Journal* 27:379–423.

Silver MA, Kastner S (2009). Topographic maps in human frontal and parietal cortex. *Trends Cogn Sci* 13:488–95.

Sompolinsky H, Yoon H, Kang K, Shamir M (2001). Population coding in neuronal systems with correlated noise. *Phys Rev E Stat Nonlin Soft Matter Phys* 64:051904.

Stein RB, Gossen ER, Jones KE (2005). Neuronal variability: noise or part of the signal? *Nat Rev Neurosci* 6:389–397.

Székely GJ, Rizzo ML, Bakirov NK (2007). Measuring and testing dependence by correlation of distances. *Ann Stat* 35:2769–2794.

Theunissen FE, Miller JP (1991). Representation of sensory information in the cricket cercal sensory system. II. Information theoretic calculation of system accuracy and optimal tuning-curve widths of four primary interneurons. *J Neurophysiol* 66:1690–703.

Tovée MJ, Rolls ET, Treves A, Bellis RP (1993). Information encoding and the responses of single neurons in the primate temporal visual cortex. *J Neurophysiol* 70:640–54.

Victor J, Purpura K (1997). Metric-space analysis of spike trains: Theory, algorithms and application. *Network* 8:127–164.

Victor JD (2005). Spike train metrics. *Curr Opin Neurobiol* 15:585–92.

Vidaurre D, Muruzábal J (2007). A quick assessment of topology preservation for SOM structures. *IEEE Trans Neural Netw* 18:1524–1528.

Villmann T, Der R, Herrmann M, Martinetz TM (1997). Topology preservation in self-organizing feature maps: exact definition and measurement. *IEEE Trans Neural Netw* 8:256–66.

Wark B, Lundstrom BN, Fairhall A (2007). Sensory adaptation. *Curr Opin Neurobiol* 17:423–9.

Wilke SD, Eurich CW (2002). Representational accuracy of stochastic neural populations. *Neural Comput* 14:155–189.

Willshaw D (2006). Analysis of mouse EphA knockins and knockouts suggests that retinal axons programme target cells to form ordered retinotopic maps. *Development* 133:2705–17.

Willshaw DJ, Sterratt DC, Teriakidis A (2014). Analysis of local and global topographic order in mouse retinocollicular maps. *J Neurosci* 34:1791–805.

Wimmer K, Hildebrandt KJ, Hennig RM, Obermayer K (2008). Adaptation and selective information transmission in the cricket auditory neuron AN2. *PLoS Comput Biol* 4:e1000182.

Woolsey C, Walzl E (1942). Topical projection of nerve fibers from local regions of the cochlea to the cerebral cortex of the cat. *Bull Johns Hopkins Hosp* 71:315–344.

Xie X (2002). Threshold behaviour of the maximum likelihood method in population decoding. *Network* 13:447–456.

Yaeli S, Meir R (2010). Error-based analysis of optimal tuning functions explains phenomena observed in sensory neurons. *Front Comput Neurosci* 4:130.

Yarrow S (2008). Accuracy measures for neural population codes with correlated variability. Master's thesis, University of Edinburgh.

Yarrow S, Challis E, Seriès P (2012). Fisher and Shannon information in finite neural populations. *Neural Comput* 24:1740–80.

Yarrow S, Razak KA, Seitz AR, Seriès P (2014). Detecting and quantifying topography in neural maps. *PLoS One* 9:e87178.

Zhang K, Sejnowski TJ (1999). Neuronal tuning: To sharpen or broaden? *Neural Comput* 11:75–84.

Zheng W (2012). Auditory map reorganization and pitch discrimination in adult rats chronically exposed to low-level ambient noise. *Front Syst Neurosci* 6:65.

Zohary E, Shadlen MN, Newsome WT (1994). Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* 370:140–143.

Zrehen S (1993). Analyzing Kohonen maps with geometry. In Gielen S, Kappen B, editors, *Proceedings of ICANN'93, International Conference on Artificial Neural Networks*, pp. 609–612. Springer.