Methods for Analyzing the Information Content of Large Neuronal Populations

by

Palipahana Pahalawaththage Chamanthi Rasangika Karunasekara

First Supervisor: Prof Stefano Panzeri

Second Supervisor: Dr Arno Onken

Submitted to Center for Mind/Brain Sciences (CIMeC)

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

at the

University of Trento

April 2016

ABSTRACT

Deciphering how neurons represent the external world is a fundamental goal in neuroscience. This requires identifying which features in the population response in a single trial are informative about the stimulus. Neurons can code stimuli using both space and time. Individual neurons show differential selectivity to certain stimuli across space at coarse time scales while representing others by modulating their activity at fine time scales. The information content in the population is modified from neural interactions across space and time. While this emphasizes the need to examine population responses across space and time, analyzing a population of hundreds of neurons is challenging when only a limited number of trials are available due to the high dimensionality of the joint spatiotemporal response space. We addressed this by introducing a novel method called space-by-time non-negative matrix factorization. The method describes the population activity with a low dimensional representation consisting of spatial modules, groups of neurons that are coactivated, and temporal modules, patterns that describe how these neurons modulate their spiking across time. The population activity in each trial is described by a set of coefficients, that indicate the level of activation of each spatial and temporal module in the trial. We used this method to analyze datasets from auditory, visual and somatosensory modalities. It identified physiologically meaningful spatial and temporal modules that described how each population coded stimuli in space and time. It further indicated the differential contributions of spatial and temporal dimensions for the population code. Particularly, the first spike latency was demonstrated to be informative at the population level. We refined the method to model the sub-Poisson, Poisson and supra-Poisson variability typically observed in spike counts. This refinement demonstrated enhanced capacity in identifying spatial and temporal modules from empirical data and indicated that the activity of a neural population code stimuli using multiple representations. Our findings indicate that our method is scalable to large populations of neurons and has the capacity to efficiently identify biologically meaningful and informative low dimensional representations.

ACKNOWLEDGEMENTS

This work would not have been possible without the support of many people. I would like to thank my primary advisor Prof. Stefano Panzeri, who always believed in me and encouraged me throughout these years. I would also like to thank Dr. Arno Onken, my second advisor, for the encouragement and collaboration during the years. The data used in this thesis include those that were kindly provided to us by Shuzo Sakata and collaborators at Strathclyde University and Tim Gollisch and colleagues from University of Gottingen. Ioannis Delis from Columbia University collaborated with us on our algorithms. I owe much gratitude to the Marie Curie Initial Training Network, Adaptive Brain Computations (ABC, grant agreement number PITN-GA-2011-290011) that supported my work. I gained much insight from the workshops, lab visits and meetings that took place and I was fortunate to be part of diverse and collaborative research community. I thank everyone in our Neural computation lab. I am very grateful to be part of our group. Many people including Sara Maistrelli, Paola Battistoni and Leah Mercanti helped and advised me in many administrative matters. I would also like to thank all my colleagues and other personnel in IIT, CIMeC and University of Trento with whom I got the opportunity to work and share experience with, who have assisted me and who were good to me in various ways during the years.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

Chapter 1: Introduction

Neural populations from sensory areas (Stopfer et al., 2003; Jones et al., 2007), motor areas (Churchland et al., 2012) to higher level regions (Durstewitz et al., 2010; Mante et al., 2013) exhibit collective dynamics in response to stimuli. Neural population activity can be characterized using two dimensions; space and time. The spatial dimension describes how neurons modulate their tuning and interactions to represent stimuli (Rigotti et al., 2013; Moreno-Bote et al., 2014; Panzeri et al., 2015; Pitkow et al., 2015). The temporal dimension describes how the collective activity of a population of neurons evolves over time, which can contain information that is lost if the fine details of the population activity across time is neglected.(Laurent, 1999; Petersen et al., 2002; Stopfer et al., 2003; Heil, 2004; Gollisch and Meister, 2008; Panzeri et al., 2010a; Zuo et al., 2015). With the recent advancements in recording methods it is now possible to record from hundreds of neurons simultaneously (Buzsáki et al., 2015). An emergent view from the insights gained from these large scale recordings is that the neural population activity consists of a limited number of stereotyped spiking patterns (Nádasdy et al., 1999). Sometimes these groups of neurons tend to fire close together in time, with the relative strength and timing of recruitment of different patterns encoding information about the stimulus features (Luczak et al., 2009; Luczak et al., 2013).

How to extract a biologically meaningful and scalable representation of neural population spike trains in space and time remains an open problem. The purpose of this thesis is to make a contribution towards finding mathematical methods that can accomplish this goal. When developing such a method, first one has to identify the key requirements that the derived representation has to fulfill. It should satisfy many requirements. First, because the brain makes decisions in single trials, it should capture information in single trial spike trains. Second, it should capture most or all information about stimuli with a small number of parameters. Third, the basis functions used to describe single-trial neural activity should be interpretable biologically: in particular, it should decompose neural activity into the constituent stereotyped patterns of firing observed in the data..

Current methods for finding low-dimensional representations of neural activity (Laubach et al., 1999a; Byron et al., 2009; Yu et al., 2009; Churchland et al., 2010; Cunningham and Byron, 2014) such as Principal Component Analysis (PCA), Independent Component Analysis (ICA), or Factor Analysis (FA) are usually applied to firing rate only, neglecting the temporal structure of spike trains, or to trial-averaged data in order to avoid the confounding effects of trial-to-trial spiking.

The central hypothesis of this thesis is that part of the problems with the methods mentioned above can be solved by describing single-trial spatiotemporal firing patterns with a machine learning technique that is well-established, yet largely unexplored for spike train analysis: Non-negative Matrix Factorization (NMF, see (Lee and Seung, 1999)). NMF linearly decomposes a non-negative dataset into a sum over non-negative basis functions using non-negative coefficients. The non-negativity constraint implemented in NMF yields several

advantages: its basis functions and coefficients are, in principle, interpretable as firing patterns and as their strength of recruitment in single trials; it generates sparse representations; and it can cope with non-orthogonal firing patterns such as the partly overlapping ones that may be generated by neural circuits with hard-wired connectivity.

In this thesis we will explore in detail the potentials of NMF for spike train analysis by studying carefully the mathematical basis for application to spike trains such as assumptions to be made regarding the spatial and temporal nature of neural responses, models or neural noise or neural variability. We introduce and refine NMF based methodologies to study population coding that address the above mentioned key requirements. We apply it systematically to many different datasets of neural responses.

Before I proceed further in studying these topics, in this introductory chapter I will review some basic empirical facts of neural population coding and some of the techniques currently used to study it. I will conclude with an overview of the other chapters of this thesis.

1.1 The brain as an information processing machine

The brain is an information-processing machine. It constructs representations of the external world from signals coming from our sensory organs. Computations on these representations lead to the way we perceive the world, remember past events, plan our future and make day-to-day decisions. Our actions and behaviors in turn are initiated from signals that our brain sends to our motor elements (Decharms and Zador, 2000). The fundamental question in neuroscience is to understand how our brain performs these functions.

Neurons are the basic information processing units in the brain. A neuron participates in information processing through changes in its membrane potential. Its membrane potential is generally at a resting value, but can change in a transient stereotypical voltage profile generating an action potential or a spike. Neurons communicate with each other by transmitting action potentials across chemical or electrical synapses. Typically, a neuron receives about 3000 - 10000 pre-synaptic connections from which about 85% are excitatory and the remaining are inhibitory (Mayhew, 1991; Shadlen and Newsome, 1998). Excitatory pre-synaptic spikes cause an increase in the membrane voltage of an excitatory neuron while inhibitory pre-synaptic spikes cause a decrease in its membrane voltage. Even though a neuron receives a large amount of inputs with strong temporal fluctuations, a neuron typically maintains a reasonably low firing rate with variable spike times through a balance in excitation and inhibition (van Vreeswijk and Sompolinsky, 1996; Shadlen and Newsome, 1998; Wehr and Zador, 2003; Okun and Lampl, 2008).

Many neurons involved in representing sensory stimuli are located in the cortex. One dataset we used was recorded from the auditory cortex and another from the somatosensory cortex of rat. The cortex is organized horizontally into six layers and vertically in a columnar architecture. In the six-layer organization, each layer contains a population of cells that are different in terms of sizes, shapes, densities, inputs and outputs. Cells are assigned to layers based on the layer that the cell body lies in. Cells in layer IV in sensory cortices receive input

from afferent thalamocortical connections. Layer V cells connect cortex to sub-cortical structures such as basal ganglia, while neurons in layer VI send efferent connections to the thalamus. In the columnar organization, a column is a vertical cluster of neurons that have similar functional roles (Mountcastle et al., 1957; Mountcastle, 1997). An example is a barrel column in rodent somatosensory cortex where inputs from the thalamus relating to a single whisker terminate at a distinct area in the layer IV of the somatosensory cortex called barrel column (Woolsey and Van der Loos, 1970). The number of neurons per column varies with the distance of the respective whisker from the ground (Meyer et al., 2013). Another example is the columnar organization found in V1 that is defined by the ocular dominance and the orientation specificity (Hubel and Wiesel, 1959, 1962; Obermayer and Blasdel, 1993). While a similar architecture is found in many other cortical regions (Mountcastle, 1997), the reasons for the existence of such columnar organization is debated (Horton and Adams, 2005).

Experimental studies have found that neuronal connectivity can be highly clustered and neurons can form sub-networks at a fine-scale (Song et al., 2005; Yoshimura and Callaway, 2005; Yoshimura et al., 2005; Perin et al., 2011). The synaptic weight that specifies the strength of the connection between the pre-synaptic neuron and the post-synaptic neuron could be strong in a few synaptic connections that are more clustered (Song et al., 2005). Interconnected sub-networks could also share a higher degree of common excitatory input (Yoshimura et al., 2005) and process related sensory information with high functional specificity (Ko et al., 2011). Such sub-networks define localized organization at a finer scale within the cortex, which can give rise to coactivations and interactions between clustered neurons. Recent studies have proposed that the population activity is constrained to display stereotypical activity patterns that are recruited to encode sensory stimuli and to coordinate motor activity (Luczak et al., 2009; Sadtler et al., 2014).

1.2 Brain states

Neurons in the resting brain show spontaneous spiking. The global structure of the spontaneous activity depends on the state of the brain. Typically, the state of the brain is defined according to the observed variations in the spiking activity on a time scale of seconds. Traditionally, it is classified into two broad categories using observations related to the sleep cycle (Steriade et al., 1993; Steriade et al., 2001). During slow-wave sleep, the spiking activity of populations of neurons show large scale structured fluctuations where periods of concerted spiking activity is interleaved with periods of silence. The periods in which neurons show activity are called up-states and the periods in which the neurons are silent are called down-states. Up-states and down-states can range in a scale of seconds under certain anesthesia (Steriade et al., 1993) to tens of milliseconds in awake animals (Luczak et al., 2009; Harris, 2013). Such fluctuations can give rise to high correlated variability between neurons (Curto et al., 2009; Pachitariu et al., 2015). When neurons show this type of activity and the summed population activity undergoes slow low frequency fluctuations (~< 4 Hz), the brain is said to be in a synchronized state. On the other hand, during waking or in rapid eye movement sleep, these slow fluctuations are no longer visible. Then the brain is said to be in a desynchronized state. The brain state at any moment can be at any point between a highly

synchronized and a highly desynchronized state. Mechanisms such as attention can cause changes in neuromodulation that result in changes of the state of the brain at a local level (Harris and Thiele, 2011). Furthermore, when experiments are conducted under anesthesia such as urethane, spontaneous changes in the brain state could be observed (Clement et al., 2008; Curto et al., 2009). One of our datasets exhibits such a spontaneous state change. The state of the brain is relevant for sensory processing because the responses of neurons to external sensory inputs are shaped from the state of the cortex. Salient stimuli such as auditory clicks and whisker deflections generate large responses irrespective of the state of the brain while temporally extended stimuli are processed efficiently only in the desynchronized state (Harris and Thiele, 2011). Thus, there is a considerable influence of the endogenous brain state on the neural activity.

Scientists conduct experiments to understand how external stimuli are presented in the brain and how computations on these representations lead to cognitive functions and motor actions. Based on what we know of the structure of the brain, how it represents the outside world and performs computations, we record responses from a population of neurons in the brain region that we believe to be involved in the task that we wish to investigate. We look at the recorded activity and try to decipher how the recorded population of neurons modulates their spiking to perform the representation or the computation using a method suitable to analyze the recorded neural activity. We further explore how the states of the brain shape this activity in time. We will now look at some ways in which neurons are known to represent the external world, or more specifically, how they represent the set of stimuli we present during the experiment. Methods that are used to analyze neural activity, which we will discuss later, are designed to suit what we know of these representations.

1.3 Neural population codes

The set of response properties that carry information about the stimuli is called the neural population code (Panzeri et al., 2015). There are many features in the recorded population responses that can convey information about stimuli (Onken et al., 2014; Panzeri et al., 2015).

Initial studies that explored neural coding were based on single neuron recordings or recordings from small neuron ensembles. Typically the response of a neuron was defined in terms of the average firing rate, using the total number of spikes of the neurons in a time window which spanned from tens to hundreds of milliseconds. Since the neural responses display variability in spiking from one trial to another, the trial averaged firing rate was typically used. The trial-averaged firing rate of each neuron was plot against the stimulus feature to form a tuning curve. A common observation was that neurons in the population selectively modulate their firing rates to different levels of the feature (Georgopoulos et al., 1983; Maunsell and Van Essen, 1983; Miller et al., 1991; Lewis and Kristan, 1998). An example is the coding of movement directions by motor neurons in the proximal arm area

(Georgopoulos et al., 1982; Georgopoulos et al., 1983; Edelman et al., 1984) shown in Figure 1.1A shows the raster responses of a cell to five trials each recorded when a monkey made hand movements from a central position to the eight equi-spaced directions indicated at the center. The firing rate of the cell shown in the tuning curve in Figure 1.1B is clearly modulated to the movement direction and defines a rate code for movement direction. In higher cortical areas neurons show mixed selectivity to a range of task related parameters simultaneously (Rigotti et al., 2013). The selective modulation of the firing rate across a neural population is a form of stimulus coding across the spatial dimension.

Parallel studies found evidence for the coding in the temporal dimension. Neurons can use precise spike timing on the order of a few milliseconds to encode stimulus information that cannot be accounted for by the firing rates alone (Panzeri et al., 2010b; Onken et al., 2014). A simple example where spike timing is used is the latency code (Gawne et al., 1996). In this code, information about stimulus features is coded by means of the time of the first spike (Johansson and Birznieks, 2004; Gollisch and Meister, 2008). An illustration of a first spike latency code is shown in Figure 1.2. The cell shown responds with four spikes to all six stimuli and thus does not show any modulation in the total spike count between stimuli. However, the first spike latency varies between stimuli and forms the latency tuning curve shown in Figure 1.2B. A coding scheme based on latency is inherently fast and can be energy efficient since only a few spikes are needed to encode information. However, they can be susceptive to spike failure and jittering. When considering full responses, studies have divided the full responses into smaller bins forming spike words, and have investigated how the information in timing change with the size of these bins (Panzeri et al., 2001).

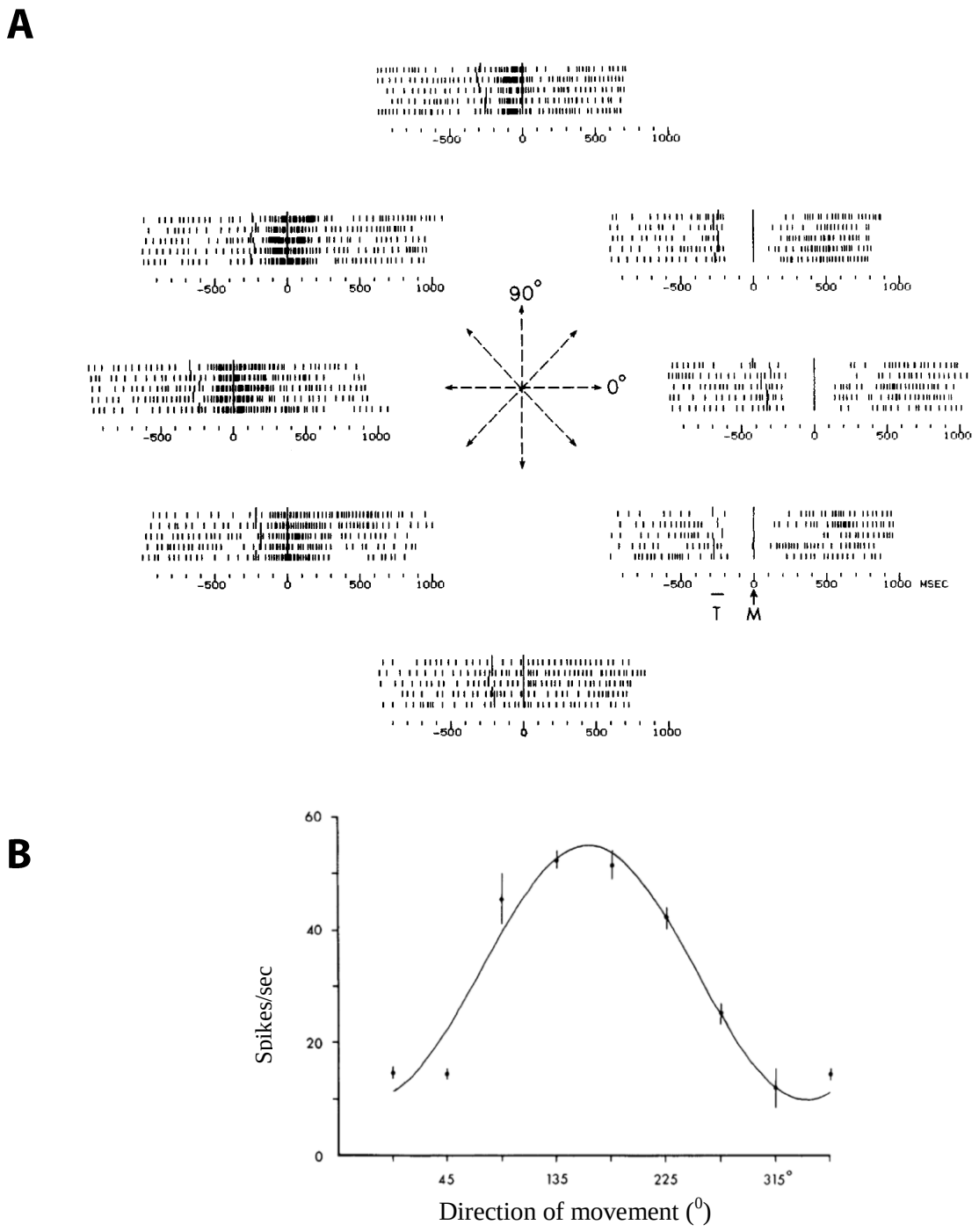

Figure 1.1: Coding of the movement direction using average firing rates A: Raster plots of responses recorded from a neuron in the proximal motor area of a monkey when the hand was moved from a central position to the periphery in each of the eight directions indicated in the center. The movement is made after the target (T) is shown at the time indicated by longer vertical lines. Trials are aligned to the movement onset M also indicated by longer vertical bars. B: The tuning curve of the cell in A showing the average firing rate of the neuron for the eight directions (The image is in (Edelman et al., 1984) and was adapted from (Georgopoulos et al., 1982b))

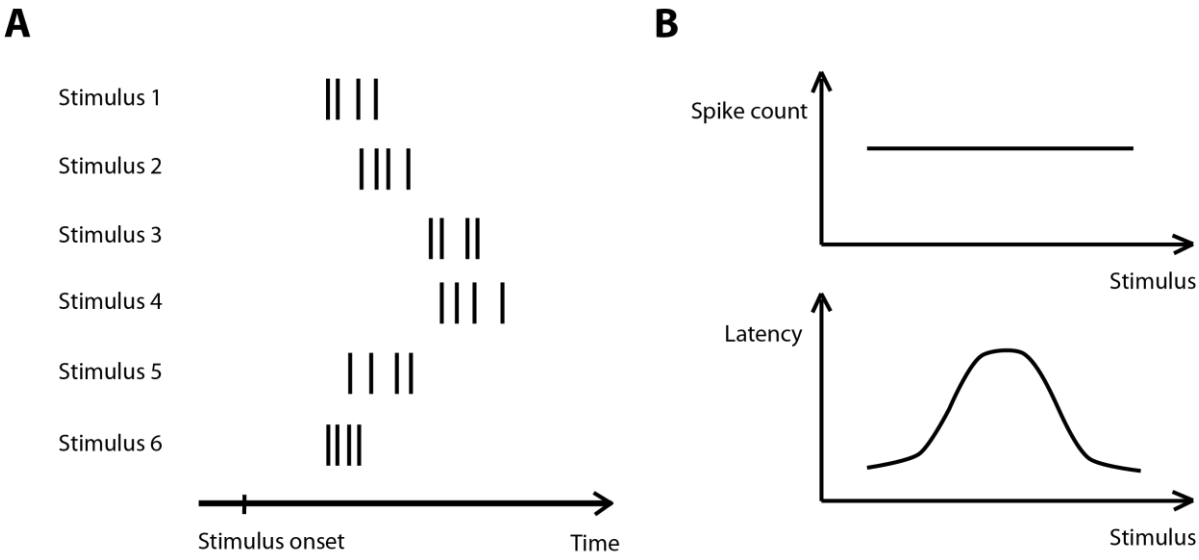

Figure 1.2: Illustration of latency code as an example of a spike timing code. A: Responses of a hypothetical neuron to six stimuli B: The tuning curves constructed using the spike counts of the cell (top) and the first spike latency (bottom).

When identifying neural codes that require the knowledge of either the stimulus onset or the segmentation of the response into informative bins, the natural question is how biologically realistic such a scheme could be. Is it possible for the brain to formulate this knowledge to use such a coding scheme? Some experimental studies have shown indications that such knowledge is available at the cellular level (Chase and Young, 2007; Onken et al., 2014; Panzeri et al., 2014). Subpopulations of neurons within the auditory cortex have been identified that respond early and reliably in a stimulus invariant way (Brasselet et al., 2012). When the response onsets of the remaining stimulus modulated neurons were set to the response onsets of these stereotyped neurons, about 95% of the spike timing information could be extracted indicating that this scheme could serve as a reliable way of identifying the stimulus onset. Similarly, in the somatosensory system, when single whiskers are deflected, the whisker position can be identified on a millisecond time scale by using the latency from the time at which a large portion of the neurons in the columns fired synchronously (Panzeri and Diamond, 2010). To address the question of how the brain could segment long time intervals into informative segments, (Kayser et al., 2009) indicated that when the spike train was partitioned into segments using the phase of the theta (2 - 6 Hz) frequency band in the local field potential forming a phase-partitioned code, it carried almost similar information as a time-partitioned code using the laboratory clock, which was large compared to information only in the total spike count. Finally, there is experimental evidence that timing codes could be read out from downstream neurons (Haddad et al., 2013; Uchida et al., 2014).

Often patterns across spatial and temporal dimensions can code information (Stopfer et al., 2003; Jones et al., 2007; Rabinovich et al., 2008; Shusterman et al., 2011; Churchland et al., 2012; Mante et al., 2013). For a population of N neurons, this is visible as trajectories of activity in the N-dimensional space and state transitions in an ensemble state space. Often

some form of dimensional reduction method is used to analyze and visualize this activity. Visualization is done in the space defined by the low-dimensional components (latent components). In the illustrative example shown in Figure 1.3, the activity of N neurons is represented in the 3-dimensional latent component space. Each stimulus evokes a different trajectory in this space. Thus, each stimulus is coded by a different pattern of population activity evolving in space and time.

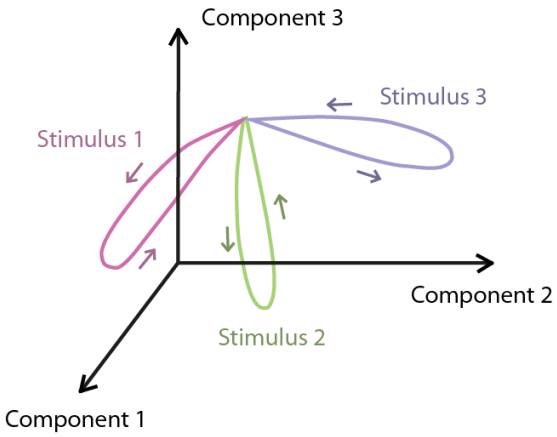

Figure 1.3: Illustration of spatiotemporal trajectories coding stimuli. The high dimensional neural activity is visualized using a dimensionality reduction method in the three dimensional latent component space. Each stimulus evokes a different trajectory in the component space.

Collective behavior becomes increasingly important as the number of recorded neurons increases. As the size of the population becomes large, spikes of individual trials can be predicted more accurately using models based on pairwise interactions between neurons compared to models that predict spikes based on external stimuli (Stevenson and Kording, 2011). Thus, we next briefly look at the effects of correlated behavior in neural coding.

1.4 Neural correlations

Correlated spiking is typically defined as signal correlations and noise correlations. Signal correlations are defined as the correlation between mean spike counts of neurons to stimuli. Correlations in the trial-to-trial variability between spike counts during a single stimulus are called as noise correlations. Much effort has been directed towards understanding the effect of noise correlations on neural coding (Averbeck et al., 2006; Latham and Roudi, 2013). This is important because correlations can potentially either increase or decrease stimulus information in the neural population when compared to when all neurons are independent (Averbeck et al., 2006) and even small correlations between neurons could have a large impact at the population level (Zohary et al., 1994). Accumulation of results from many theoretical studies indicate that the effect of correlations depends on the correlation structure

(Abbott and Dayan, 1999; Sompolinsky et al., 2001; Wilke and Eurich, 2002; Shamir and Sompolinsky, 2006; Josic et al., 2009; Ecker et al., 2011; Moreno-Bote et al., 2014). When a population has similar tuning curves and has positive noise correlations, correlations limit the information coding capability (Sompolinsky et al., 2001), while such a limit does not exist when the neurons have heterogeneous tuning (Shamir and Sompolinsky, 2006). (Moreno-Bote et al., 2014) formulated that correlation structures that limit the level of information coding has a component in the covariance matrix that is proportional to the product of the derivatives of the population responses. An illustrative example from (Moreno-Bote et al., 2014) is shown in Figure 1.4. It shows the mean population activity $f(s)$ in the N-dimensional space. The distribution of the trial-to-trial variability is shown by the yellow curve. When this distribution has a small curvature along $f(s)$, it can be approximated with the blue distribution that lies on the tangent to the mean population response indicating that the covariance matrix is proportional to $\mathbf{f'f'}^{T}$.

Correlational structures are not static. Experimental studies indicate that they can change dynamically with variations in attention and brain states (Steinmetz et al., 2000; Gutnisky and Dragoi, 2008; Cohen and Maunsell, 2009; Curto et al., 2009; Pachitariu et al., 2015). Certain correlation structures that show stimulus dependent changes can give substantial increases in the level of information coding (Franke et al., 2016; Zylberberg et al., 2016). Thus, noise correlations can shape neural coding in diverse ways. The presence of correlations that affect neural coding places a practical constraint known as the curse of dimensionality against increasing the spatiotemporal space in all methods that use the probability distribution of the population responses conditioned on a stimulus. We explain this in more detail in the next section.

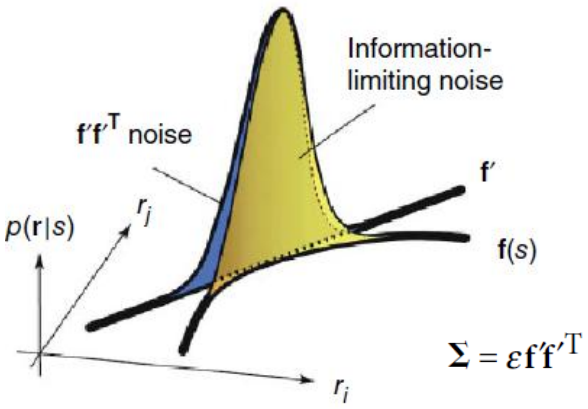

Figure 1.4: Information limiting correlations An illustration of how the structure of the covariance matrix can limit information capacity (Adapted with permission from (Moreno-Bote et al., 2014)).

1.5 Methods to analyze neural populations

Now we will look at methods that are commonly used to analyze spatiotemporal activity of neurons. In this thesis we propose methods based on non-negative matrix factorization (NMF) to analyze population activity. Although NMF based methods are well established to study large scale datasets (Cichocki et al., 2009), these methods have only been used in a few studies to analyze neural data (Kim et al., 2005; Overduin et al., 2015; Wei et al., 2015). Thus in this section we review other commonly used methods to analyze neural activity and identify the contribution that our method provides compared to other methods.

Since spike times are discrete events, from the statistical point of view, a spike train is a point process (van Vreeswijk, 2010). Multiple simultaneously recorded (parallel) spike trains form a multi-dimensional point process time series (Brown et al., 2004). Methods for analyzing these processes are still under development. The key challenge in this quest is that the number of parameters that needs to be estimated for a given method increases exponentially when the dimensionality of the response space, as specified by the number of neurons and the duration to be analyzed, increases. This is known as the curse of dimensionality. For example, in an experiment where N is the number of recorded neurons and the spikes of each neuron are discretized into T time bins, there are 2^{NT} possible spatiotemporal spike patterns that could appear in any trial. If there are 100 neurons and 100 time bins, this would give $2^{100x100} \approx 10^{60}$ possible patterns. If one wants to know whether there are correlated firing patterns between neurons, theoretically, the probability of seeing a particular pattern could be compared to the probability of the particular pattern occurring by chance if all neurons spiked independently. However, it is impossible to calculate the probability of seeing a pattern just by counting the number of times the pattern appears in experimental data because there are 10^{60} possible patterns and it is impossible to record enough trials in an experiment. Therefore, from this practical limitation on the amount of data that is available, there is a limit to the number of parameters or, equivalently, the level of detail that can actually be estimated from the available data. Different methods employ different techniques and assumptions to overcome this problem, for example, by assuming an underlying statistical model of the data, by using a dimensionality reduction technique or by pooling data in one spatiotemporal dimension.

In the following, we first briefly summarize methods that use statistical testing to identify salient spatial and temporal patterns from data. Then we detail model-based methods to explore population activity focusing mainly on the generalized linear model and the maximum entropy model. Next we look at decoding and information theory, which are two general methods that can be used to study neural coding in space and time. Finally we look into dimensionality reduction methods commonly used in neuroscience studies.

1.5.1 Statistical methods

The presence of spatial, temporal or spatiotemporal activity from data can be identified using some form of statistical test. Task specific synchronization time points from the visual cortex (Maldonado et al., 2008), motor cortex (Riehle et al., 2000; Kilavik et al., 2009) and prefrontal cortex (Grun et al., 2002) were identified using unitary event analysis (Grun, 2009; Grün et al., 2010). a method that evaluates how synchronous spiking events at the population level change across time by comparing the number of spike synchronizations that occur in a specified time window with the number of spike synchronizations expected to occur if the spike trains of the neurons were independent. Optimized search algorithms for large scale datasets, such as used for frequent itemset mining, can aid statistical testing to identify groupings of neurons that occur more than a specified number of times (Picado-Muino et al., 2013; Torre et al., 2013).

1.5.2 Model-based methods

More often experimental studies use model based approaches to study population activity. Two commonly used models are generalized linear models (GLM) and maximum entropy models (MEM). They model the probability distribution of response patterns that is conditional on a combination of past and/or present stimulus, spiking history and spiking interactions. The model parameters are estimated from the data to obtain the conditional response probability distribution of the data.

Generalized linear models (GLM)

GLMs are phenomenological models that generally assume that the responses of the neurons in the current time bin are independent of each other. The response of one neuron in the current time bin may depend on the current and past stimulus and the past spiking history of the neuron itself and other neurons as shown in Figure 1.5. The probability of the response of the i^{th} neuron conditional on the stimulus s is given as (Truccolo et al., 2005; Pillow et al., 2008; Latham and Roudi, 2013),

$$q\left(r_i(t)\,|\,s\right)=\frac{1}{Z}\exp\left[K_i[s]r_i(t)+\sum_{t'<t}h_i(t')r_i(t-t')+\sum_{t'<t,j\neq i}j_{ij}(t')r_j(t-t')\right] \quad (1.1)$$

The first term inside the exponential takes into account the dependency of the response on the stimulus using the stimulus filter $K_i[s]$. The second term uses the post-spike filter $h_i(t')$ to model effects such as refractoriness, burstiness and adaptation from the past spiking by the neuron. The third term captures the effects from spiking of other cells in the recent past. Z is a normalizing term. Parameters of the model that have to be estimated from the data are those associated with the three filters. They are usually estimated through maximum likelihood (Paninski, 2004; Truccolo et al., 2005). Once they are estimated and the model is fitted, it is able to predict the probability of spiking of a neuron when the stimulus and past spiking

histories are known. It has been used in many experimental studies investigating different cortical areas such as retina (Pillow et al., 2008), lateral geniculate nucleus (Babadi et al., 2010), motor cortex (Truccolo et al., 2010) and auditory cortex (Calabrese et al., 2011). The model with and without the dependencies on the spike trains of other neurons can be used to decode the stimuli as a way of estimating the effect of correlations (Pillow et al., 2008). GLM models were generalized to model the dependency of the neural activity on unobserved internal and external states in (Lawhern et al., 2010; Escola et al., 2011; Pfau et al., 2013). We will discuss state-spaced based analysis methods in section 1.5.5.

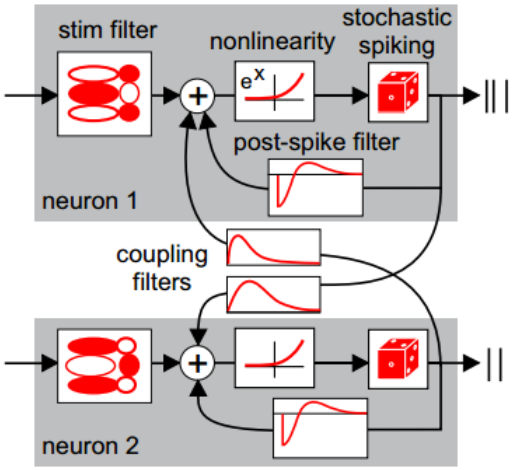

Figure 1.5: Illustration of the generalized linear model (GLM) for two neuron network. GLM model the response of a neuron conditional on a stimulus. The response of each neuron depends on the stimulus through the stimulus filter the dynamics of its own past spiking through the post-spike filter and the spiking dynamics of the rest of the network through coupling filters. The instantaneous firing rate is produced using a nonlinearity on the summed filter outputs. Spikes are generated using the instantaneous firing rate. (Adapted with permission from (Pillow et al., 2008))

Maximum entropy models (MEM)

A MEM is the least structured model (and thus having the maximum entropy) giving the probability distribution of the population response in one time bin that is obtained subject to some predefined constraints (Jaynes, 1957; Schneidman et al., 2006; Shlens et al., 2006). Typically, the response is defined using a small time window where a neuron spikes at most once. Then the response is constructed as a binary word where one denotes the spiking of the neuron. Then if the mean firing rates and pairwise correlations are used as the constraints on the model, the probability distribution of the population response conditional on the stimulus is given by (Latham and Roudi, 2013),

$$q\left(\bar{r}(t)\,|\,s\right)=\frac{1}{Z(t)}\exp\left[\sum_i h_i[s]r_i(t)+\sum_{j<i}J_{ij}[s]r_i(t)r_j(t)\right] \qquad (1.2)$$

The first term constrains the model to have the same mean firing rates for each neuron as in the data while the second term constraints the model to have the same pairwise correlations between pairs of neurons as in the data. $Z(t)$ is a normalization factor called partition function. The model in this formulation treats the responses as stationary in time and thus does not consider the temporal dimension. However, it has been extended to include the temporal dimension by estimating the spatiotemporal probability distribution instead of the spatial distribution in (Nasser et al., 2013). Once the model is fitted, the effect of correlations could be investigated for example in an information theoretic framework using the reduction of the entropy of the probability distribution when the correlations are included in the model compared to when only the mean firing rates are included (Schneidman et al., 2006), or using mutual information between the responses and stimuli with different order of correlations included in the model (Ince et al., 2010). MEM models have been used to investigate correlations in the retina (Schneidman et al., 2006; Shlens et al., 2006; Ganmor et al., 2011; Granot-Atedgi et al., 2013; Tkacik et al., 2014), in V1 (Ohiorhenuan et al., 2010), motor cortex (Truccolo et al., 2010) and somatosensory cortex (Ince et al., 2010). As pointed out in (Roudi et al., 2009), the bin size is important for MEM models and predictions made for larger populations using models fitted to small populations or to larger time bins while assuming stationary may not always be true.

1.5.3 Decoding

Other general approaches to explore the activity of neural populations include decoding and information theory. Decoding takes a set of population response features in a trial and predicts which stimulus elicited them. In order to make the predictions, the decoder is first trained using a part of data that is not used for making predictions. Decoding is a generic method that can take into account different response features in space and time such as firing rates (Jazayeri and Movshon, 2006; Wu et al., 2006), spike patterns (Brown et al., 1998) and correlations (Pillow et al., 2008). By evaluating how the inclusion of different response properties changes the stimulus predictions, it is possible to identify how a particular response property contributes to stimulus encoding. Many decoding algorithms such as nearest-neighbor algorithms, Fisher linear discriminant algorithms, support vector machines, Bayesian decoding algorithms, and maximum-likelihood decoding algorithms exist (Dayan and Abbott, 2001; Quian Quiroga and Panzeri, 2009). The decoding performance may be dependent on the particular decoding algorithm and different algorithms could give different results. The algorithm could fail to decode the stimuli if the response space is very large or if incorrect assumptions are made about the data (Quian Quiroga and Panzeri, 2009). When the response space is large, it is possible to use a dimensionality reduction method to identify representative features from the data. The ideal dimensionality reduction method to use

would be one that gives reduced representations that are representative of the property being investigated and are easily interpretable in relation to the original data.

1.5.4 Mutual information

Mutual information (MI) is another generic method commonly used to study population activity. MI $I\left(S;\bar{R}\right)$ between a set of population response features $\bar{R}$ and the stimuli set S quantifies the reduction of uncertainty of a stimulus that can be gained from observing the response features in a single trial (Ince et al., 2010). It is calculated using the probability of each stimulus $P\left(s\right)$, the probability of observing the response features across all stimuli $P\left(\bar{r}\right)$ and the conditional probability of observing the response features for a given stimulus $P\left(\bar{r}\,|\,s\right)$ using,

$$I\left(S;\bar{R}\right)=\sum_{s\in S}P\left(s\right)\sum_{\bar{r}\in\bar{R}}P\left(\bar{r}\,|\,s\right)\log_2\frac{P\left(\bar{r}\,|\,s\right)}{P\left(\bar{r}\right)} \tag{1.3}$$

MI naturally accounts for all orders of interactions in the population (Ince et al., 2010). Breakdown of the full MI into subcomponents in (Pola et al., 2003) as shown below quantifies the impact of different correlative interactions in the population.

$$I\left(S;\bar{R}\right)=I_{lin}+I_{sig-sim}+I_{cor-ind}+I_{cor-dep} \tag{1.4}$$

I_{lin} quantifies the information encoded if all cells were independent. $I_{sig-sim}$ specifies how much of the information is redundant due to signal correlations that arise because the mean responses of different neurons have similar selectivity for stimuli. $I_{cor-ind}$ and $I_{cor-dep}$ specify the information gain or loss due to noise correlations between neurons; $I_{cor-ind}$ is due to correlations that do not depend on the stimulus while $I_{cor-dep}$ arises because of correlations that depend on the stimulus.

Theoretically, MI is a generic method that can be used to evaluate any particular response feature to determine whether it is important in encoding stimuli. MI can be evaluated for different response definitions (for example, binning spikes in one 100 ms bin and thereby evaluating the MI in firing rate vs. binning spikes in 5 ms bins and thereby evaluating the MI in spike precision). However, when the response dimension (the number of neurons and the number of time bins) is high, MI suffers from the curse of dimensionality because the number of trials is not sufficient to evaluate the probabilities accurately. This gives rise to a systematic error (or a bias) between the true MI and the estimated MI from the limited number of trials available. Various bias correction techniques exist (Panzeri et al., 2007; Ince et al., 2010). However, these methods still set a limit for about 35 spatiotemporal

combinations (for example, 7 time bins in 5 neurons) when working with a typical number of trials (~30-60 trials) (Onken et al., 2014). Therefore, it is difficult to directly generalize MI calculation to high dimensional spatiotemporal response patterns.

Some typical approaches to circumvent this problem are either to use a pooling strategy, to use an approximate probability distribution or to use a transformation or a dimensionality reduction method to reduce the dimensionality of the response space. In the first approach, all spikes of one neuron in a trial could be pooled together (i.e. pooled across time) or the responses of all neurons in one time bin could be pooled together (i.e. pooled across space) (Arabzadeh et al., 2004; Ince et al., 2013). In these cases, the population response $\bar{R}$ is transformed to a pooled response $f(\bar{R})$ as $\bar{R} \rightarrow f(\bar{R})$ (Quian Quiroga and Panzeri, 2009). Since according to the data processing inequality, any transformation can only decrease the information in the original responses (Cover and Thomas, 2006), the MI calculated from pooling across dimensions gives a lower estimate of the information in the population. The pooling approach is only suitable if the pooled dimension has little contribution to the information encoding of the stimulus and is not suitable as a general analysis of the population activity. Under the second approach, a model such as MEM could be used as an approximate conditional probability distribution (Ince et al., 2010; Ohiorhenuan et al., 2010) or MI can be calculated under a framework of underlying Gaussian probability distributions (Yu et al., 2010; Crumiller et al., 2011). In these cases, how close the calculated MI would be to the true MI depends on how close the approximate probability distribution is to the true probability distribution. In the third approach, a decoder is trained to predict the stimulus from response features. The predicted stimulus of a decoder S^p is used as a transformation of the population response $\bar{R}$ as given by $\bar{R} \rightarrow S^p$ (Quian Quiroga and Panzeri, 2009) to get a lower estimation for MI. Alternatively, the response dimensionality can be reduced by a dimensionality reduction method such as principal component analysis (Optican and Richmond, 1987; Zuo et al., 2015). To directly relate MI values to response features, the dimensionality reduction method has to give a meaningful representation of the data.

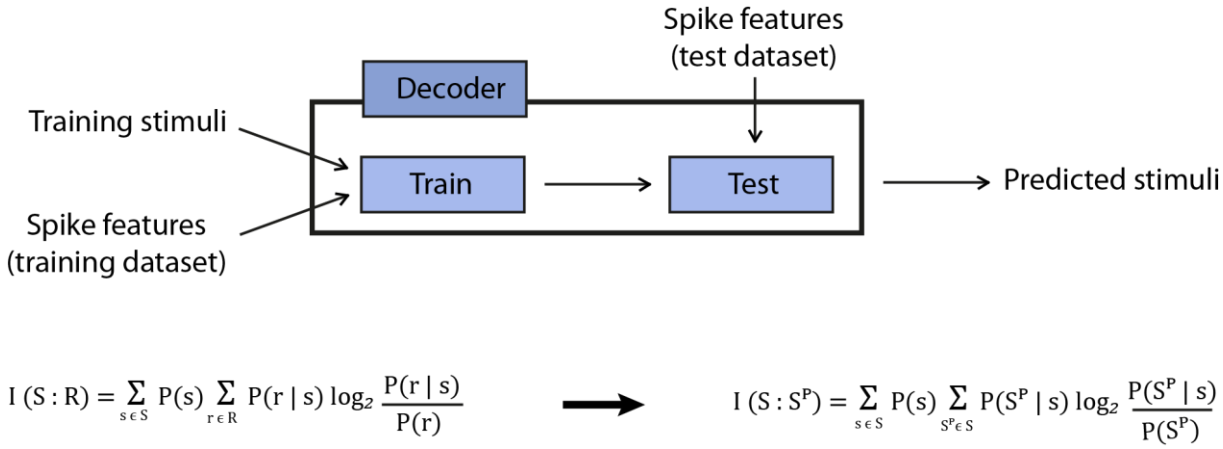

Figure 1.6: Illustration of extending mutual information (MI) to larger populations using a decoding framework. The population response $\bar{R}$ is transformed into the predicted response S^p from the decoder which is used for the MI calculation.

1.5.5 Dimensionality reduction methods

Dimensionality reduction methods aim to describe a dataset that has a large number of data items using a few parameters (often called modules, bases, factors, hidden components or latent components) that are representative of the property being investigated (Cunningham and Byron, 2014). They are becoming increasingly important with the growth in the number of neurons that can be recorded simultaneously. Each of these methods is formulated to optimize a criterion that is specific to the method. Depending on this criterion, different methods can provide different viewpoints into a single large scale dataset. When using to analyze neural activity in space and time, dimensionality reduction methods can be broadly classified into two categories; static dimensionality reduction methods and dynamic methods (Roweis and Ghahramani, 1999; Churchland et al., 2007). Static dimensionality reduction methods first perform dimensionality reduction and subsequently formulate the time progression of the activity using the time progression of the reduced dimensional latent components. Dynamical methods on the other hand explicitly model the time progression of the neural activity, typically as the progression through a set of states.

Next we will discuss three commonly used dimensionality reduction methods; principal component analysis (PCA), independent component analysis (ICA) and factor analysis (FA) and several dynamic dimensionality reduction methods.

Static dimensionality reduction methods

Principal component analysis (PCA)

PCA extracts a set of low-dimensional components that best capture the variance in the data (Jolliffe, 2002). The components are ordered such that the first PCA component captures the highest level of variability in the dataset. The remaining components are ordered in the decreasing order of variance. All components are orthogonal to each other. A simple example of PCA applied on the spike counts recorded from two auditory cortical neurons is shown in Figure 1.7A. PCA finds that the direction that has the maximum variance in the spike counts d_1 and the direction orthogonal to it d_2. d_1 is the low dimensional component that best preserves the covariance between the spike counts.

PCA and its variants have been widely used in neuroscience (Chapin and Nicolelis, 1999; Hu et al., 2005; Mazor and Laurent, 2005; Peyrache et al., 2010; Lopes-dos-Santos et al., 2011; Ahrens et al., 2012; Churchland et al., 2012). A model of PCA using Hebbian learning in linear neurons was implemented in (Oja, 1982). When applying to neural spike counts, PCA is typically preceded by a processing step such as z-scoring of the spike counts, temporal smoothing of spike counts, trial averaging or a square-root transformation (Cunningham and Byron, 2014). This is because PCA is designed to capture variability, it can capture spiking variability, firing rate variability or trial-to-trial variability that is not representative of the

population activity, but can confound the interpretation of the low dimensional decomposition. The low dimensional representation can also be biased by the most active neurons since the variance of spike counts increases with the mean spike count (Byron et al., 2009; Cunningham and Byron, 2014).

A variant of PCA, demixed PCA (Kobak et al., 2014) is designed to address the mixed selectivity displayed by neurons in higher cortical areas (Rigotti et al., 2013). It first marginalizes the response into independent parts that takes into account the time-varying, stimulus-dependent, decision-dependent components and their interaction. This is equivalent to the linear model used in the analysis of variance (ANOVA) test. The dimensionality reduction is performed on this marginalization with the constraint that the variance in each direction to be from only one component of the marginalization. The orthogonality constraint in PCA is relaxed in demixed PCA to allow for the demixing.

In general, identifying what the components derived from PCA represent in terms of spike counts can be difficult because they consist of both positive and negative values, whereas neural spike counts contain only non-negative values. This is more easily understood in relation to an illustration from (Lee and Seung, 1999), shown in Figure 1.7B, where PCA is applied to a dataset containing facial images. The derived module images (eigen images) contain both positive and negative values which makes it difficult to understand that they represents in terms of the visible features in the original dataset.

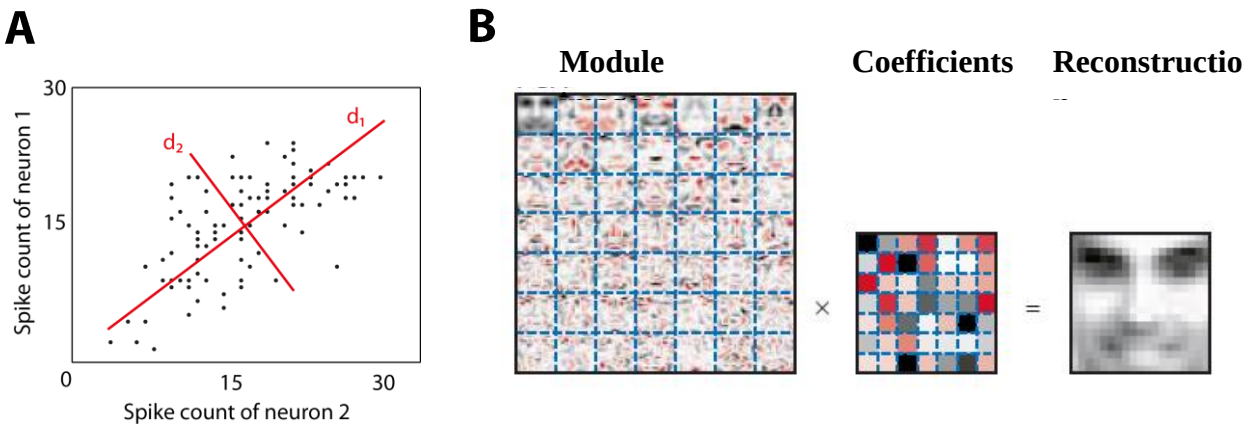

Figure 1.7: Principal component analysis (PCA). A: Illustration of PCA on spike counts recorded from two auditory cortical neurons. PCA finds the direction that captures the highest variance in the spike counts d_1 and the direction orthogonal to it d_2. B: PCA applied to a database containing facial images extracts a set of orthonormal module images (eigen images) that best capture the variance in image pixels and a set of coefficients that indicate the contribution of each module to reconstruct an image as a linear combination of the module images. (Adapted with permission from (Lee and Seung, 1999))

Independent component analysis (ICA)

Independent component analysis (ICA) is used to find independent sources of activity from a mixture of activities from different sources when the sources are non-Gaussian (Comon, 1994; Hyvärinen and Oja, 2000; Brown et al., 2001). Typically, this is represented as,

$$\mathbf{s} = \mathbf{W}\mathbf{x} \tag{1.5}$$

where the matrix $\mathbf{s}$ contains the independent sources to be estimated, the matrix $\mathbf{x}$ contains the observed data values and $\mathbf{W}$ is the demixing matrix.

Typically several preprocessing steps such as centering and whitening are performed prior to applying ICA on the data (Hyvärinen and Oja, 2000). Whitening is a linear transformation on data to derive a matrix $\tilde{\mathbf{x}}$ which contains uncorrelated components that have unit variance. Thus PCA can be used as a whitening method and ICA can be considered as an extension of PCA. Both PCA and ICA examine the dependency structure in the data. While PCA transforms data to remove the second order dependencies and decorrelates the data, ICA removes dependencies of all orders. In practice, when the sources are not strictly independent, ICA finds sources that are as independent as possible. It is based on the central limit theorem which says that the sum of a set of independent random variables will be more Gaussian than the variables themselves. ICA starts with a random vector $\mathbf{W}$ and rotates its axes to minimize a measure that specifies the Gaussianity such as mutual information, negentropy, kurtosis or maximum likelihood (Hyvärinen and Oja, 2000; Lopes-dos-Santos et al., 2013).

ICA has been used to identify neural ensembles with correlated firing (Laubach et al., 1999a; Laubach et al., 2000), analyze local field potentials (Makarov et al., 2010) and mass-scale neural signals (Makeig et al., 1997; Demirci et al., 2009; LeVan et al., 2010). In (Laubach et al., 1999a), where correlated firing was generated from common input, ICA was compared with PCA and was found to have a higher performance compared to PCA for identifying correlated neural ensembles. By observing that simple cells in the visual cortex that have distributed sparse representations of natural scenes using line and edge selectivities similar to independent components derived when applying ICA on natural images (Bell and Sejnowski, 1997; Vinje and Gallant, 2000), (Bell and Sejnowski, 1997) proposed that these cells could perform ICA on natural image information they receive from the retina. Recent modeling studies (Clopath et al., 2009; Savin et al., 2010) have implemented ICA based neuron models using spike timing dependent plasticity to separate correlated sources.

Factor analysis (FA)

FA assumes that the noise in the spike counts is Gaussian distributed and extracts a low-dimensional representation that models the variance in the spike counts that is shared across neurons. Simultaneously, the variance in the spike counts that is independent across neurons is identified and removed (Roweis and Ghahramani, 1999; Santhanam et al., 2009; Cunningham and Byron, 2014). Thus the method models the trial-to-trial variability in neural spike counts. More specifically, the probability distribution of the observed spike counts $\mathbf{y}$ of n neurons for a given stimulus s is modeled using k Gaussian distributed independent factors (latent components) with zero mean and unit variance $\mathbf{x}$, i.e. $p(\mathbf{x}) = N(\mathbf{x}; \mathbf{0}, \mathbf{I})$, such that (Santhanam et al., 2009),

$$p(\mathbf{y} \mid \mathbf{x}, s) = N(\mathbf{y}; \boldsymbol{\mu}_s + \mathbf{C}_s \mathbf{x}, \mathbf{R}_s) \tag{1.6}$$

and the likelihood of the spike counts for the stimulus s is,

$$p(\mathbf{y} \mid s) = N(\mathbf{y}; \boldsymbol{\mu}_s, \mathbf{C}_s \mathbf{C}_s^{'} + \mathbf{R}_s) \tag{1.7}$$

The latent variables $\mathbf{x}$ are considered to define the state of the population. $\boldsymbol{\mu}_s$ is the mean number of spikes generated for the stimulus s. $\mathbf{C}_s$ is the $n \times k$ matrix that maps the observations to the latent components. $\mathbf{R}_s$ is a diagonal matrix that captures the trial-to-trial variability in spike counts that is independent across neurons. This noise is attributed to the biophysical spiking noise and other non-shared sources of variability that is private to each neuron. The covariance in $\mathbf{y}$ for the stimulus s is $\mathbf{C}_s \mathbf{C}_s^{'} + \mathbf{R}_s$. $\mathbf{C}_s \mathbf{C}_s^{'}$ is the shared variability across the population of neurons. This shared variability is considered to be due to the firing rate variability between neurons. FA is closely related to sensible PCA (Roweis, 1998) or probabilistic PCA (Tipping and Bishop, 1999). The difference between the two methods is that FA allows the independent noise in $\mathbf{R}_s$ to vary between neurons while the PCA approaches constrain it to be the same for all neurons (Santhanam et al., 2009). In general, FA works best when using bin sizes > 150 ms (Santhanam et al., 2009). Both Gaussian and Poisson noise models have been evaluated for spike count data and the Gaussian noise model achieved higher performance compared to the Poisson model (Santhanam et al., 2009).

Gaussian process factor analysis (GPFA, (Byron et al., 2009; Yu et al., 2009)) is an extension of FA that is essentially a collection of FA models applied as one FA model for each time point in a trial. Then the neural states $\mathbf{x}_{:,t}$ at different time points t are related through Gaussian processes. The Gaussianity gives rise to smooth trajectories of activity in time. The i^{th} neural state $\mathbf{x}_{i,:}$ is modeled as $p(\mathbf{x}_{i,:}) = N(\mathbf{x}_{:,t}; \mathbf{0}, K_i)$, where the $T \times T$ matrix K_i defines the covariance of the states between different time points. K_i is typically modeled as a squared exponential function given by,

$$K_i\left(t_1,t_2\right)=\sigma_{f,i}^2\exp\left(-\frac{\left(t_1-t_2\right)^2}{2\tau_i^2}\right)+\sigma_{n,i}^2.\delta_{t_1,t_2} \tag{1.8}$$

The covariance of $K_i\left(t_1,t_2\right)$ is defined from its signal variance $\sigma_{f,i}^2\in\mathbb{R}_+$, characteristic time scale $\tau_i\in\mathbb{R}_+$ and the noise variance in the Gaussian process $\sigma_{n,i}^2\in\mathbb{R}_+$. The prior distribution of $\mathbf{x}_{:,t}$ is set to $p\left(\mathbf{x}_{:,t}\right)=N\left(\mathbf{x}_{:,t};\mathbf{0},\mathbf{I}\right)$ and the temporal smoothness is enforced by selecting the noise variance in the Gaussian process to be small $\left(\sigma_{n,i}^2=10^{-3}\right)$. Visualization of the model is done after a post-processing step in which the columns of $\mathbf{C}_s$ are orthonomalized.

FA based models have been used to study low dimensional state trajectories in premotor and motor cortical neurons in monkeys and to model the shared firing rate variability of neurons in V1, premotor and motor cortical neurons in monkeys (Byron et al., 2009; Yu et al., 2009; Churchland et al., 2010; Afshar et al., 2011; Ecker et al., 2014; Sadtler et al., 2014).

Dynamic methods

Dynamical methods (Radons et al., 1994; Abeles et al., 1995; Seidemann et al., 1996; Byron et al., 2005) explicitly model the time progression of the neural activity using a low dimensional dynamical process that moves through a state space. State space contains all possible states that the linear dynamical model could take. These states are thought to be related to different factors. They could relate to the experiment as well as other internal brain states. States defined by the experiment will depend on the details of each trial. For example, each trial could consists of a fixation period after which a cue is presented that is followed by a preparatory period. At the end of the preparatory period, a signal is given to make a movement and then the movement is made. Thus the activity of the recorded population would change during the trial depending on whether the current time is in the fixation period, preparatory period or the period of the movement (Byron et al., 2009). Furthermore, the neural activity would also change other internal states that are related to attention, fatigue and decreased motivation (Lawhern et al., 2010). Dynamical models aim to model this using a collection of hidden states. As the trial progresses the neural activity moves from one state to another. Depending on the internal states, the progression of the states could vary from one trial to another. Thus such models inherently include non-stationarity in the trial-to-trial activity. Dynamical models that analyze neural population activity in this manner include hidden Markov models that typically model transitions between states to be discrete events, linear and non-linear continuous dynamical models (Yu et al., 2009; Cunningham and Byron, 2014). These models are data driven unsupervised data analysis methods. Many of them contain explicit noise models and are used for single trial analysis. However, in general, they have been used at relatively long time scales of tens to hundreds of milliseconds.

Hidden Markov models (HMM)

These models define one possible way in which a dynamical system could evolve across a set of unobserved (hidden) states in time, where the subsequent state that the system could take depends only on the that the system is currently in (this memorylessness defines the Markovian property). According to these models, the spiking activity of N neurons at any time instance t, $\mathbf{y}_t$, is in one of K discrete hidden states $\mathbf{x}_t$ and makes a transition from one state to another according to (Escola et al., 2011),

$$p\left(\mathbf{x}_t \mid \mathbf{x}_{[0:t-1]}, \mathbf{y}_{[0:t-1]}\right) = p\left(\mathbf{x}_t \mid \mathbf{x}_{t-1}\right) \tag{1.9}$$

Thus, the future state is dependent only on the current state and is independent of the activity in the current point in time and the history of the past states. This memorylessness is a Markovian property. Furthermore, the probability of the transition from state m to n, a_{mn} (Figure 1.8), is constant (homogenous) at each time point with $0 \le a_{mn} \le 1$ and $\sum_{n=1}^{k} a_{mn} = 1$. The state transition matrix $\mathbf{A}$ contains the elements a_{mn}. For the states to be persistent across time, the diagonal elements of $\mathbf{A}$ are set close to unity.

The probability of spiking, or the emission probability is modeled to be dependent only on the current state i.e. $p\left(\mathbf{y}_t \mid \mathbf{x}_{[0:t]}, \mathbf{y}_{[0:t-1]}\right) = p\left(\mathbf{y}_t \mid \mathbf{x}_t\right)$. Then, if the number of spikes range between 0 to C, the probability of observing c spikes in state $\mathbf{m}$, $\eta_{mc}, \eta_{mc} \in \boldsymbol{\eta}$ is given by, $\eta_{mc} \equiv p\left(\mathbf{y}_t = c \mid \mathbf{x}_t = \mathbf{m}\right)$, where $0 \le \eta_{mc} \le 1$ and $\sum_{n=1}^{C} \eta_{mc} = 1$. Typically, the spike counts of each neuron are assumed to be Poisson distributed (Abeles et al., 1995; Gat et al., 1997).

This gives the HMM defined by,

$$p\left(\mathbf{y}, \mathbf{x}\right) = p(\mathbf{x}_0) \prod_{t=1}^{T} p(\mathbf{x}_t \mid \mathbf{x}_{t-1}) \prod_{t=1}^{T} p(\mathbf{y}_t \mid \mathbf{x}_t) \tag{1.10}$$

where $\mathbf{x}_0$ is the initial state defined by the probability distribution $\boldsymbol{\pi}$ such that $\boldsymbol{\pi}_m \equiv p\left(\mathbf{x}_0 = \mathbf{m}\right)$. The parameters of the model, $\mathbf{A}$, $\boldsymbol{\eta}$ and $\boldsymbol{\pi}$ are found using maximum likelihood estimation. The likelihood of the data is typically estimated using the Baum-Welch algorithm (Baum et al., 1970). (Escola et al., 2011) extended the discrete HMMs defined using GLMs to continuous HMMs. HMMs have been applied for single trial analysis of neural responses in a range of studies including the frontal cortex in monkeys (Abeles et al., 1995; Seidemann et al., 1996; Ponce-Alvarez et al., 2012), V1 in monkeys (Radons et al., 1994), premotor cortex in monkeys (Kemere et al., 2008), parietal cortex in monkeys (Bollimunta et al., 2012), somatosensory cortex in monkeys (Ponce-Alvarez et al., 2012), associative cortex in monkeys (Gat et al., 1997), rat gustatory cortex (Jones et al., 2007; Escola et al., 2011; Sadacca et al., 2016), premotor neurons in songbirds (Danóczy and Hahnloser, 2006).

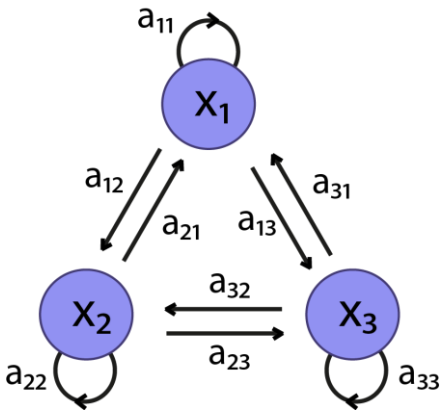

Figure 1.8: State transition diagram for a model with three hidden states. The state transition probabilities a_{mn} are specified by the state transition matrix $\mathbf{A}$.

Linear dynamical models

Kalman filter and autoregressive process are two linear dynamical systems. Many of these models are based on Gaussian noise models. We will now look at each of them briefly.

Kalman filter models

The Kalman filter is a linear dynamical model describing the neural activity in terms of the observed variables (such as hand movements in a motor task) and a set of hidden states using a Gaussian noise model (Wu et al., 2009; Wu and Liu, 2015). According to this model, the neural activity $\mathbf{y}_t$ at a time t of N neurons is related to k_o number of observed variables $\mathbf{x}_t$ and k_{uo} number of unobserved hidden states $\mathbf{n}_t$ at time t by the measurement equation

$$\mathbf{y}_t = \mathbf{H}\mathbf{x}_t + \mathbf{G}\mathbf{n}_t + \mathbf{v}_t \tag{1.11}$$

and the state transition is defined by the system equation

$$\begin{pmatrix} \mathbf{x}_{t+1} \\ \mathbf{n}_{t+1} \end{pmatrix} = \mathbf{A}\begin{pmatrix} \mathbf{x}_t \\ \mathbf{n}_t \end{pmatrix} + \mathbf{w}_t \tag{1.12}$$

The $N \times k_o$ matrix $\mathbf{H}$ and the $N \times k_{uo}$ matrix $\mathbf{G}$ are coefficient matrices that respectively relate the observed variables and the unobserved variables to the spike counts. $\mathbf{A}$ is the $(k_0 + k_{uo}) \times (k_0 + k_{uo})$ state transition matrix similar to that for HMMs. $\mathbf{v}_t$ and $\mathbf{w}_t$ are independent noise terms that are taken to be Gaussian distributed such that $p(\mathbf{v}_t) = N(\mathbf{0}, \mathbf{Q})$

and $p(\mathbf{w}_t) = N(\mathbf{0}, \mathbf{W})$ where the $N \times N$ matrix $\mathbf{Q}$ and the $(k_0 + k_{uo}) \times (k_0 + k_{uo})$ matrix $\mathbf{W}$ define the covariance of the noise. The initial state $\mathbf{n}_0$ is considered to be Gaussian distributed according to $p(\mathbf{n}_0) = N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ such that $\boldsymbol{\mu}$ is the k_{uo} dimensional vector specifying the mean of the initial state and $\boldsymbol{\Sigma}$ is the $k_{uo} \times k_{uo}$ dimensional covariance matrix. The parameters of the model $\mathbf{H}, \mathbf{G}, \mathbf{A}, \mathbf{W}$ and $\mathbf{Q}$ are estimated by fitting the model to the data using maximum likelihood estimation (Wu et al., 2009). Kalman filter based models have been used to study the neural activity of motor neurons with and without the inclusion of hidden states (Wu et al., 2004; Wu et al., 2006; Wu et al., 2009).

Autoregressive models

First-order autoregressive models are typically used to model linear dynamical processes (Smith and Brown, 2003; Kulkarni and Paninski, 2007; Lawhern et al., 2010; Buesing et al., 2012a, b). The initial hidden state $\mathbf{x}_0$ of a model that has k states is considered to be Gaussian distributed according to $p(\mathbf{x}_0) = N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ such that $\boldsymbol{\mu}$ is the k dimensional vector specifying the mean of the initial state and $\boldsymbol{\Sigma}$ is the $k \times k$ dimensional covariance matrix (Buesing et al., 2012a). The state transition is modeled by a Gaussian process

$$p(\mathbf{x}_{t+1} \mid \mathbf{x}_t) = N(\mathbf{A}\mathbf{x}_t, \mathbf{Q}) \tag{1.13}$$

where the $k \times k$ matrix $\mathbf{A}$ models the temporal dependence of the process and $\mathbf{Q}$ is the covariance of the state transitions. The hidden state is related to the neural activity through a q-dimensional variable $\mathbf{z}_t$ where $\mathbf{z}_t = \mathbf{C}\mathbf{x}_t + \mathbf{d}$. $\mathbf{C}$ is a loading matrix that relates $\mathbf{x}_t$ to $\mathbf{z}_t$ and $\mathbf{d}$ is the mean parameter. The neural activity is related to $\mathbf{z}_t$ through another Gaussian process where

$$p(\mathbf{y}_t \mid \mathbf{z}_t) = N(\mathbf{z}_t, \mathbf{R}) \tag{1.14}$$

$\mathbf{R}$ is the covariance of the observed neural activity at the state $\mathbf{z}_t$. With sufficient time $(t \to \infty)$, and when $|a_{mn}| < 1, a_{mn} \in \mathbf{A}$, the model achieves a stationary state (Yu et al., 2009). Autoregressive models have been used to incorporate hidden states into the GLMs in (Lawhern et al., 2010). They have been useful to analyze population activity in the motor cortex in monkeys (Lawhern et al., 2010; Buesing et al., 2012b) and premotor cortex in monkeys (Buesing et al., 2012b).

Summary

Advances in multi-electrode recording techniques and spike sorting methods currently allow reliable recordings from populations of hundreds of neurons simultaneously (Buzsaki, 2004; Buzsáki et al., 2015). Analyzing large ensembles is important because the brain has access to millions of neurons at the same time. Furthermore, neuroimaging methods such as fMRI allow the study of a variety of cognitive and pathological conditions noninvasively, which is often the only option when studying the human brain. However, these methods are based on large scale neuronal activity. Predicting the activity at the single cell level that gives arise to these large scale measures can only be accomplished if we understand how neurons encode information at the level of large populations.

Neurons can code information in spatial and temporal dimensions concurrently. Thus, ideally, we require methods to analyze their responses concurrently in space and time. There is a wide range of methods currently developed. However, many methods that require the evaluation of conditional probability distributions of the population responses to stimuli are constrained by the curse of dimensionality when extending to spatiotemporal analysis of large scale datasets. They often employ a range of dimensionality reduction methods to extract meaningful features from these datasets. While there is a range of dimensionality reduction methods that have been used successfully in neuroscience, it is not easy to interpret the extracted features in terms of the original data and not all methods have been used to extract information in short time scales of a few milliseconds. This indicates that there is still the need for methods that could derive representations that can give intuitively meaningful insights into understanding complex activity patterns.

1.6 Overview of the thesis

We summarized commonly used methods to study neural population coding. As we discussed, many methods are constrained by the curse of dimensionality when extending to the high dimensional spatiotemporal response space and require dimensionality reduction methods. This thesis adapts a new dimensionality reduction method, non-negative matrix factorization (NMF, (Lee and Seung, 1999)), to extract meaningful patterns of activity from the responses of large populations of neurons concurrently in space and time. NMF is a method that has been widely used in many areas within the machine learning community (Cichocki et al., 2009), but with very little application in neuroscience (Kim et al., 2005; Wei et al., 2015; Pnevmatikakis et al., 2016; Onken et al., In preparation). The thesis is organized as follows.

In chapter 2, we adapt two NMF methods, spatiotemporal NMF and space-by-time NMF, to analyze the spiking activity of large neural populations in space and time. We apply these methods on three large scale datasets from three sensory modalities; 1) a dataset recorded from rat auditory cortex (A1) to long tones and clicks (kindly provided by Shuzo Sakata and collaborators at Strathclyde University), 2) a dataset recorded from salamander retinal ganglion cells to static images and movies (kindly provided by Tim Gollisch and colleagues

from University of Gottingen) and 3) a dataset recorded from rat barrel cortex to whisker deflections (Petersen et al., 2001). We explain in detail how our new method represents these datasets, its capability to discriminate stimuli and report findings on the spatiotemporal neural codes used by the populations.

In chapter 3, we extend space-by-time NMF to model sub-Poisson, Poisson and supra-Poisson variability in neural data, thus optimizing the model performance even further. We validate our new algorithms rigorously using statistical simulations, network simulations and applying them on our somatosensory and auditory datasets. We report the results from validation of the new update rules and detail further insight we gained in using space-by-time NMF to analyze cortical spike trains.

Finally, we conclude the thesis in chapter 4 with a discussion that summarizes and discusses the main findings from our work and elaborate on implications that our new method has on analyzing large scale neural datasets.

Chapter 2: Non-negative matrix factorization to extract spatiotemporal spike patterns

2.1 Abstract

Understanding the neural population code requires identification of salient and informative patterns of activity from large scale recordings. In this chapter we explore the feasibility of using non-negative matrix factorization (NMF) to extract informative spatiotemporal patterns of activity from recordings of spike trains from neural populations. We investigate two variations of NMF. The first, spatiotemporal NMF that identifies recurrent spatiotemporal spike patterns has been used in a few studies to study neural activity (Kim et al., 2005; Overduin et al., 2015; Wei et al., 2015). We introduce a new NMF based method space-by-time NMF (Delis et al., 2014) to study population activity.

This chapter is structured as follows. It begins with a short introduction to the importance of spatial and temporal dimensions for population coding. Next, we present a short introduction to NMF and describe the two methodologies that we use to analyze population spike trains. This is followed by a comprehensive analysis performed over three sensory datasets comprising of auditory, touch and visual modalities. We conclude the chapter with a discussion of the insight we gained through our methods and implications of using NMF to study neural activity.

2.2 Introduction

Neural populations from sensory areas (Stopfer et al., 2003; Jones et al., 2007), motor areas (Churchland et al., 2012) to higher level regions (Durstewitz et al., 2010; Mante et al., 2013) encode sensory stimuli or task-related variables using neuronal population responses with complex spatial and temporal dynamics, such as trajectories of population activity evolving across time. Understanding how these sensory, motor or task-related variables are encoded in neural population activity, that is understanding the neural population code, is a prerequisite to understand any brain function. It is important, for example, to understand what parts of neural population responses should be plotted and further analyzed in neurocognitive studies, or to understand the aspects of neural activity that must be crucially be well explained by computational models aiming at describing specific brain functions.

As we discussed in detail in section 1.3, a neural population can encode information through activity patterns across spatial and temporal dimensions. Here, the spatial dimension refers to encoding information across groups of neurons by exhibiting differential selectivity between groups to different attributes. Such subsets of neurons are often locally organized in specific brain region (Hubel and Wiesel, 1962; Petersen, 2007; Schreiner et al., 2011) but concurrent coordination of neuronal groups across multiple brain regions are also observed for higher level functions (Hoffman and McNaughton, 2002). The temporal dimension involves

modulation of spiking activity across time (Petersen et al., 2001; Panzeri et al., 2010b; Onken et al., 2014). Spatiotemporal dynamics typically observed in populations require concurrent analysis of both dimensions. As we illustrated in section 1.5, the number of parameters that has to be estimated for a given analysis method increases exponentially when the dimensionality of the response space, as specified by the number of neurons and the duration to be analyzed, increases, but is constrained by the limited amount of trials available to estimate them (Onken et al., 2014). One common approach to overcome this curse of dimensionality is to use a dimension reduction technique (Laubach et al., 1999b; Stopfer et al., 2003; Byron et al., 2009; Churchland et al., 2012; Mazzoni et al., 2013; Zuo et al., 2015).

Dimensionality reduction methods (section 1.5.5, (Cunningham and Byron, 2014)) are mathematical procedures that describe a dataset with the smallest number of parameters and at minimal information loss. Specifically, they are designed to optimize an objective specific to each method and extract a reduced number of latent or hidden components that offer a certain viewpoint of the dataset based on the objective that the particular method optimizes. Dimension reduction methods such as principal component analysis (PCA), independent component analysis (ICA) and factor analysis (FA) have been very useful to study neural coding, the patterns they extracts from a dataset are often difficult to interpret in terms of spike counts in the original dataset. This is mainly because the patterns contain negative values which could be difficult to relate to the spike counts in the dataset.

We investigated the feasibility of using another dimension reduction method, non-negative matrix factorization (NMF), to identify salient spatial and temporal patterns from the spiking activity in large populations of neurons. NMF (Lee and Seung, 1999) is a method used in the machine learning community (Cichocki et al., 2009) to study large-scale datasets such as document and email corpuses (Xu et al., 2003; Berry and Browne, 2005) and microarray data (Brunet et al., 2004; Carmona-Saez et al., 2006). However, it is only beginning to be used to analyze neural data (Kim et al., 2005; Wei et al., 2015; Pnevmatikakis et al., 2016; Onken et al., In preparation) and limited detail is available on the low-dimensional representation with which it can represent spike data.

We conducted a comprehensive study about what type of spatial and temporal patterns NMF could extract concurrently from large neural datasets and how well the extracted patterns describe stimulus related attributes under a decoding framework. We considered two possible variants of NMF that could be used to study population spike trains. Keeping to the terminology from (Delis et al., 2014), we refer to these methods as spatiotemporal NMF and space-by-time NMF. We used them in three neural datasets; 1) a dataset recorded from rat auditory cortex (A1) to long tones and clicks (kindly provided by Shuzo Sakata and collaborators at Strathclyde University), 2) a dataset recorded from salamander retinal ganglion cells to static images and movies (kindly provided by Tim Gollisch and colleagues from University of Gottingen) and 3) a dataset recorded from rat barrel cortex to whisker deflections (Petersen et al., 2001). The first two datasets consists of simultaneous multielectrode recordings while the latter dataset is a pooling of recordings over multiple sessions. These provide an understanding of the low-dimensional representation used by NMF to represent large-scale spike trains.

This chapter is organized as follows. It starts with a brief general introduction to NMF. Then the two methods we use to analyze data, spatiotemporal NMF and space-by-time NMF are described in detail. This is followed by findings from the three datasets using NMF to identify concurrent low-dimensional spatial and temporal patterns and its ability to discriminate stimuli in comparison to using information from only one dimension. The discussion at the end summarizes and elaborates our findings and lays out the reasons for the series of studies we conducted to improve the method to analyze neural population responses.

2.3 Non-negative matrix factorization

NMF (Lee and Seung, 1999) is a data-driven dimension reduction method that can be applied on large scale datasets containing only non-negative elements. It can extract patterns that are reliably present in the dataset at a global scale. Typically, the full dataset is arranged into a matrix. NMF can decompose this matrix into one or more matrices that contain patterns which are representative of the datasets (we refer to these matrices as module matrices) and one matrix that contains a set of weights with which each element in the original dataset could be reconstructed as a linearly weighted sum of the extracted patterns (we refer to this matrix as the coefficient matrix). The factorization process is performed iteratively where in each iteration, the factorized matrices are updated using a set of update rules designed to minimize a dissimilarity measure that quantifies the error between the original dataset and the linearly reconstructed dataset. The update rules are constrained such that the factorized matrices remain non-negative throughout the updating process. This non-negativity constraint together with the linear reconstruction give rise to identification of patterns that are sparse and part-based (Lee and Seung, 1999). They are often found to have clear and intuitive meaning with respect to the original data in many areas that have used NMF as a dimension reduction method. A few representative data types include facial images (Lee and Seung, 1999; Guillamet and Vitria, 2002), document and email corpuses (Xu et al., 2003; Berry and Browne, 2005), microarray data (Brunet et al., 2004; Carmona-Saez et al., 2006), electromyographic activity (d'Avella et al., 2003; Delis et al., 2013; Delis et al., 2014) and music (Smaragdis and Brown, 2003; Févotte et al., 2009).

NMF in its basic two-factor decomposition (Lee and Seung, 1999) factorizes a $N \times T$ matrix $\mathbf{R}$ into a $N \times K$ module matrix $\mathbf{W}$ and a $K \times T$ coefficient matrix $\mathbf{H}$ where,

$$\mathbf{R} \approx \mathbf{WH} \tag{2.1}$$

and $K \ll \min(T, N)$. The module matrix $\mathbf{W}$ consists of the low-dimensional patterns extracted from the data. These patterns are described based on the full dataset and are item-independent. The coefficient matrix $\mathbf{H}$ specifies the contribution of each module necessary to make a linear reconstruction of an item in the original dataset. Thus, unlike the module matrix, the coefficient matrix is item dependent and scales in size according to the number of items in the dataset.

In the example shown in Figure 2.1, NMF was applied on a facial image dataset (Lee and Seung, 1999). The data matrix $\mathbf{R}$ was constructed such that each column in $\mathbf{R}$ contained the intensity of the vectorized pixels in one image. Each module extracted into $\mathbf{W}$ matrix is an image consisting of a global facial feature in the dataset such as eyes, nose or lips. The coefficients in the $\mathbf{H}$ matrix indicate the level of the presence of each facial feature within each image. The reconstruction of an image in the dataset $\mathbf{r}_i$ is obtained as $\mathbf{r}_i \approx \sum_{k=1}^{K} h_k \mathbf{W}$.

(Ding et al., 2006) extended the two-factor NMF to a tri-factor decomposition through

$$\mathbf{R} \approx \mathbf{W}_1 \mathbf{H} \mathbf{W}_2 \tag{2.2}$$

with $\mathbf{W}_1^T \mathbf{W}_1 = \mathbf{I}$ and $\mathbf{W}_2^T \mathbf{W}_2 = \mathbf{I}$. The $N \times K_1$ matrix $\mathbf{W}_1$ clusters pixels into facial features based on pixel co-occurrences in different images and $K_2 \times T$ matrix $\mathbf{W}_2$ clusters images into groups based on the presence of similar pixels in the images. A common use of tri-factorization is for text mining (Ding et al., 2006; Wang et al., 2011). In this case, $\mathbf{R}$ is a term-document matrix built using the frequency of words in a document corpus with words as its rows and documents as its columns. Then, $\mathbf{W}_1$ clusters words into topics based on word co-occurrence across documents and $\mathbf{W}_2$ clusters documents into different groups based on the presence of similar words in the documents. The ability of the three-factor decomposition to co-cluster features from both dimensions of the data matrix into separate matrices makes it a natural choice for concurrent identification of spatial and temporal activity patterns from a data matrix comprising of spatiotemporal population responses.

2.4 NMF for spatiotemporal spike activity

When applying NMF for spike data from N neurons, we first bin each spike train of s $(1 \le s \le S)$ trials into non-overlapping T timebins creating a $N \times Ts$ response matrix $\mathbf{R}$.

We applied the following two variants of NMF to explore population spike counts concurrently in space and time.

1. Spatiotemporal NMF

2. Space-by-time NMF

In both cases, NMF models the population response of each trial using a linear combination of trial-invariant modules.

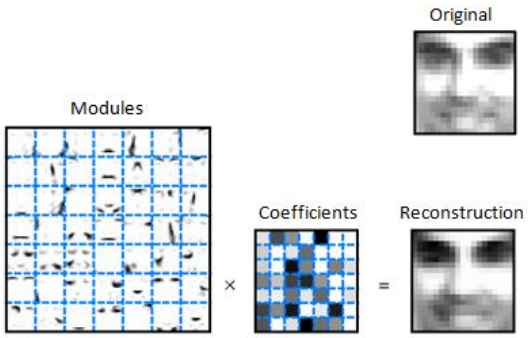

Figure 2.1: Illustration of two-factor NMF applied on a facial image dataset and the reconstruction of an image using underlying modules. The dataset consisted of 2429 facial images of 19 x 19 pixels each. (adapted with permission from (Lee and Seung, 1999))

Spatiotemporal NMF

The spatiotemporal NMF decomposition (d Avella and Tresch, 2002), shown in Figure 2.2A is obtained by first arranging the time bins of all neurons in one trial as columns of matrix **R** and then using equation (2.1) to decompose **R**. The resulting $N \times T$ module matrix **W** contains spatiotemporal modules as its columns. These are then reshaped appropriately to obtain K spatiotemporal modules. These spatiotemporal modules are trial-invariant. Each module describes a time-varying pattern of activity in the population. Trial-to-trial activity is described by the $T \times S$ coefficient matrix **H** containing a weight for each spatiotemporal module indicating the contribution of the particular activity pattern in the module to the population response of one trial. In this low-dimensional representation, the dimensionality of each trial is reduced from NT parameters to K parameters.

Space-by-time NMF

This algorithm is a variant of the tri-factor NMF model that (Delis et al., 2014) initially proposed as sample-based non-negative matrix tri-factorization (sNM3F) to analyze electromyography data (Delis et al., 2014). It concurrently decomposes the spatial and temporal dimensions into two trial-invariant matrices; a $L \times N$ spatial module matrix and a $T \times P$ temporal module matrix as shown in Figure 2.2B using,

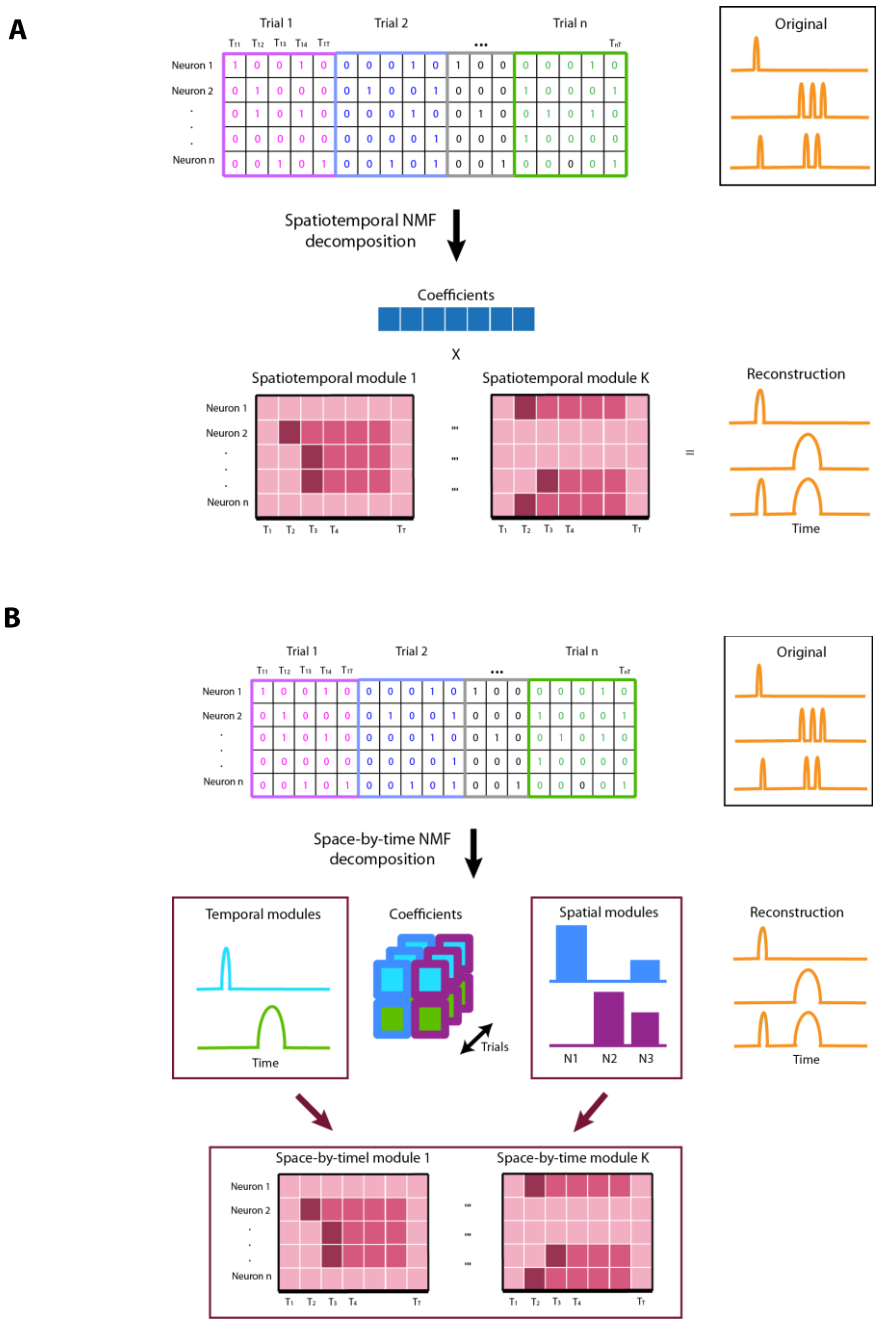

Figure 2.2: Illustration of spatiotemporal NMF and space-by-time NMF models. A: Spatiotemporal NMF model. Spatiotemporal NMF decomposes the dataset into a set of trial-invariant spatiotemporal modules and a trial-variant coefficient matrix. Each coefficient describes the activation level of the respective spatiotemporal module in a trial. B: Space-by-time NMF model. Space-by-time NMF model decomposes the data matrix into a trial invariant spatial module matrix and a temporal module matrix together with a trial-variant coefficient matrix. The spatial module matrix consists of groups of neurons that tend to fire together. The temporal module matrix consists of temporal activation patterns of neurons. Each coefficient of the coefficient matrix indicates the level of activation of the corresponding spatial and temporal module combination (i.e. one space-by-time module) in a trial.

$$\mathbf{r}^s(t) \approx \sum_{i=1}^{P}\sum_{j=1}^{L} \mathbf{w}_{tem,i}(t) h_{i,j}^s \mathbf{w}_{spa,j} \qquad (2.3)$$

where $\mathbf{w}_{spa,j}$ is the j^{th} row of the spatial module matrix $\mathbf{W}_{spa}$ and $\mathbf{w}_{tem,i}$ is the i^{th} column of the temporal module matrix $\mathbf{W}_{tem}$. The activity of a single-trial is specified by the $P \times L$ coefficient matrix $\mathbf{h}^s$. The value $h_{i,j}^s$ in the coefficient matrix indicates the activation level of the i^{th} temporal module and j^{th} spatial module in trial s. In the space-by-time representation, the dimensionality of each trial is reduced from NT to PL parameters.

The complete space-by-time NMF algorithm is given in Algorithm 1.

Algorithm 1: Space-by-time NMF algorithm

Let $\mathbf{R} = \begin{pmatrix} \mathbf{r}^1 \\ \mathbf{r}^2 \\ \vdots \\ \mathbf{r}^s \end{pmatrix}$. Define $\mathbf{R}' = \left(\mathbf{r}^1 \mathbf{r}^2 ... \mathbf{r}^S\right)$ to be the block transpose of $\mathbf{R}$, where $\left(\mathbf{R}'\right)' = \mathbf{R}$

Initialize $\mathbf{W}_{tem}^1 > \mathbf{0}$, $\left(\mathbf{H}'\right)^1 = \left[\left(\mathbf{h}^1\right)^1 \left(\mathbf{h}^2\right)^1 \ldots \left(\mathbf{h}^S\right)^1\right] > \mathbf{0}$, and $\mathbf{W}_{spa}^1 > 0$.

Iterate

1. Given $\mathbf{H}'$ and $\mathbf{W}_{tem}^k$, update $\mathbf{W}_{spa}^k$:

   a. Formulate $\left(\mathbf{G}'\right)^k = \mathbf{W}_{tem}^k \left(\mathbf{H}'\right)^k$

   b. Obtain the block transpose $\mathbf{G}^k$

   c. Update $\mathbf{W}_{spa}^k$ to improve the approximation $\mathbf{R}^k \approx \mathbf{G}^k \mathbf{W}_{spa}^k$

$$\mathbf{W}_{spa,i,j}^{k+1} \leftarrow \mathbf{W}_{spa,i,j}^k \frac{\left(\left(\mathbf{G}^k\right)^T \mathbf{R}\right)_{i,j}}{\left(\left(\mathbf{G}^k\right)^T \mathbf{G}^k \mathbf{W}_{spa}^k\right)_{i,j}}, \forall i \in [1 \ldots L], j \in [1 \ldots N]$$

2. Given $\mathbf{H}'$ and $\mathbf{W}_{spa}^{k+1}$, update $\mathbf{W}_{tem}^k$:

   a. Get the block transpose $\mathbf{H}^k$

   b. Calculate $\mathbf{P}^k = \mathbf{H}^k \mathbf{W}_{spa}^{k+1}$

c. Obtain the block transpose $\left(\mathbf{P}^k\right)'$

d. Update $\mathbf{W}_{tem}^k$ to improve the approximation $\mathbf{R}' \approx \mathbf{W}_{tem}^k \left(\mathbf{P}^k\right)'$

$$\mathbf{W}_{tem,i,j}^{k+1} \leftarrow \mathbf{W}_{tem,i,j}^k \frac{\left(\mathbf{R}'\mathbf{P}^k\right)_{i,j}}{\left(\mathbf{W}_{tem}^k \left(\mathbf{P}^k\right)' \mathbf{P}^k\right)_{i,j}}, \forall i \in [1 \ldots T], j \in [1 \ldots P]$$

3. Given $\mathbf{W}_{spa}^{k+1}$ and $\mathbf{W}_{tem}^{k+1}$, update $\mathbf{H}$:

Update $\left(\mathbf{H}^s\right)^k, \forall s \in [1 \ldots S]$ to improve the approximation $\mathbf{R}^s \approx \mathbf{W}_{tem}^{k+1} \left(\mathbf{H}^s\right)^k \mathbf{W}_{spa}^{k+1}$

$$\left(\mathbf{H}_{i,j}^s\right)^{k+1} \leftarrow \left(\mathbf{H}_{i,j}^s\right)^k \frac{\left(\left(\mathbf{W}_{tem}^{k+1}\right)^T \mathbf{R}^s \left(\mathbf{W}_{spa}^{k+1}\right)^T\right)_{i,j}}{\left(\left(\mathbf{W}_{tem}^{k+1}\right)^T \mathbf{W}_{tem}^{k+1} \left(\mathbf{H}^s\right)^k \mathbf{W}_{spa}^{k+1} \left(\mathbf{W}_{spa}^{k+1}\right)^T\right)_{i,j}} \forall i \in [1 \ldots P], j \in [1 \ldots L]$$

4. Estimate the dissimilarity $\mathbf{E}^2$ between dataset $\mathbf{R}$ and the total reconstruction $\sum_s \mathbf{W}_{tem}^{k+1} \left(\mathbf{H}^s\right)^{k+1} \mathbf{W}_{spa}^{k+1}$ using Euclidean distance,

$$\left(\mathbf{E}^2\right)^{k+1} = \sum_s \left(\left(\mathbf{E}^2\right)^s\right)^{k+1} = \sum_s \left\| \mathbf{R}^s - \mathbf{W}_{tem}^{k+1} \left(\mathbf{H}^s\right)^{k+1} \mathbf{W}_{spa}^{k+1} \right\|^2$$

until $\left(E^2\right)^{k+1} - \left(E^2\right)^k < \left(E^2\right)_{tolarence}$ or $k+1 >$ max number of iterations

---

2.5 Representation of spatiotemporal neural responses using NMF

We applied NMF to three large scale neural datasets recorded from neurons coding three sensory modalities; a dataset recorded from rat auditory cortex to long tones and clicks, a dataset recorded from salamander retinal ganglion cells to static images and movies (Onken et al., In preparation) and a dataset recorded from rat barrel cortex to whisker reflections (Petersen et al., 2001). The first two datasets consisted of simultaneous multi-electrode recordings while the third dataset contained a collection of records from multiple experimental sessions.

2.5.1 Using NMF to study population coding of sounds in rat auditory cortex

Experimental details

The dataset was recorded from the auditory (A1) cortex of an anesthetized adult Sprague-Dawley rat. Experiments were performed in accordance with the United Kingdom Animals (Scientific Procedures) Act of 1986 and were approved by the United Kingdom Home Office and the Ethical Committee of Strathclyde University. The experimental procedures were similar to that described in (Sakata and Harris, 2009; Kayser et al., 2015). The animal was anesthetized with 1.5 g/kg urethane. Lidocaine (2%, 0.1–0.3 mg) was administered subcutaneously near the site of incision. A feedback temperature controller was used to maintain the body temperature at 37°C. To facilitate acoustic stimulation, a head post was fixed to the frontal bone using bone screws, and the animal was placed in a custom head restraint that left the ears unobstructed. After reflecting the left temporalis muscle, the bone over the left A1 was removed and a small duratomy was performed. Simultaneous activity of neurons was recorded with a 8 channel ‘‘silicon probe’’ (NeuroNexus Technologies). The brain was covered with 1% agar/0.1 M PBS to reduce pulsation and to keep the cortical surface moisturized during recording. There was a waiting period of ∼30 min prior to the start of recordings. Broadband signals (0.07 Hz to 8 kHz) were amplified (x 1000) using a Plexon system (HST/32V-G20 and PBX3) relative to a cerebellar bone screw and were digitized at 20 kHz (PXI; National Instruments).

Acoustic stimuli were generated through a multi-function data acquisition board (NI-PCI-6221; National Instruments) and were presented at a sampling rate of 96 kHz using a speaker driver (ED1; Tucker-Davis Technologies) and using free-field speakers (ES1; Tucker-Davis Technologies) located ∼10 cm in front of the animal. Sound presentation was calibrated using a pressure microphone (PS9200KIT-1/4; ACO Pacific) to ensure linear transfer and calibrated intensity. Tone stimuli consisted of nine long pure tones (500 ms long with 10 ms cosine ramps, 1/2 octave steps, 2 - 32 kHz, 30 and 60 dB SPL). Click stimuli comprised of 1 s long repetitive click sequences (5 ms broadband noise with 1 ms cosine ramps at 70 dB SPL) of frequencies 4, 8, 16, 32 and 64 Hz. Each stimulus was repeated over 100 trials. Spike detection and sorting was done offline using freely available software EToS version 3

(http://etos.sourceforge.net) and Klustakwik (http://klustakwik.sourceforge.net). Only cells with isolation distances $\geq$ 20 were used in the analysis. This gave 85 well-isolated cells.

Data analysis

We binned spikes of each cell into 5 ms timebins and created a 85 x 180,000 spatiotemporal matrix of spike counts in which all trials were concatenated sequentially. We then applied space-by-time NMF and spatiotemporal NMF. We identified the ability of the extracted modules to describe physiological data. We evaluated the ability of the coefficient matrix derived by NMF to describe trial-by-trial activity through a cross validated decoding procedure. The performance of the decoder indicates the concurrent contribution of space-and-time dimensions towards identifying stimuli. To study the contribution of the spatial dimension and the temporal dimension separately to the decoding performance, we used a permutation procedure (Figure 2.6A). To identify the decoding performance when only spatial dimension carry information, we randomly permuted the spikes of each trial while keeping the spatial structure intact in each time bin prior to applying NMF. Only stimulus selective modulation of the firing rates of the neurons remained after processing the data in this manner thus consisting only information available in the spatial dimension. To obtain the decoding performance when using only the temporal dimension, we permuted the spike trains across neurons in each trial while keeping the order of spikes in each train intact. Only the temporal ordering of spikes remain in the data after processing the data in the manner thus containing only temporal information..

Decoding

We applied the following decoding procedure. First we generated twenty trial-randomized datasets and divided each dataset into an equal sized training set and test set. Then we extracted the underlying temporal and spatial module structure in the training dataset together with the activation coefficients for each trial. We then fixed the derived temporal and spatial modules of the training set and updated only the coefficient matrix for the test set. We used the estimated training set coefficients to train a linear discriminant classifier and used it to decode the test set trials. We identified the optimal decomposition to be the one which gives the best test set decoding performance with the least number of modules.

Signal and noise correlations

We calculated the signal correlation and noise correlation between pairs of neurons grouped into spatial modules to better understand the composition of spatial modules in terms of measures used to describe neural variability. We assigned neurons to spatial modules using a threshold on the weights of the neurons in a spatial module. Then we identified neurons pairs that shared one or more spatial modules. We call these pairs of neurons "within module

pairs". Then we randomly sampled an equal number of pairs of neurons from all pairs of neurons in the dataset. We call these pairs of neurons "across module pairs". We computed signal and noise correlations for both neuron groups using the total spike counts in each trial. The threshold used to assign neurons to modules is an unknown value. We chose the threshold to be 0.4 times the standard deviation of the weights in each module. This threshold classified 50% and 37% of all pairs of neurons as within module pairs for responses to click sequences and long tones respectively.

Space-by-time NMF module structure

We first examine the module structure we obtained for the responses of the 85 neurons when stimulated with click sequences. Two example population responses to 4 Hz click sequence are shown in Figure 2.3A. The population response is inhomogeneous across repeating clicks in the sequence during a single trial and also has a large degree of trial-to-trial variability in which periods of population activity fluctuates between periods of silence. As described in section 1.2, similar fluctuations of activity is observed when the cortex is in a synchronized state under anesthesia (Petersen et al., 2003; Curto et al., 2009; Luczak et al., 2009; Pachitariu et al., 2015). During this synchronized state, which can last for long time durations in the order of seconds, periods of concerted population activity, in the order of hundreds of milliseconds, appear between time periods of population silence (Curto et al., 2009). The activated periods are called up-states while the inactivated periods are referred to as down-states. Similar up-down activity states are found in urathene-anesthetized rats when stimulated with click trains (Curto et al., 2009). Even though there is a large variability in individual trials, the trial-averaged population responses in Figure 2.3A show clear stimulus modulation. There is a high degree of synchronized activity in the population for the onset response to each click in low frequency click trains. Population onset responses to repeating clicks diminish as the frequency of the click train increases, while the onset response to the first click in the sequence remains stable across frequencies.

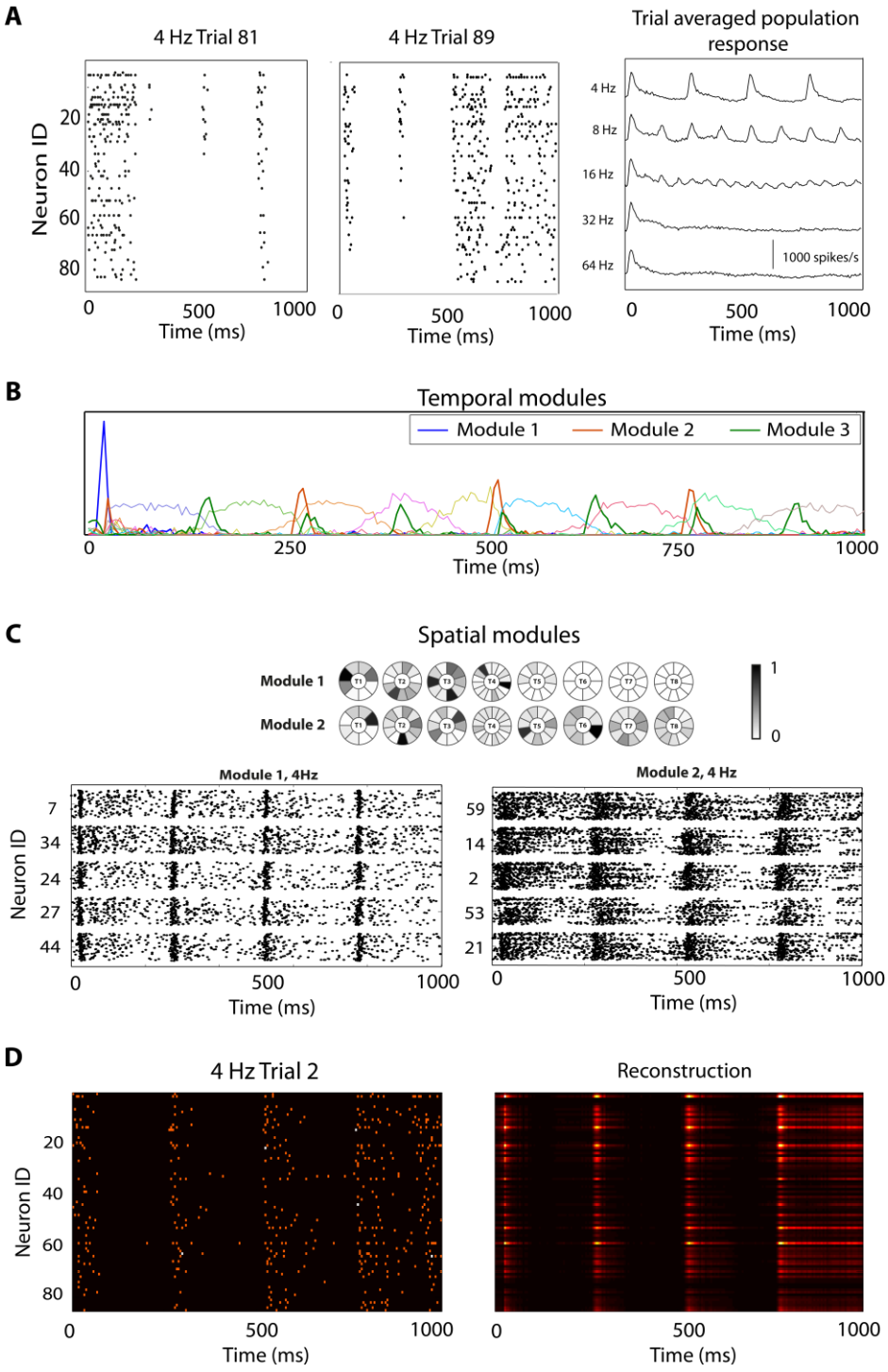

Figure 2.3:Space-by-time module structure for responses of 85 A1 neurons to click sequences. A: The population responses for two trials during 4 Hz sequences (top left and middle) and the trial-averaged population responses to the five click sequences (top right). B: Temporal modules extracted by space-by-time NMF from the responses of the same neurons to click sequences are shown at the bottom. C: Spatial modules extracted by space-by-time NMF. The raster plots below spatial modules show spiking activity of the five neurons that are assigned the highest weight in each spatial module for the 4 Hz click sequence. D: An example reconstruction (right) of the population response during the second trial of the 4 Hz sequence (left) using the low-dimensional description of the trial.

Temporal modules describing this population activity are shown in Figure 2.3B. Modules 1 - 3 represent salient stimulus related activity patterns from the trial-averaged population response. The first module that spans 15 - 30 ms of the response with a peak at 25 ms describes the initial response of neurons to click trains. The second module mainly describes the responses to repeating clicks in the 4 Hz sequence while the third module mainly indicates the onset responses to repeating clicks from 8 Hz sequences that are not present in the 4 Hz sequences. The temporal differences between the second module and the third module indicate that there is a late component in the population response to 8 Hz clicks compared to 4 Hz clicks. The allocation of temporal modules in this manner by dividing the full response into the response to the first click and to the subsequent clicks appears to enable the model to capture common stimulus related variability with the ability to distinguish between click sequences. The remaining modules in the decomposition describe the activity that occurs between the onset responses to 4 Hz and 8 Hz sequences and capture responses to repeating clicks in the 16 Hz sequences that occurred in these periods. They further count spikes in these intervals for 32 and 64 Hz sequences. Separate modules are allocated to describe the activity in these intervals rather than using a single module that span across all intervals because the population activity has a structured trial-to-trial variation.

Neurons in A1 are usually characterized using parameters such as the characteristic frequency, best frequency and the frequency bandwidth of the neuron (Schreiner et al., 2011). The characteristic frequency is the frequency in which a neuron produces a response at the lowest intensity from all the frequencies used in the experiment. The best frequency of a neuron is the frequency at which the maximum response is obtained at a given sound pressure level. The frequency bandwidth is the frequency range that produces an excitatory response at a given sound pressure level. Neurons in Al are often located as localized groups that have similar characteristic frequencies (Kilgard and Merzenich, 1999; Rutkowski et al., 2003). When moving along the posterior to anterior direction in the rat A1, the characteristic frequency of the neurons change from low frequency to high frequency (Sally and Kelly, 1988) indicating the tonotopic organization in A1. Figure 2.3C shows spatial modules derived from the population responses mapped onto a schematic figure of the eight tetrode configuration used during the recording session. Tetrode 1 contained neurons with characteristic frequencies in the high frequency regions while tetrode eight contained neurons with their characteristic frequencies in low frequency range. The first module consists of neurons from the first four tetrodes while the second contains neurons from all tetrodes. The two raster plots at the bottom of the panel show the responses of the five neurons that have the highest weight in each spatial module for the 4 Hz click sequence. This clearly shows that the first spatial module consists of neurons that have short onset responses while the second spatial module groups neurons that also display a sustained response. The co-activation of neurons across the tonotopic axis is expected when stimulating with clicks since a single click contains a large range from the frequency spectrum. Thus we observe that the spatial modules obtained with space-by-time NMF groups neurons that tend to fire together and have similar firing profiles. We then investigated whether this would give rise to covariations between neurons. We quantified signal and noise correlations between pairs of neurons that were identified as within modules and between modules. The neurons pairs within modules had a

higher positive signal correlation of 0.24 ± 0.63 compared to 0.15 ± 0.62 in pairs of neurons between modules (Wilcoxon rank sum test, $p<0.001$). Noise correlations in neurons pairs within modules (0.37 ± 0.22) was slightly, but significantly higher than pairs of neurons between modules (0.27 ± 0.23, Wilcoxon rank sum test, $p<0.05$). This indicates that the similarity in the firing profiles of neurons in modules reflect tendency to fire together in terms of correlated variability within and across stimuli.

Next, we analyzed the responses of the same neurons to long pure tones of 500 ms duration. Responses to long tones indicated of a state change during the experimental session. During the first half of the experiment, the responses of the neurons showed synchronized activity consisting of up-down states (Figure 2.4A). However during the second half of the experiment, the responses were more desynchronized (Figure 2.4B). When we analyzed the power in the multi unit activity (MUA, the details are given in Appendix 2) we found that the ratio of power in 0 - 5 Hz frequency range to that in 0 - 50 Hz frequency range was lower in trials 61 - 100 compared to trials 1 - 100 (two-tailed t-test, $p < 0.005$) indicating that the synchronization level decreased during the experiment (Curto et al., 2009). Changes in signal and noise correlations accompanied the change in the synchronization level (Figure 2.4B,C, (Pachitariu et al., 2015)). However, these changes showed differences between pairs of neurons recorded from different tetrodes. Both signal and noise correlations decreased in pairs of neurons recorded from tetrodes located in the anterior part of the cortex (two tailed t-test, $p < 0.05$) while pairs of neurons recorded from the posterior most two tetrodes did not show any changes in their correlation level.

We applied NMF to the binned responses from all trials, but trials from both states distributed across training and test datasets. Typically, the number of modules that give the optimal low-dimensional structure describing the responses of the neurons is unknown at the start of the analysis. Thus we initially apply NMF over all decompositions within a plausible range of decompositions suited for the experimental context. In this analysis we obtained decompositions that had 1 - 5 temporal modules and 1 - 20 spatial modules. This range of decompositions was chosen for the following reasons. Neurons responding to pure tones have temporal profiles that are described as phasic, phasic-tonic or tonic (Chimoto et al., 2002). Thus the possible range of temporal patterns exhibited by neurons to pure tones is low and could be expected to be captured well using 1 - 5 temporal modules. The tonotopic organization in A1 could be expected to give rise to spatial modules that groups neurons with similar frequency preferences (Sally and Kelly, 1988; Stiebler et al., 1997; Bizley et al., 2005; Schreiner et al., 2011). This would suggest that the number of spatial modules would be in the range of the number of tetrodes used for the recording. Considering that the neurons grouped into a single spatial module also have similarity in temporal firing profiles, the number of spatial modules could be expected to be in the range 1 - 20. Thus we obtained the range of decompositions spanning 1 - 5 temporal modules and 1 - 20 spatial modules.

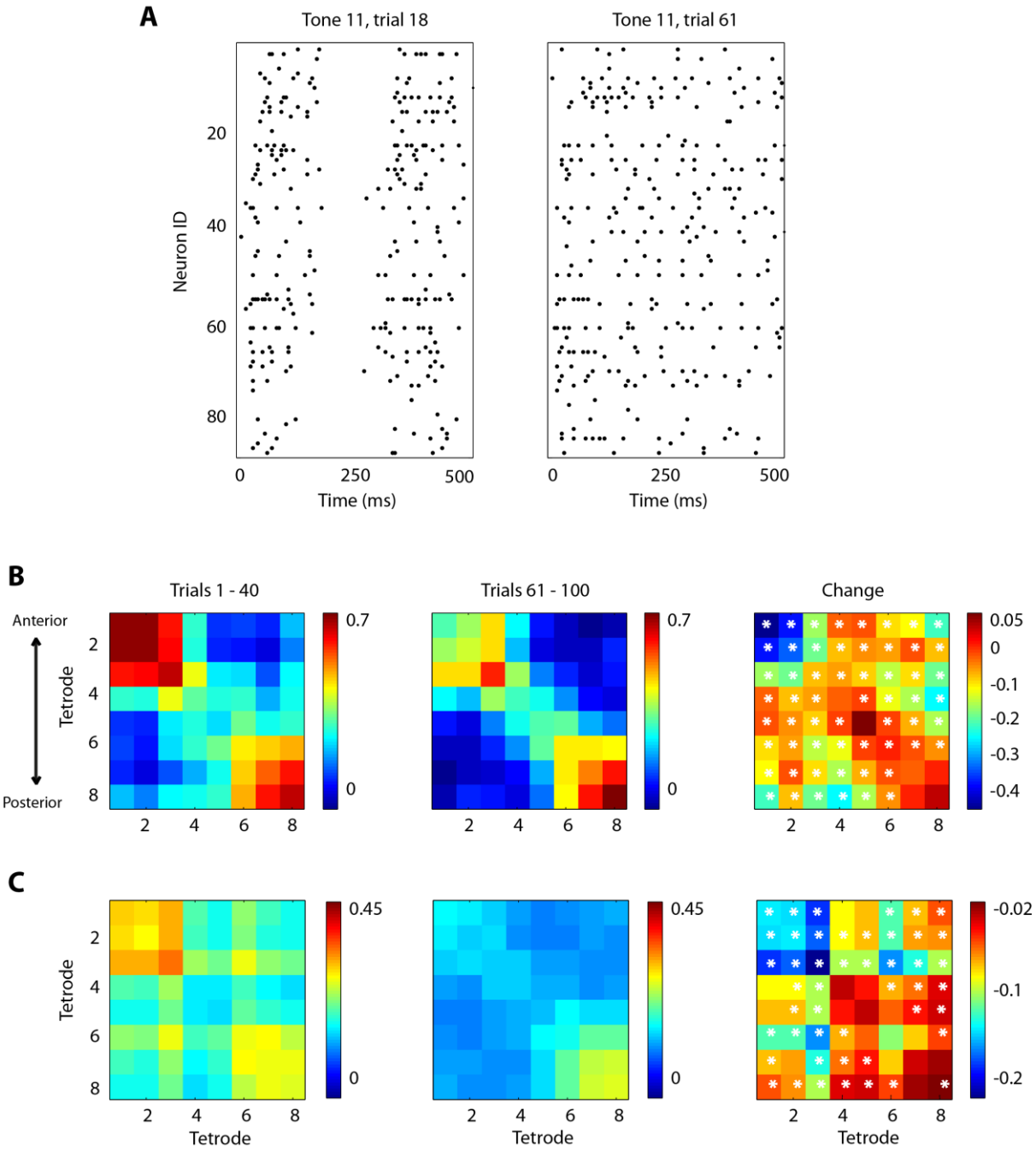

Figure 2.4: Changes observed in the neural responses of auditory neurons between the two levels of synchronization. A: Raster response of the population activity in a representative trial during the synchronized state (left) and during the desynchronized state (right). B: Signal correlations between pairs of neurons recorded from each tetrode. The figure on the left shows signal correlations between neurons in trials 1 - 40 while the middle figure shows the same for trials 61 - 100. The figure on the right indicates the change in signal correlation when the state became desynchronized. * show significant changes (two-tailed t-test, $p < 0.05$). CB: Noise correlations averaged across tones in pairs of neurons recorded from each tetrode using the same figure arrangement as in A.

The temporal modules extracted by NMF for three temporal modules and 20 spatial modules decomposition is shown in Figure 2.5A. The first module spans 15 - 25 ms period after application of the stimulus while the second temporal module spans the subsequent 25 - 120 ms period. Together, these describe the phasic response of a phasic-tonic cell. The temporal difference between the peak values of these two modules describes the activity of neurons with different onset latencies. The third module describes the sustained component of a phasic-tonic cell. The coefficients in the insets of the two figures below describe how the shown responses of the two neurons are encoded by the model. There is a high activation level in the first temporal module for the response of the neuron 15 representing its short onset response during the time window of the module. In contrast, the second and the third temporal modules carry high weight for the tonic firing neuron 11. The activation level of the coefficients of this neuron is approximately proportional to the number of spikes in the time window of the respective module. This give rise to a higher activation level of the third temporal module compared to the second temporal module. Thus, the sustained component has a role of rate coding while the precise short duration modules signify the contribution from precise spikes at the onset. In the two temporal modules, 20 spatial modules decomposition shown in Figure 2.5B, the phasic response is extracted as a separate module by space-by-time NMF. Therefore, the temporal module structure derived from NMF is suited to describe all types of temporal activity patterns observed in auditory neuron responses to tones.

We typically choose the optimal decomposition to be the one which has the optimal decoding performance. In this dataset, the number of correctly identified trials initially increased at a high rate with the increase in the number of spatial modules for a given number of temporal modules (Figure 2.5C). When reaching five spatial modules (which had a decoding performance of 45.7 % $\pm$ 1.24 %), the slope changed sharply and the increase in the performance occurred at a lower rate. At 20 spatial modules, the performance was 60.11 % $\pm$ 2.33%. Previous studies using NMF to analyze muscle synergies show a change in the slope of the performance measure used when the number of modules increases (d'Avella et al., 2003; Cheung et al., 2005) and the number of decompositions at which the change in the slope occurs is often taken to be the optimal decomposition. However, with regard to tonal responses, the difference in the decoding performances between the maximum number of spatial modules we extracted and the module at which the slope of the curve changes was considerable (close to 15%). When we examined the spatial modules for the five spatial module decomposition (Figure 2.5D), we found that the modules had a coarse organization. The first three spatial modules composed of neurons mainly from tetrodes 1 - 4, while the last spatial module composed of neurons mainly from tetrodes 6 - 8. Five neurons were selectively given higher weight in the modules. Four of these neurons (neuron 2, 11, 14 and 59) are among the five neurons with the highest stimuli-averaged firing rate. The 20 spatial module decomposition, which had the highest decoding performance, indicated that the weights of the neurons in spatial modules gradually shifted from neurons with characteristic frequencies in the high frequency regions to neurons with their best frequencies in the low frequency region. However, half of the spatial modules were assigned almost only to one neuron (Figure 2.5E) thus implying that the stimuli could be better described with less

emphasis on the population activity. When we compared the signal correlations of pairs of neurons within modules to pairs of neurons across modules, pairs of neurons within modules had a higher positive signal correlation of 0.43 ± 0.38 compared to 0.25 ± 0.39 of pairs of neurons between modules (Wilcoxon rank sum test, $p<0.001$). When we investigated noise correlations, we found that noise correlations in pairs of neurons located within modules (0.26 ± 0.26) were again positive and higher than those for pairs of neurons located between modules (0.14 ± 0.25, Wilcoxon rank sum test, $p<0.05$). This indicates that although many spatial modules derived for this dataset are sparse, pairs of neurons classified into spatial modules have similar tendency to fire. We discuss this finding in detail in the discussion.

Contribution of space and time dimensions for discriminability

Figure 2.6B indicates that the stimulus discriminability is higher for both datasets, when the spatial and temporal dimension are taken together in comparison to what can be achieved when considering either the spatial dimension or the temporal dimension independently (Wilcoxon rank sum test, $p < 0.001$). For tones, in which nine tones between 2 - 32 kHz frequency range with 0.5 octave separation were applied at 30 dB and 60 dB sound pressure levels, the loss of structure in the spatial dimension resulted in a large decrease in the ability to discriminate compared to when the full spatial and temporal information is present (Wilcoxon rank sum test, $p < 0.001$). This can be understood in the context of the tonotopy in the A1 neurons. The characteristic frequency of the neurons in the dataset varied from 40.03 ± 23.22 kHz in the first tetrode to 1.76 ± 0.6 kHz in the eight tetrode. Thus, the variation of frequencies along the anterior-posterior axis could contribute largely towards encoding tone frequencies and intensities. In the click sequences, the temporal dimension gives better performance in terms of identifying click trains compared to when only using the spatial dimension (Wilcoxon rank sum test, $p < 0.001$). This could be mainly due to the temporal nature of the stimuli; click sequences differ based on the repetition frequency and each click contains a large frequency range that evoke activity from a large portion of neurons in the population simultaneously.

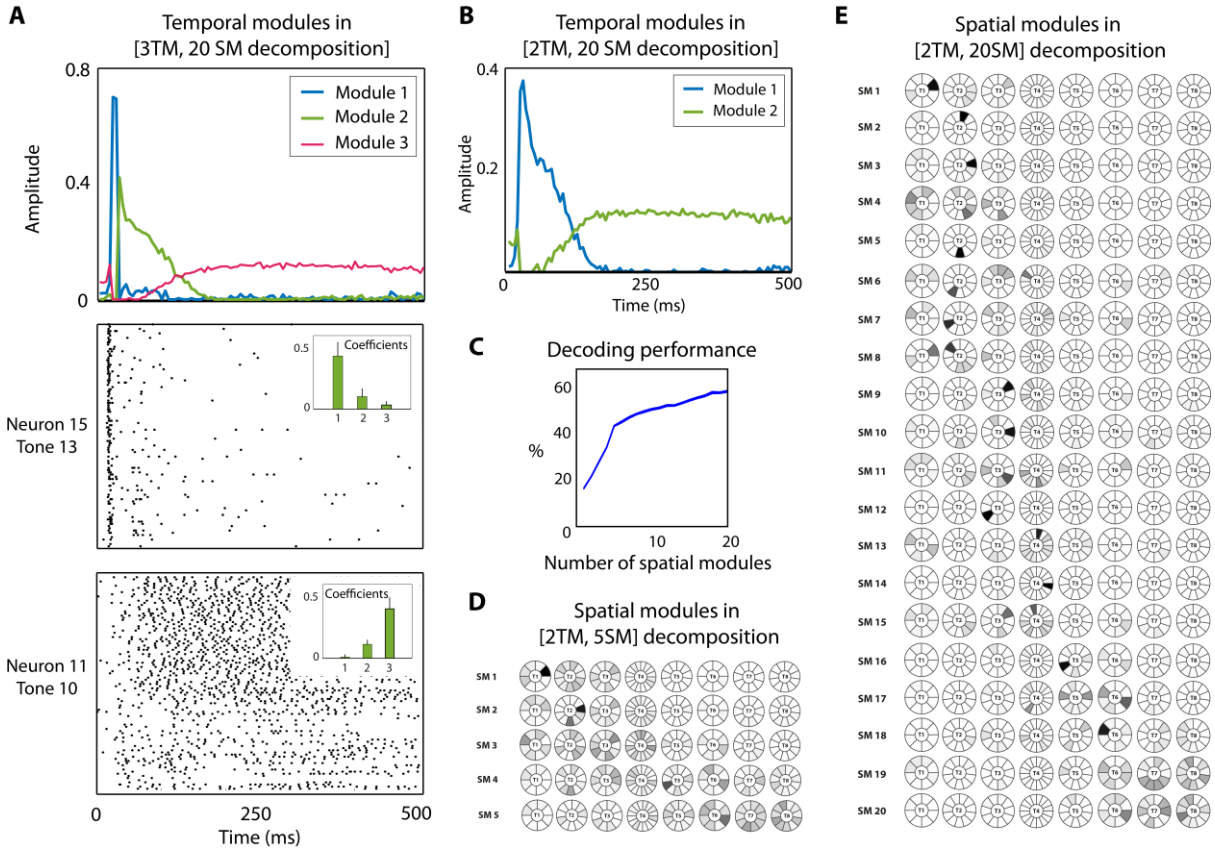

Figure 2.5: Space-by-time module structure for responses of 85 A1 neurons to long tones. A: Temporal modules extracted by space-by-time NMF for decomposition with three temporal modules and 20 spatial modules (top) and raster plots showing responses of two neurons to two tones (middle and bottom). Each raster plot contains spikes of the neuron during 100 trials of the indicated tone. Insets in the raster plots show the activation coefficients corresponding to each temporal module that describe the activity of the neuron. B: Temporal modules extracted by space-by-time NMF for decomposition with two temporal modules and 20 spatial modules C: The variation of the test set decoding performance for two temporal module decompositions when the number of spatial modules increases. D: Spatial modules extracted by space-by-time NMF for the decomposition with two temporal modules and five spatial modules depicted on a schematic illustration of the eight tetrodes (T) used in the recording. Each row shows the weights of the neurons in one spatial module (SM). The weight of each neuron in spatial module takes a value between 0 and 1. All neurons recorded from one tetrode are shown around the respective tetrode. E: Spatial modules extracted by space-by-time NMF for the decomposition with two temporal modules and 20 spatial modules

.

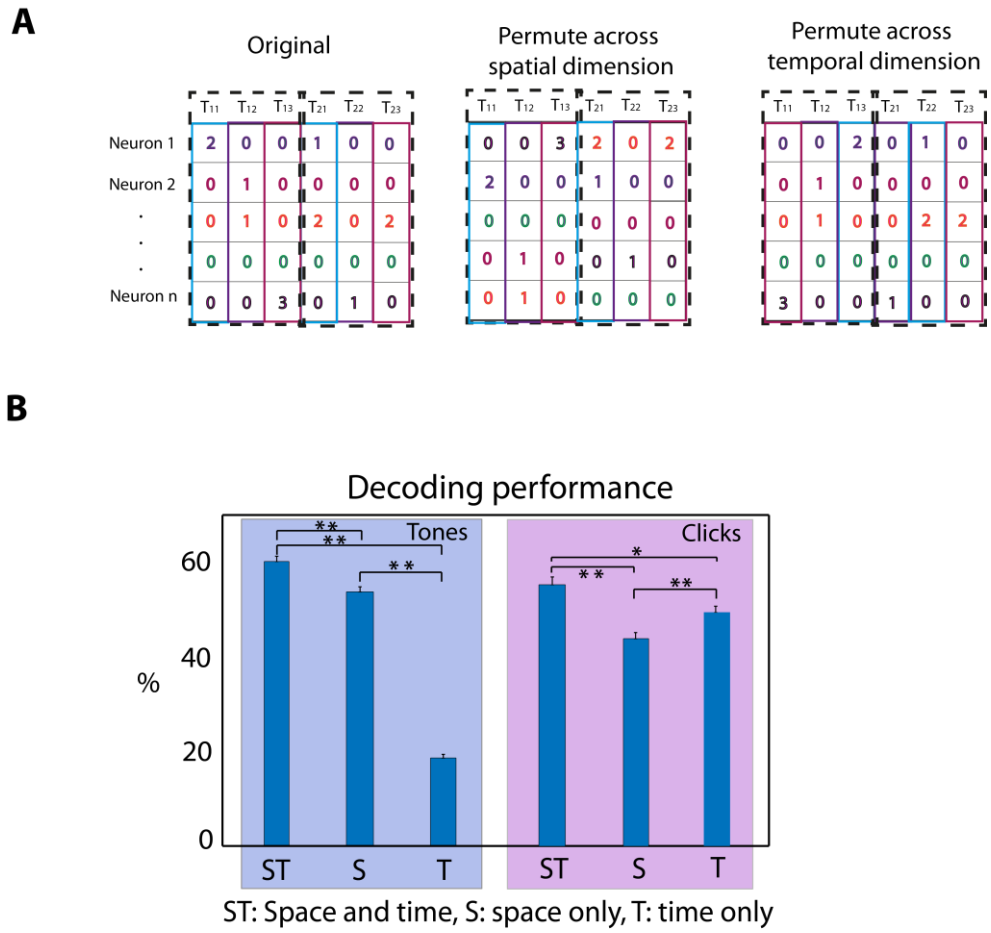

Figure 2.6: Stimulus discrimination capability of the low-dimensional representations extracted by NMF. A: An illustration of the permutation procedure we implemented to obtain information present only in one dimension. To obtain the information present only in the spatial dimension, we randomly permute the spike trains between neurons while keeping the temporal structure of each trial intact. To eliminate information in the time dimension, we randomly permute the time bins while keeping the spatial relationships in spikes intact. B: The decoding performance obtained when using information present in both spatial and temporal dimensions, either of the spatial or temporal dimensions for the neural responses to tones and clicks. (**: $p < 0.001$ and *: $p<0.05$; Wilcoxon rank sum test, erorr bars show s.d.)

2.5.2 Using NMF to study population coding of stimulus location in rat somatosensory cortex

Experimental details

The dataset recorded from somatosensory neurons was previously analyzed in several studies (Panzeri et al., 2001; Petersen et al., 2001) and the experimental methods are fully described in (Lebedev et al., 2000). Experiments were performed in accordance with the NIH and international standards on the use of experimental animals. Twenty-two adult male Wistar rats (weighing ~350 g) were used. The animal was anesthetized with urethane (1.5 g/kg body weight, i.p.). The animal was placed in a stereotactic apparatus (Narishige, Tokyo) and left somatosensory cortex exposed by a 4 mm diameter craniotomy. During the experiment, the body temperature was maintained near 37.5°C and the depth of anesthesia was held consistent by monitoring hindpaw withdrawal, corneal reflex, and respiration rate. The animal was perfused with saline followed by 4% paraformaldehyde at the end of the experiment. A flattened slab of neocortex was frozen, cut into 40 μm tangential sections after postfixation in 20% sucrose. To visualize barrel-columns in layer IV it was processed to label nitric oxide synthase activity (Valtschanoff et al., 1993).

Whiskers C1 – C3, D1 – D3, and E1 – E3 were stimulated independently at 3 mm from their base using a piezoelectric wafer (Morgan Matroc, Bedford, OH) that was controlled by a voltage pulse generator (A.M.P.I., Jerusalem). The stimulus was a step function of 80 μm amplitude and 100 ms duration delivered at a rate of 1 Hz, 50 times for each vibrissa.

Recording was done with an array of six tungsten electrodes, arranged either as a single row or as a 2 × 3 matrix, with 300 ± 50 μm horizontal separation between adjacent electrode tips. It was advanced into the cortical barrel field, focused on barrel-column D2 (typically 1–2 electrodes penetrated barrel D2) in 100 μm steps with an effort to sample recording sites throughout the cortical depth although the majority of neurons were likely to have been located between 300–950 μm.

The recorded activity was amplified, band-pass filtered between 300–7500 Hz and were digitized at 25 KHz, 32 points per waveform, time-stamped with 0.1 ms precision (Datawave, Boulder, CO), and stored on a Pentium PC for offline analysis. A custom template-matching algorithm identified spikes using differences in shape and amplitude. Typically 1 – 2 single units were identified from each electrode. The full dataset contained 100 neurons each recorded from cortical barrel-columns D1 - D3.

Data analysis

We binned spikes in the 0 - 100 ms period after stimulus onset into 5 ms time bins and applied space-by-time NMF and spatiotemporal NMF. We adopted a modified decoding

procedure from that used in section 2.5.1 that is better suited for the low number of trials available in the somatosensory dataset, which is detailed below.

We calculated the signal correlation and noise correlation between pairs of neurons grouped into spatial modules in a similar procedure as in section 2.5.1. Due to the nature of spatial modules obtained in responses to whisker deflections, we chose the threshold such that approximately half of all pairs of neurons were assigned as within module pairs to obtain conservative estimates of signal and noise correlations.

Decoding

We first generated twenty trial randomized datasets. We divided each dataset into a training dataset with 40 trials and a test with eight trials. We used the training set to identify the optimal decomposition with the following procedure. We divided the trials in the training set into ten equal sized partitions of four trials each and implemented a ten-fold cross validated decoding procedure. For each fold, we kept the four trials of one partition as the test set and obtained NMF modules from the 36 trials of the remaining partitions. We used the coefficients identified in the training set and test set to train a linear discriminant classifier. We repeated this ten times with non-overlapping test set trials during each fold and calculated the overall decoding performance to be the mean number of trials correctly assigned to each stimulus over the ten folds. Using this performance, we identified the optimal decomposition to be the decomposition that gives the highest performance with the least number of modules. When multiple decompositions gave similar decoding performance, we selected the decomposition that had the lowest number of parameters. Once the optimal decomposition was selected, we identified the fold in the optimal decomposition that had a decoding performance closest to the mean value. We used the modules obtained in this fold to extract coefficients for the eight trials separated as the test set to obtain the decoding performance corresponding to the dataset.

Spatiotemporal NMF module structure

The module structure identified by spatiotemporal NMF is shown in Figure 2.7B where each line is the color coded activity pattern of the respective neuron in the module. Spatiotemporal NMF identified the first three modules to be the population activity for D1 - D3 stimuli. They include neurons from barrel columns of non-principle whiskers that responded in addition to neurons from the barrel-column of the principle whisker. In each module, the temporal activity to non-principle whisker is delayed with respect to the activity for the principle whisker. This is in accordance with the spiking activity in our data (Figure 2.6A, D) and previous reports in the literature (Armstrong-James and Fox, 1987; Petersen and Diamond, 2000). The last module grouped population activity observed when stimulating non-principle whiskers C1 - C3 and E1 - E3.

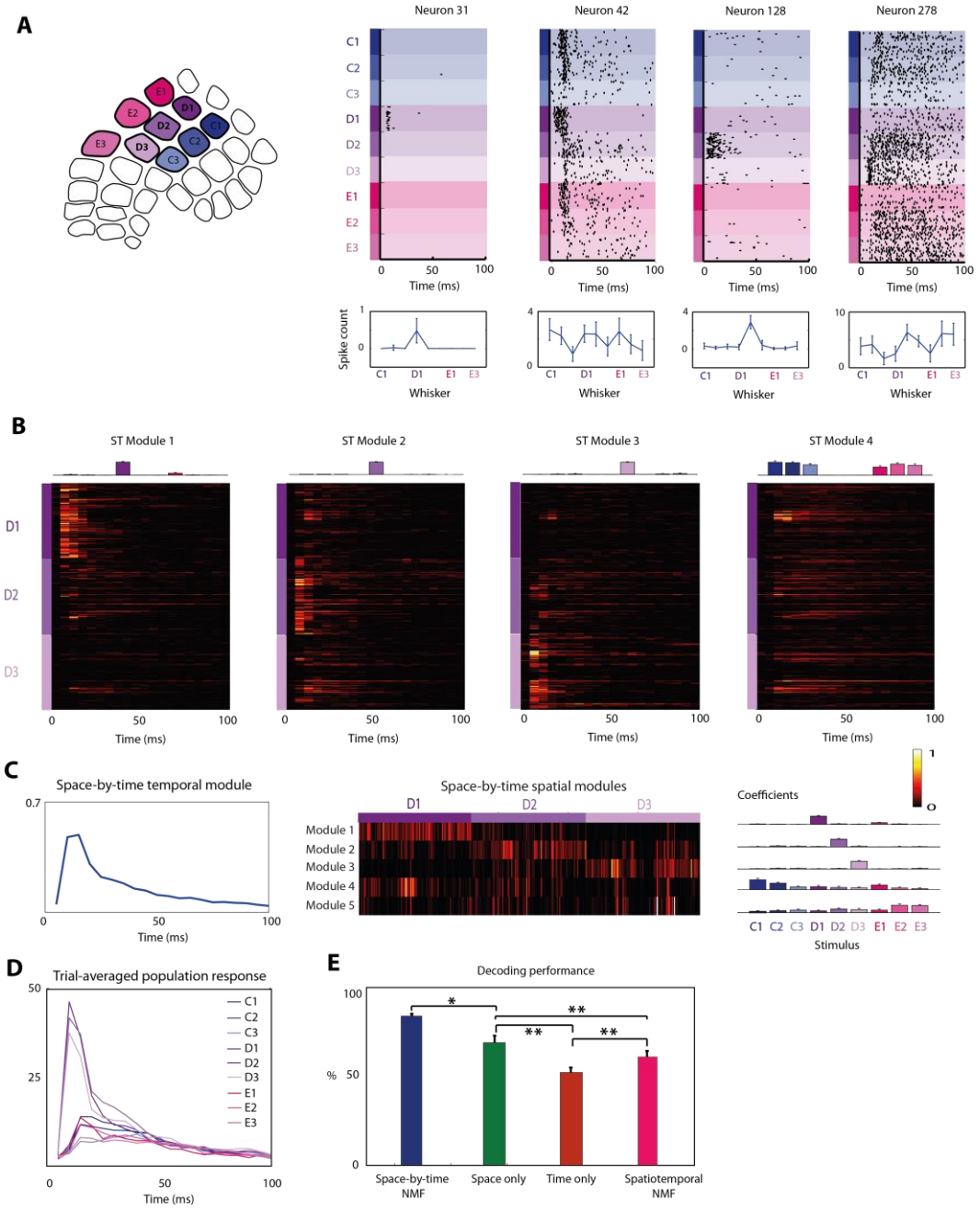

Figure 2.7: Module structure and the decoding performance of spatiotemporal NMF and space-by-time NMF on the response of 300 neurons to whisker deflections. A: Responses were recorded from 100 neurons each from barrel columns D1 - D3 (left). Example responses of four neurons are shown in the raster plots (right) with the tuning curves of each neuron shown at the bottom. B: Modules identified by spatiotemporal NMF. Each horizontal line shows the color coded temporal pattern of one neuron. Barrel columns of neurons are indicated on the right. The activation level of each module for each stimulus is shown above each module. C: Temporal module (left) and spatial modules (right) extracted from space-by-time NMF. Coefficients show the activation level of each spatial module for each stimulus. E: Trial-averaged population responses for each whisker deflection. E: Decoding performance. (**: $p < 0.001$ and *: $p<0.05$; Wilcoxon rank sum test, erorr bars show s.d.)

Space-by-time NMF module structure

Figure 2.7C shows five spatial and one temporal modules extracted by space-by-time NMF. Spatial modules identified by space-by-time NMF have clear composition with respect to neuron responses. The first three modules correspond to neurons responding mainly to D1 - D3 stimulations. Last two modules group neurons responding to non-principle whisker deflections in addition to principle whisker deflections. We found that the signal correlations in pairs of neurons within modules (0.3 ± 0.39) was positive and higher than pairs of neurons between modules (0.18 ± 0.41) (Wilcoxon rank sum test, $p < 0.001$) indicating that the neurons in the same module shared similar tuning to whisker location. However, noise correlations between pairs of neurons in the two categories were not significantly different for six stimuli. This could be understood because the dataset is a collection of independent recordings.

The temporal module in the optimal decomposition has an elevated activation level at the start of the trial. Space-by-time modules are generated using separate temporal and spatial modules by multiplying each spatial module with the temporal module (refer to the illustration in Figure 2.2B). Space-time modules that correspond to one temporal module and five spatial module decomposition are shown in Figure 2.7B. As can be seen in these plots, the high initial activation level of the temporal module results in a peak activity in time bins two and three in each space-by-time module (corresponding to 5 - 15 ms of the population response). Thus, it captures the highly reliable onset response of the neurons responding to principle whisker between 5 - 15 ms and the late onset of the responses to the non-principle whisker initiating at 10 - 15 ms (Figure 2.7D).

Two example reconstructions of the population response in two trials are given in Figure 2.8C-D. Each reconstruction is formed as a weighted linear summation of the space-time modules in Figure 2.8B. The weights used for each space-time module, which were obtained from the respective trial in the coefficient matrix, are indicated within brackets. Trial 8 of D1 whisker deflection is reconstructed with the contributions from space-time modules one and four that mainly compose of neurons from D1 barrel-column. Late responding neurons from D2 and D3 barrel-columns also have their peak activity value during 5 - 15 ms in the reconstruction. This is because only one temporal module is available to describe the activity of all neurons. However, the weights assigned to the second and the fourth space-time modules depict that these late responses are lower than that assigned to the first space-time module that describes the principle whisker responses. This indicates that the low-dimension description, incorporates information that has been shown to be present in the first spike latencies (Panzeri et al., 2001; Petersen et al., 2001) to a certain extent using the shape of the temporal module.

The reconstruction of a trial from C1 whisker deflection is shown in Figure 2.8D. This non-principle whisker response is mainly formed using fourth and fifth space-by-time modules. In the reconstructed trial, neurons from D3 barrel-column that respond to the whisker deflection have high values even though the responses occur later in the trial. This is because they discharge more spikes during the trial.

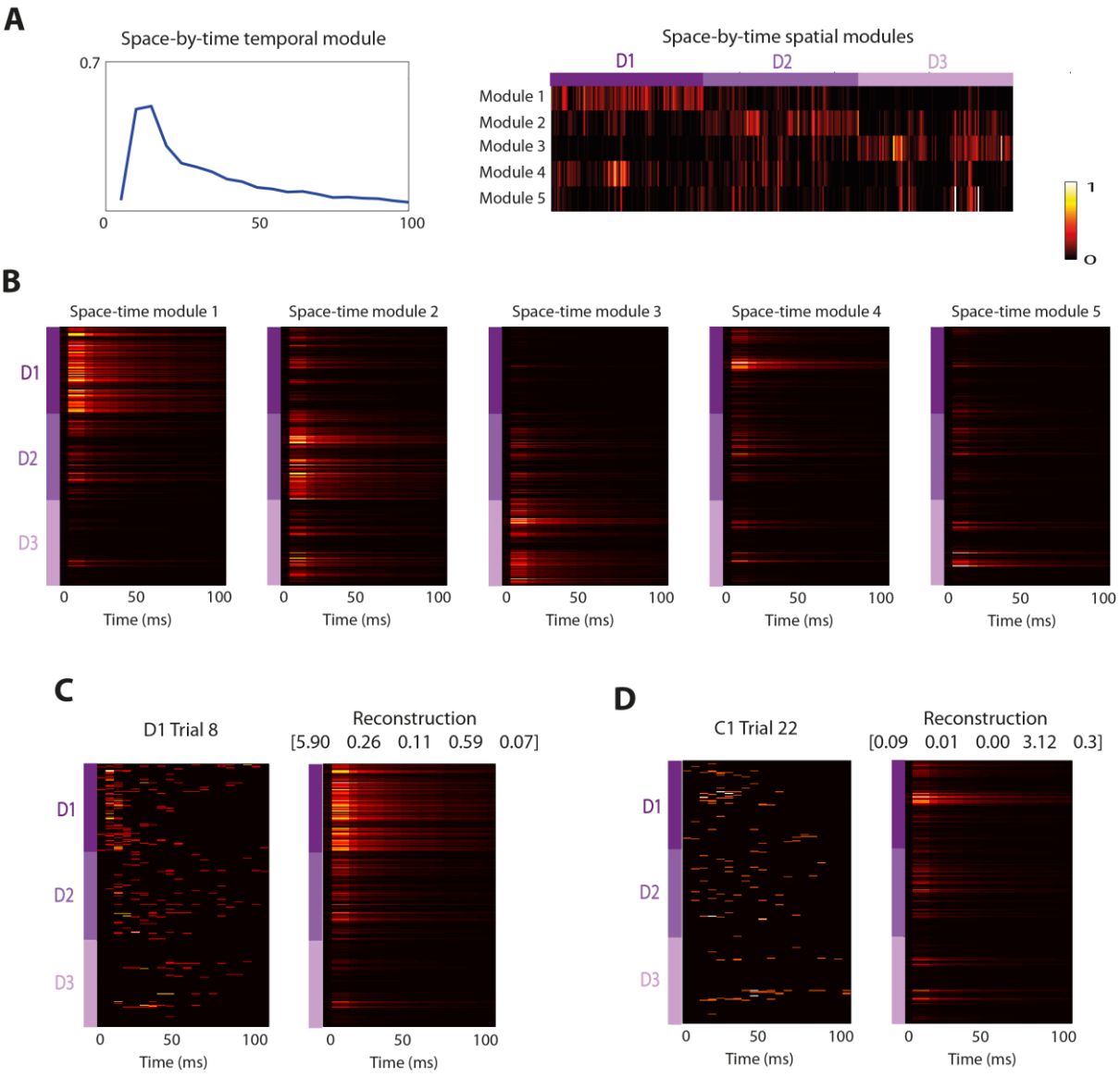

Figure 2.8: Reconstruction of population responses using space-by-time NMF A: Temporal module (left) and spatial modules (right) extracted from space-by-time NMF. B: Space-by-time modules that are described by the spatial and temporal modules. C: The population response for eighth trial of D1 whisker deflection (left) is reconstructed (right) using the five coefficients (indicated within brackets) that correspond to the five space-by-time modules in B. D: Reconstruction (right) of the population response (left) for 22nd trial of C1 whisker deflection.

Contribution of space and time dimensions to identification of stimulus location

From the two NMF methods that use information from both spatial and temporal dimensions, space-by-time NMF gave a higher decoding performance compared to spatiotemporal NMF (Figure 2.7E, Wilcoxon rank sum test, $p < 0.001$). This could be mainly attributed to the higher flexibility with which the population response could be described when spatial and temporal dimensions are separated in space-by-time NMF. Coefficients for the third and the fourth spatial modules in the space-by-time NMF have an inverse relationship. When whisker deflection moves from C1 to E1, the activation of the third spatial module becomes progressively lower (with the exception of E1 whisker) while the activation of the fourth spatial module becomes increasingly higher. The separation of the spatial and temporal dimensions enables space-by-time NMF to model a range of population responses by differentially weighting the space-time modules. However, spatiotemporal NMF requires more modules to generate the same range of variability in the population response. We limited the total number of modules extracted to seven modules. Increasing the number modules can lead to over-fitting. Due to the variability in the firing rates between neurons and since NMF update rules are designed to minimize an error measure describing the difference between the original dataset and the low-dimensional reconstruction, increasing the number of modules could lead to the formation of modules that capture the activity of high firing neurons rather than accounting for the population activity as a whole. In this particular case, where the dataset contain pooled responses obtained over multiple recording sessions, if there is some statistical difference between the responses from different sessions, there is a further possibility that the modules could be optimized to fit the trials recorded in a subset of recordings rather than the entire dataset. We did not have the details of the recording sessions to verify this possibility. Thus, we limited the maximum number of spatiotemporal modules and the maximum number of spatial modules to seven to obtain a conservative estimate of the decoding performance.

According to Figure 2.7E, using the information in both space and time gives a performance advantage of 14.9 ± 2.51 % compared to using only the spatial dimension (Wilcoxon rank sum test, $p < 0.05$). This indicates that the space-and-time representation has a component that cannot be represented only from the firing rates of the neurons alone. The only form that the one temporal module decomposition incorporates temporal information is through the temporal weighting of spikes. This mechanism gives higher importance to spikes that occur at the beginning of the trial compared to spikes that occur during the latter part of the trial. (Petersen et al., 2001) showed that at the level of pairs of neurons in the same dataset, 82-85 % of the total information in spike timing is contained in the timing of individual spikes and about 91 % of this information is contained in the timing of the first spike. This is mainly because the responses to principle whisker occurs prior to non-principle whiskers in coding stimulus location in somatosensory system (raster plots in Figure 2.6A,(Armstrong-James and Fox, 1987)). Thus, temporally weighing of spikes by space-by-time NMF enables us to incorporate this spike timing information with the information in neuron-to-neuron differences in firing rates at the population level to increase the ability to discriminate whiskers. Since the optimal decoding performance is obtained with only one temporal

module rather than using multiple temporal modules, this suggests that for these responses, a mechanism that could weigh spikes depending on their time of occurrence during the entire response period could discriminate stimuli better than a mechanism that can analyze the weighted spike counts in multiple temporal windows separately.

Figure 2.7E show that the difference in the decoding performance when using only information present in the temporal dimension compared to using information in both space-and-time is 31.73 ± 4.6 % (Wilcoxon rank sum test, $p < 0.05$). This indicates that the space-and-time representation has a component that is not represented by time alone and that this component is twice as large as that when only using the spatial dimension. However, considering that the temporal dimension can alone discriminate more than half the trials (decoding performance of 52.2 ± 2.8 %), coding of the stimulus location has a high redundancy in space and time in the population.

2.5.3 Using NMF to study encoding of visual information by populations of retinal ganglion cells

Experimental procedures

The datasets were recorded from retinal ganglion cells of dark-adapted *axolotl salamanders* (Ambystoma mexicanum; pigmented wild type) of either sex. All experimental procedures were performed in accordance with institutional guidelines of the University Medical Center Göttingen. Recordings were made at room temperature while supplying the retina with oxygenated Ringer's solution. Spike detection and sorting was carried out by an expectation-maximization algorithm for a Gaussian mixture model (Pouzat et al., 2002). The receptive fields of the cells were identified stimulating with spatiotemporal white noise and using the spike-triggered average (Chichilnisky, 2001). The spike-triggered average was separated into a spatial and temporal component using singular value decomposition(Gauthier et al., 2009). A two-dimensional Gaussian function was fitted to the spatial receptive field component to determine the center, size, and shape of the receptive field.

Visual stimuli were projected onto the photoreceptor layer of the retina using a gamma-corrected miniature OLED display (600 × 800 pixels) with monochromatic white light. A telecentric lens demagnified the stimuli to a pixel size of 7.5 µm × 7.5 µm. Average light intensity for all stimuli on the retinal surface was approximately 2.6 mW/m2, in the photopic range, on the retinal surface. Stimuli were presented with a custom-made software package developed with C++ and OpenGL.

Datasets were recorded for three types of stimuli; flashed natural images, shifted natural images and square wave gratings.

In the experiments in which natural images were flashed on the retina, 60 images were used to stimulate the retinal ganglion cells. The images (256 x 256 pixels) were selected from "McGill Calibrated Colour Image Database" http://tabby.vision.mcgill.ca/html/browsedownload.html (Olmos, 2004) and consisted of natural and artificial scenes spanning a field of view of 20° - 40°. RGB values of each image were converted to gray scale performing a weighted average across the three color channels. They were normalized using the know exposure time of the camera. Then the pixel values were linearly related to the absolute luminance values and the standard deviation was set to 50% of the mean intensity. To ensure that the maximal pixel values are within the physically available range of the display, pixel values that deviated > 100% from the mean in either direction were clipped. To minimize the artifacts induced by this clipping, we only selected images that had only few clipped pixels (i.e., not more than 0.035% of the pixels). Images are presented individually for 200 ms each in a pseudo-random sequence, with an inter-stimulus-interval of 800 ms. The datasets recorded for flashed natural images consisted of two sessions from two retinas that had 38 and 49 retinal ganglion cells.

In the experiments where shifted natural images were used, the image set was constructed by taking 9 different natural images from the "McGill Calibrated Colour Image Database"

(Olmos, 2004), and then presenting 9 different versions of each image obtained by spatially shifting each of them by 90 µm step (much smaller than the typical RF size). This gave us 81 images and each image was presented 19 times. Images were presented individually for 200 ms each in a pseudo-random sequence, with an inter-stimulus-interval of 800 ms in which a full field gray stimulus was presented. For data analysis, we used the first 300 ms of neural activity after stimulus onset. We recorded data from two retinas (n=23 and n=37 cells, respectively).

In different recording sessions, we used square-wave gratings of 900 µm spatial period and 60% visual contrast. Responses were recorded from 54 neurons to 60 shifted versions of the same grating, uniformly covering the complete range of spatial phases of the grating. Stimulus presentation and data analysis was done analogous for natural images.

Data analysis

When analyzing responses to flashed natural images, we used 30 trials from each recording and we used all 19 trials when analyzing responses to shifted gratings. In both cases, we used the first 300 ms after stimulus onset from the retinal ganglion cell activity and binned the responses into 10 ms time bins. To identify the effectiveness with which the concurrently extracted low-dimensional structure that contains both spatial and temporal information can be used to identify stimuli, we implemented a cross validated decoding procedure similar to that used for somatosensory data in section 2.5.2. To quantify the specific contribution of the spatial and temporal dimensions separately to the total information, we used the permutation procedure described in section 2.5.1.

To understand whether timing itself carried information, we computed the information carried by the first-spike latency of the firing of each neuron in response to the images. To compute this information we decoded the spike trains with space-by-time NMF performed exactly as above but applied to spike trains in which we deleted all spikes apart from the first one for each neuron in each trial.

Decoding

We divided each dataset into equal sized training dataset and test set with 30 trials each and used the training set to identify the optimal decomposition using a procedure similar to that in section 2.5.2. Briefly, we used leave-one-out cross validation with a linear discriminant classifier to obtain the optimal decomposition using the training set. We selected the optimal decomposition to be the one that gave the maximum average decoding performance in the leave-one-out decoding procedure. If multiple decompositions gave maximum performance, we selected the decomposition that had the maximum performance and the minimum number of parameters.

In addition to the NMF analysis, we evaluated the performance of a rank order decoder to discriminate stimuli. This decoder is based on the relative order of the first-spike latency of each neuron (Rullen and Thorpe, 2001; Johansson and Birznieks, 2004; Saal et al., 2009; Panzeri and Diamond, 2010). We applied the same procedure that was used in (Panzeri and Diamond, 2010) which is also very similar to that of (Johansson and Birznieks, 2004): In each trial, we ranked neurons according to their first-spike latency (using the MATLAB's *tiedrank* function). For each grating phase, we then constructed a template of the population rank order. We did this by first calculating the mean rank of each neuron over the training trials and then we calculated the tie-corrected rank order of the means. For each test trial, we also ranked neurons according to their first-spike latency and then calculated the Spearman rank correlations between the test trial rank order and each template. We selected the stimulus with the highest correlation as the decoder output.

Space-by-time module structure

The temporal modules extracted by space-by-time NMF from the neural responses to flashed natural images is shown in Figure 2.9A. They consist of two short duration modules centered at 140 ms and 160 ms and spanning ~50 ms. These describe the differences in response latencies of neurons in the dataset. The third module spans the 140 ms of the sustained response. Thus, space-by-time NMF describes the population activity of retinal ganglion cells natural images using relatively coarse time scale at the onset and a rate based modulation in the sustained response.

Figure 2.9B shows the eight spatial modules extracted by the algorithm. Each module is mapped onto the receptive fields of the respective cells in Figure 2.9C. The same weighting of each cell in the module in Figure 2.9B is also used in Figure 2.9C. This shows that the spatial modules appear to be composed of neurons that have nearby receptive fields, suggesting that similar firing profiles across time arise because nearby neurons receive more common stimulation from the natural images and have higher connectivity (Masland, 2012).

When full field gratings were flashed with different spatial phases, the first spike latency of neural responses depended on the spatial phase. This effect is visible in the three representative retinal ganglion cell responses shown in Figure 2.10A. This phase-dependence of latency is revealed even more clearly in Figure 2.10B, in which first-spike latencies of these neurons are shown in different colors for different grating phases. Figure 2.10C shows the temporal modules we obtained by applying space-by-time NMF. There is a clear relation between these modules and the spike latencies shown in Figure 2.10B. Temporal modules have their peaks at time points that corresponds very well with the first spike latencies of the neurons, suggesting that these modules describe latency differences and are suited to describe differences in responses to different stimuli.

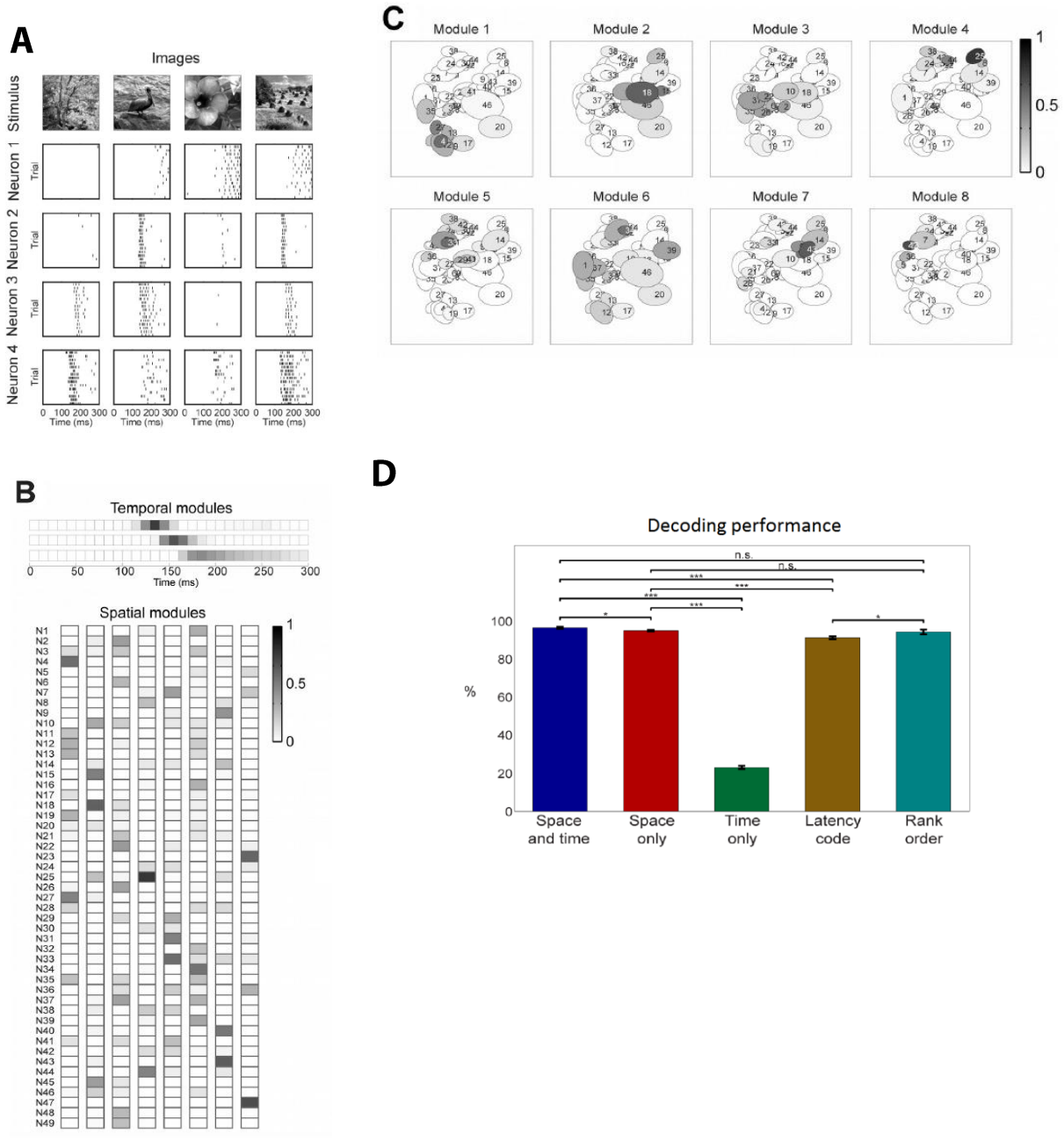

Figure 2.9: Analysis of neural responses of 49 retinal ganglion cells for flashed natural images. A: Responses of four example neurons to four natural images. B: Temporal and spatial modules extracted from space-by-time NMF. C: The composition of neurons in each spatial module depicted in the receptive field of each neuron where the weights of each neuron are set to the same values as in B. E: The decoding performance.

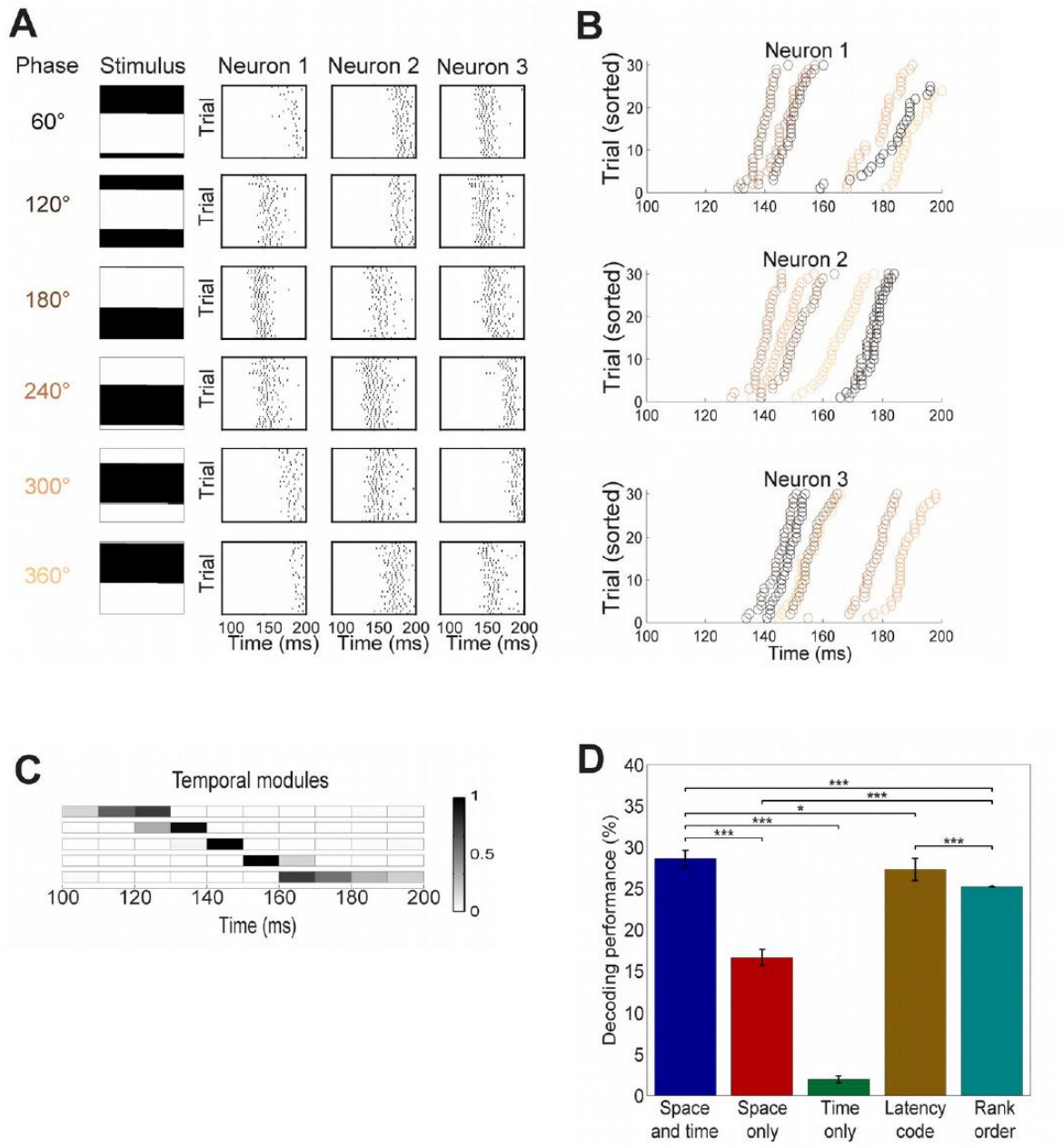

Figure 2.10: Analysis of neural responses of 54 retinal ganglion cells for full-field gratings. A: Responses of three example neurons to full field grating with six spatial phases. Each column shows the responses of the respective neuron to different spatial phases of the grating (left) B: The first spike latencies of three representative neurons. Trials are sorted for each neuron individually and only the period 100 - 200 ms is shown for clarity. C: Temporal modules extracted by space-by-time NMF for retinal ganglion cell responses to natural images. D: Decoding performance. (***: $p < 0.001$; two-tailed t-test. Error bars indicate s.e.m.)

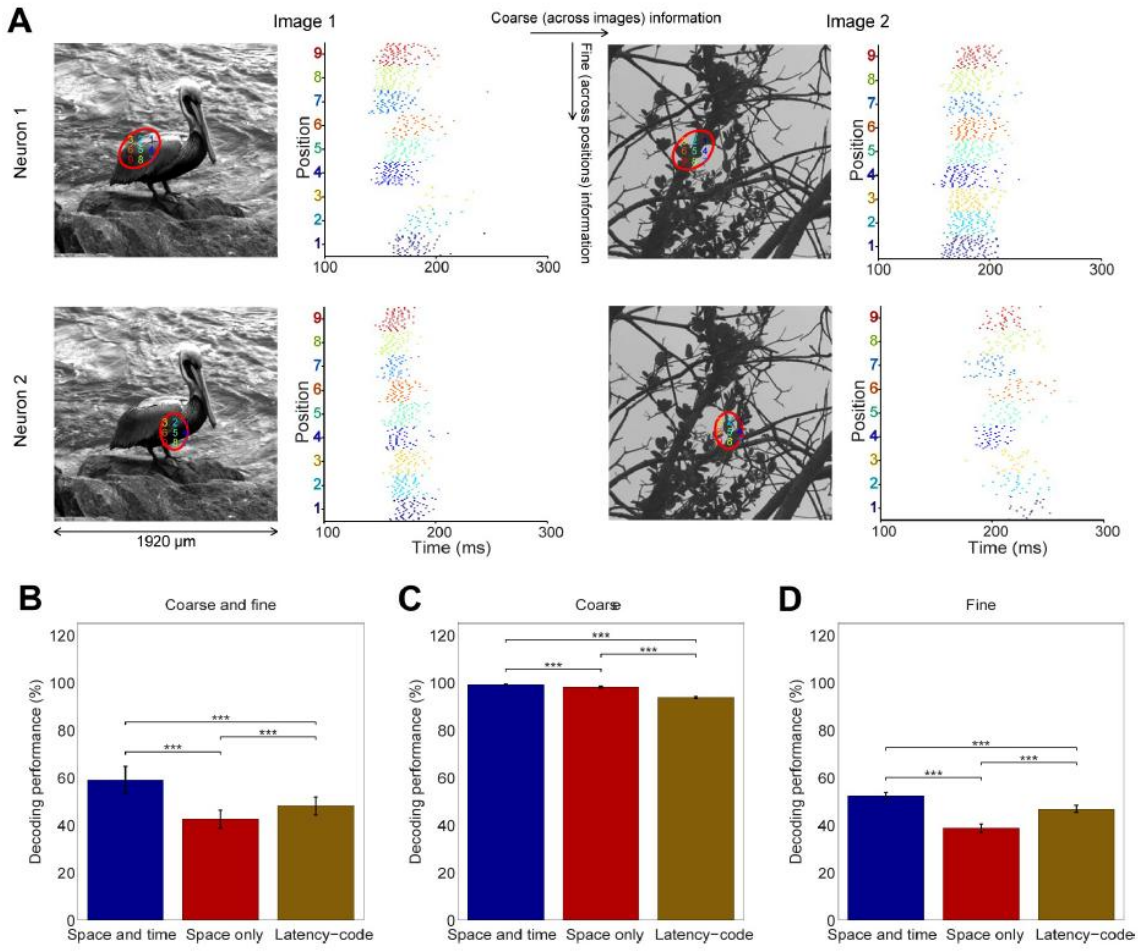

Figure 2.11: Analysis of neural responses to shifted natural images. A: The receptive fields of two representative neurons are shown in red on two images used (left) the raster responses of the neurons (right) showing only 0 - 300 ms of the response. The numbers in the ellipse denote the directions from the center of the ellipse in which the image was shifted. The responses of the neurons for these shifts are shown in different colours on the raster plot. B - D: Comparison of decoding performance when considering full information in space and time in all spikes (space and time), when considering only information in all spikes in the spatial dimension (space only) and when the responses of each trial contains the first spike of each neurons (Latency-code). The performance of decoding image ID and image position is shown in B. C shows the performance of decoding image ID for each position, averaged over positions while D shows the performance of decoding image position for each image id, averaged over images (***p<0.001; two-tailed t-test. Error bars indicate s.e.m.)

We found that both signal and noise correlations were positive and significantly larger for within-modules pairs of neurons than for across-modules pairs of neurons (one-tailed t-test, $p < 0.001$). The stronger signal and noise correlations in the "within modules" group indicate that spatial NMF modules represent ensembles of neurons that tend to fire together.

Contribution of space and time dimensions to stimulus discriminability

Using static images flashed on to the retina, we studied how neurons in the retina code image detail. We flashed the static images onto the retina for 200 ms after a 800 ms period of uniform gray screen. Suddenly flashing these images is an experimental paradigm often used to assess how the retina processes new visual information after a saccade or head movement (Gollisch and Meister, 2008). The spatial and temporal patterns found from the population activity in the retinal ganglion cells have near perfect discrimination of natural images as shown in Figure 2.9C. The activity within cells forming the spatial dimension had a performance that is only slightly, but significantly, lower (two-tailed t-test, $p < 0.05$). However, eliminating the spatial dimension generates a large decrease in the performance. indicating the importance of the spatial organization to code images in retinal ganglion cells. However, the brain many not rely only on the spatial dimension for stimulus discrimination. We performed two further analyses to identify whether a timing related code could also represent stimuli.

First, we computed the information carried by the first-spike latency of each neuron in response to the images. We removed all spikes except the first spike of each neuron in each trial and applied space-by-time NMF. For better comparison with the performance we obtained when using the full information in space and time, we used the same number of spatial and temporal modules that gave the optimal performance for the full dataset. As shown in Figure 2.9C, we found that decoding performance in the latency-code is very close to the performances when using information in both space-and-time and space-only, indicating that, at the population level almost all information is indeed redundantly carried by both neuron identity and first-spike latency (5%± 0.3% drop in performance compared to when using space-by-time information).

We then decoded the first-spike latency information using a rank-order decoder, which evaluates relative differences of first spike latencies in the population (Johansson and Birznieks, 2004; Saal et al., 2009; Panzeri and Diamond, 2010). Rank-order decoding led to a slightly higher decoding performance than that obtained with space-by-time NMF (94% vs. 91%, $p < 0.05$), suggesting that our approach is competitive with current methods. The information in first-spike latencies could be decoded by a downstream neural system that does not have independent knowledge of the stimulus time.

We tested the capability of space-by-time NMF to detect a strong impact of timing by using full-field gratings with varying spatial phases that is known to exhibit very strong timing dependent information in retinal ganglion cell responses in the form of first spike latencies and relative timing (Gollisch and Meister, 2008). We then applied space-by-time NMF and the permutation procedure in order to understand the relative importance of space and time in the population coding of spatial phase of images. Consistent with the visual inspection of the responses, this analysis showed a strong decrease in decoding (performance drop of 42% ± 1.37%, $p < 0.001$, Figure 2.10D) when destroying spike timing information. This means that the retinal population code of spatial phase contains crucial information on spike timing that cannot possibly be recovered from time-average firing rates only. Visual inspection of

responses, as well as previous analysis of small populations (Gollisch and Meister, 2008) suggests that first-spike latency is a key component of this population code. To test this hypothesis at the larger population level of tens of cells, we applied NMF and estimated the decoding performance on spike trains that contained only the first spike of each neuron. We found that first latencies carried almost all information contained in the full spike trains (27% vs. 29%, $p < 0.05$), demonstrating that the information carried by later spikes about image spatial phase is redundant to that already carried by first spikes. Decoding the first-spike latency information using a rank-order decoder led to a slightly lower decoding performance than that obtained with space-by-time NMF (25% vs. 27%, $p < 0.001$), demonstrating again that our approach is competitive with current methods and that information in first-spike latencies could be decoded by a downstream neural system that does not have independent knowledge of the stimulus time.

The importance of spike timing for coding small image differences was tested in the above and (to our knowledge) in previous experiments only using artificial grating stimuli. To verify this hypothesis in a natural image context, we performed a new experiment in which we simultaneously recorded RGC responses for a set of 81 natural images that contained both coarse and fine differences in within-RF image features. We first decoded which of these 81 stimuli was being presented. When decoding the stimulus identity, the discrimination between both coarse image features (differences between different images) and fine image features (differences between shifts of the same image) is required. As shown in Figure 2.11, we found that neglecting spike times and using only the spatial dimension led to a large loss of decoding performance (decoding that used only information in the spatial dimension was 16% less accurate than space-by-time decoding, $p < 0.001$). To investigate whether spike times and spike rates contributed differentially to fine and coarse image coding, we separated out their relative contribution to coarse and fine image coding. We first considered a set of nine stimuli, obtained by grouping all nine shifts of an image into a single stimulus class. Decoding these stimuli required the discrimination between image IDs (using only coarse image differences). In full agreement with the previous analysis obtained with the set of 60 different natural images, we found that this "coarse image decoding" had high performance when only using firing rates and not considering spike times(the decrease in the decoding performance when using only information in the spatial dimension was only 1% compared to when using the full information in space and time, $p < 0.001$, Figure 2.11C). Then we decoded fine image features using a set of nine stimuli, in which we used the nine shifts of a single image. This required identification of small shifts in the same image (using only fine image differences). In this case, and in agreement with the grating results, we found (Figure 2.11D) that this "fine image decoding" required spike timing: neglecting spike times led to a large loss of decoding performance (decoding using only information in the spatial dimension resulted in a 14% decrease in decoding compared to when using the full information, $p < 0.001$). We note that in all cases the latency code (evaluated with space-by-time NMF) carried a large fraction of the total information and was larger than spike count information for all decoding needing fine discrimination (Figure 2.11).

These results confirm that at the population level spike timing and the first spike latency is an important component used by retinal cells to code fine spatial differences in image features.

2.6 Discussion

We identified globally visible and recurring spatial and temporal patterns in three different large-scale neural datasets using NMF, a dimension reduction method that has been successfully used to analyze large-scale datasets in many domains, but has very little application in electrophysiological data to date. We investigated the ability of these concurrent spatial and temporal patterns to discriminate stimuli in comparison to what can be achieved when using either the spatial or the temporal dimensions alone.

In all three sensory datasets we analyzed; activity of auditory cortical neurons in response to long tones and clicks, the activity of somatosensory neurons to whisker deflections and the activity of retinal ganglion cells to natural images, NMF extracted physiologically meaningful patterns in a data driven manner. We compared two variants of NMF to study population responses concurrently in space and time: spatiotemporal NMF and space-by-time NMF. As observed for neural activity in the neurons from barrel-columns to whisker stimulations, space-by-time NMF gave higher performance in decoding stimuli compared to space-by-time NMF.

We analyzed how information in spatial and temporal dimensions is used by the three neural populations to code stimuli. When we analyzed the responses of retinal ganglion cells to flashed images, we found that coarse image features could be discriminated using a rate code while discriminating fine features required the availability of timing information above that available in the rate code. The temporal dimension and the reliability of onset responses were also important for whisker location coding by somatosensory neurons. We further describe the analysis of the somatosensory dataset in chapter 3 and discuss these results in detail in chapter 4.

We noted that the organization of spatial modules identified using responses to long tones was not completely in line with our expectations drawn from the tonotopy. If the activity of auditory neurons to long tones followed the tonotopical structure, we would expect the spatial modules to be composed of neurons localized between adjacent tetrodes. However, the spatial modules were sparser than expected based on the tonotopy. Neurons with high firing rates were allocated high weights and separated into individual spatial modules. The decoding performance continued to increase when the number of spatial modules increased. There are several properties in the tone dataset that could lead to these results

First, there is a relatively high degree of firing rate variability between neurons in the population (mean firing rate of 5.2 ± 4.48 spikes/s). Many applications that use update rules based on the Euclidean distance preprocess the data to minimize the variance in the dataset (Cichocki et al., 2009). Separating high firing neurons into individual spatial modules implies that the firing rate distribution of the dataset could potentially be a factor that may bias the algorithm. Such biasness towards high firing neurons is observed in other dimension reduction methods such as principal component analysis, where the responses are often z-scored in a preprocessing step (Cunningham and Byron, 2014).

Second, all NMF algorithms are iterative algorithms. Typically, the working matrices are initialized using random values and updated in each iteration using update rules that are designed to minimize a specific error measure. The space-by-time NMF algorithm we implemented uses the Euclidean distance (the squared error between the original dataset and the reconstructed dataset) to iteratively optimize the working matrices. According to the probabilistic interpretation of NMF (chapter 3, (Févotte and Cemgil, 2009), NMF update rules derived using the Euclidean distance have the implicit assumption that the variability in the data is Gaussian distributed. The neural data is more likely to have Poisson distributed variability (Koch, 2004) and update rules accounting for other noise models may be more appropriate for analyzing spike counts.

We observed a change in the synchronization level during the experiment, likely due to urethane anesthesia used in the experiment (Clement et al., 2008). The structure of the signal and noise correlations changed between the two states (Marguet and Harris, 2011; Pachitariu et al., 2015) and differentially between pairs of neurons recorded from different tetrodes. As detailed in Appendix A2, the relationship between the noise correlations and the geometric mean firing rates was also differentially modulated in pairs of neurons recorded from anterior tetrodes compared to those recorded from posterior tetrodes.

In summary, we suggest that the following could be the potential causes that gave rise to the high degree of sparsity we observed in spatial modules when analyzing responses to long tones; a) update rules of the space-by-time NMF being developed to suit datasets with Gaussian distributed variability whereas neural data often show sub-Poisson, Poisson or supra-Poisson variability, b) the sensitivity of the current update rules to neurons with high firing rates, c) the change in the firing characteristics and the correlation structure between neurons during the change in the synchronization state, and d) the firing reliability of neurons across tone frequencies.

In order to understand and improve the mechanisms with which NMF extracts a low dimensional structure from neural responses, we changed the update rules of the space-by-time NMF to suit sub-Poisson, Poisson and supra-Poisson variability present in neural responses. We further carried out statistical simulations as well as simulations using a conductance-based integrated-and-fire network to understand the influence of the properties of the population responses on the performance of space-by-time NMF. The details of this work are described in chapter 3.

Chapter 3: Extension of space-by-time NMF to model variability in neural spike trains

3.1 Abstract

In the previous chapter, we adapted two variants of NMF, spatiotemporal NMF and space-by-time NMF (originally proposed as sample-based non-negative matrix tri-factorization (sNM3F) (Delis et al., 2014)) to analyze large neural datasets concurrently in space and time. As discussed in detail in chapter 2, we found that NMF is able to extract a low-dimensional representation of a neural dataset that is intuitive and easy to understand, is representative of physiologically relevant details and can naturally account for certain properties such as those of onset responses. The space-by-time NMF model we introduced assumes that the spike counts are Gaussian distributed. In this chapter we extend it to model sub-Poisson, Poisson and supra-Poisson variability observed in experimental recordings from neural populations.

This chapter is organized as follows. We begin with a brief introduction followed by the details of the process through which we obtained the new algorithms. Then we report how we validated our new algorithms methodologically through statistical simulations, network simulation and by analyzing the auditory and somatosensory datasets. We report the insight we gained using our new algorithms and conclude the chapter with a discussion of the implications of our findings.

3.2 Introduction

When using any model to analyze data, a primary requirement is that it should be suited to account for the statistical properties of the dataset. Or in other words, the structure of the variability in the data should be appropriately captured by the model's assumptions. NMF is an iterative algorithm that typically initializes the working matrices to random values and iteratively optimizes them using a set of update rules. These update rules are designed to minimize a dissimilarity measure (also referred to as a distance, divergence or cost function) between the reconstructed dataset and the original dataset. The choice of the dissimilarity measure is based upon explicit assumptions about the nature of the noise in the data. When the applied dissimilarity measure optimally matches the noise distribution in the data, NMF can be formulated in a probabilistic framework as a composite generative model of the data (Lee and Seung, 1999; Févotte and Cemgil, 2009). In this formulation, an item in the original dataset is generated by adding noise to the reconstructed element using the probability distribution that describes the variability in the data. Furthermore, in this setting, minimizing the respective dissimilarity measure through NMF is equivalent to maximum likelihood estimation under the corresponding composite generative noise model (Févotte & Cemgil, 2009). Thus, NMF algorithms can integrate knowledge about the variability in the data into the factorization process through the update rules used in each iteration.

The initial space-by-time NMF model that we introduced and used in chapter 2 is based on a Gaussian generative model. This model assumes that the trial-to-trial variability in spike counts is distributed according to a Gaussian distribution that has a mean equal to the reconstructed spike count. However, spike counts are discrete quantities, and their variability is related to the mean spike counts as typically described using the Fano factor, defined as the ratio between the variance of the spike count to the mean spike count. Fano factors observed in spike counts of neurons span a continuum ranging from ~0.1 in retinal ganglion cells (Kara et al., 2000) to ~2.9 for cortical neurons (Oram et al., 1999). Poisson distribution is closely associated with modeling neural variability. This is because it has a variance to mean ratio of one and thus could be used to model spike counts that have a Fano factor of one (Figure 3.1). Spike counts that have a Fano factor of less than one, i.e. more reliable, are described as sub-Poisson while those with a Fano factor of greater than one are described as being supra-Poisson. The Fano factor can dynamically change during a trial, a phenomena often observed at the stimulus onset (Churchland et al., 2010). Given the range of heterogeneity present in the neural data, the non-negative discrete nature of the spike counts, and considering the sparsity we observed in the spatial modules corresponding to population responses for long tones, we wondered whether the ability of the space-by-time NMF to represent neural population responses could be improved by explicitly incorporating the knowledge about the neural variability into the update process. Thus, we optimized space-by-time NMF to model the trial-to-trial variability in neural data. We used Bregman divergences as the dissimilarity measure and utilized the generalization of update rules proposed by (Dhillon & Sra, 2005; Sra & Dhillon, 2006). We evaluated the new update rules with a range of simulations using spike counts generated using statistical models, a conductance-based integrate and fire network and applied them to somatosensory and auditory data that we analyzed in chapter 2.

This chapter is organized as follows. It begins with a short introduction to the probabilistic formulation of NMF. Then we detail how space-by-time NMF was extended to model neural variability. This is followed by the results we obtained from a detailed verification process. Finally, we discuss the findings from the new update rules and their implications when using space-by-time NMF as an analysis method.

3.3 Extension of space-by-time NMF

Before detailing how we modified the space-by-time NMF to model neural variability we will lay the ground for our work by briefly detailing the probabilistic formulation of NMF. We start with an introduction to the dissimilarity measures used in NMF in section 3.3.1 and describe how NMF could be interpreted as a generative model under the optimal noise distribution in section 3.3.2. We identify noise distributions to model neural variability in section 3.3.3. In section 3.3.4, we identify optimal Bregman divergences corresponding to these noise distributions. Finally, we derive the new update rules for space-by-time NMF in section 3.3.5.

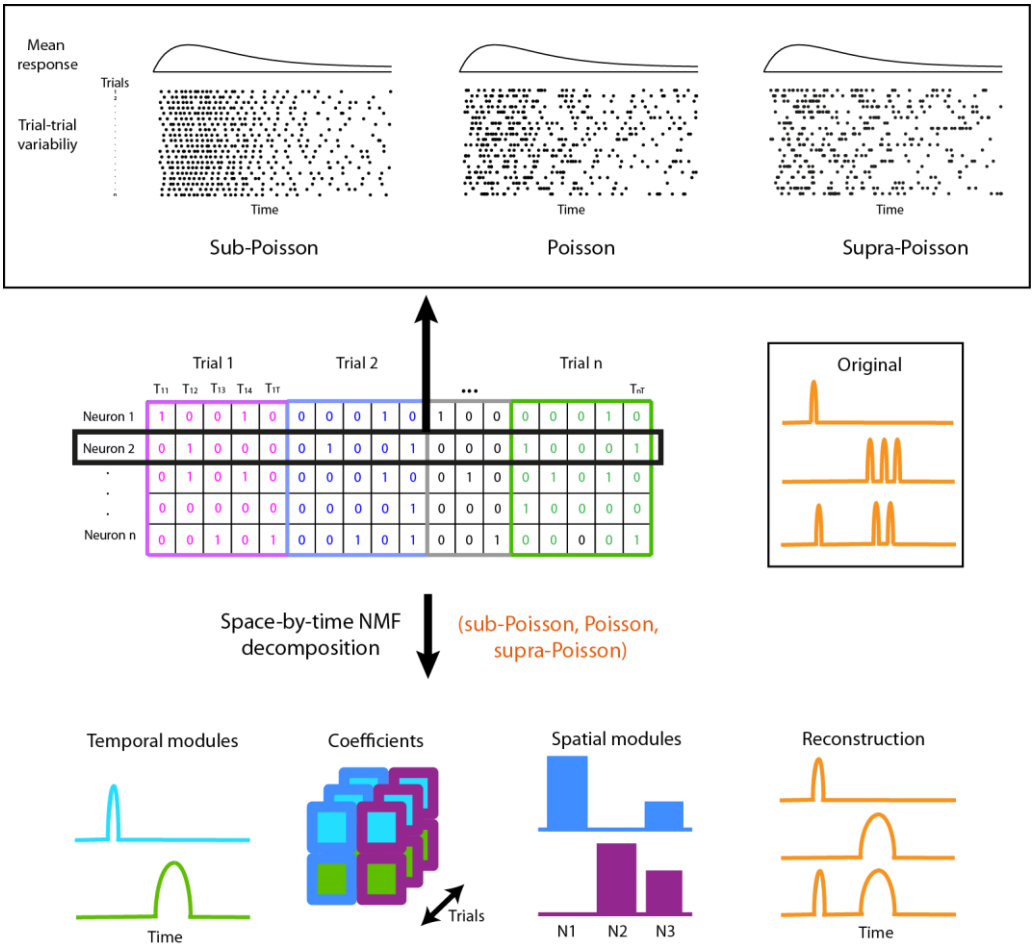

Figure 3.1: Illustration of space-by-time NMF under different noise models. Variability in neural spike counts is described as sub-Poisson, Poisson or supra-Poisson. Space-by-time update rules model these using different update rules.

3.3.1 Dissimilarity measures used in NMF

NMF is an iterative algorithm. At the start of the iterative process, the working matrices are initialized with uniformly distributed random values or values optimized using a method such as singular value decomposition (Boutsidis and Gallopoulos, 2008), spherical k-mean clustering (Wild et al., 2004) among others (Xue et al., 2008; Janecek and Tan, 2011). During each iteration the working matrices are updated to minimize a dissimilarity measure that specifies the difference between the original dataset $\mathbf{R}$ and the reconstructed dataset $\hat{\mathbf{R}}$. Typically, this dissimilarity measure is individually evaluated for each spike count $r_{t,n}$ of neuron n in time bin t in the dataset under the assumption that the total dissimilarity is separable.

$$D\left(\mathbf{R},\hat{\mathbf{R}}\right)=\sum_{n=1}^{n}\sum_{t=1}^{T}d\left(r_{t,n},\hat{r}_{t,n}\right) \tag{3.1}$$

The most commonly used dissimilarity measure is the Euclidean distance.

$$D_E\left(\mathbf{R},\hat{\mathbf{R}}\right)=\sum_{n=1}^{n}\sum_{t=1}^{T}\left(r_{t,n}-\hat{r}_{t,n}\right)^2 \tag{3.2}$$

Generalized Kullback-Leibler divergence (I-divergence) is the optimal dissimilarity measure when the data is Poisson distributed (Févotte and Cemgil, 2009) such as pixelwise noise in image datasets (Lee and Seung, 1999).

$$D_{KL}\left(\mathbf{R},\hat{\mathbf{R}}\right)=\sum_{n=1}^{N}\sum_{t=1}^{T}r_{t,n}\log\frac{r_{t,n}}{\hat{r}_{t,n}}-r_{t,n}+\hat{r}_{t,n} \tag{3.3}$$

Itakura-Saito (IS) divergence is used to analyze audio power spectrums (Févotte et al., 2009).

$$D_{IS}\left(\mathbf{R},\hat{\mathbf{R}}\right)=\sum_{n=1}^{N}\sum_{t=1}^{T}\frac{r_{t,n}}{\hat{r}_{t,n}}-\log\frac{r_{t,n}}{\hat{r}_{t,n}}-1 \tag{3.4}$$

Many families of divergences such as Csiszár f-divergences, Bregman divergences and Beta-divergences have been studied as dissimilarity measures for NMF under a variety of application settings. A comprehensive review of the properties of these measures are given in (Cichocki et al., 2009). We used Bregman divergences for the extension of space-by-time NMF to model neural variability. A short introduction to Bregman divergences is given in section 3.2.4.

3.3.2 NMF as a generative model

As summarized in (Févotte and Cemgil, 2009), NMF is formulated in a probabilistic setting when using certain noise models. Typically, each data point r_{tn} is modeled as $r_{tn}=\sum_k s_{k,tn}$, where $s_{k,tn}$ is a hidden source identified by NMF such that $s_{k,tn}\sim p\left(s_{k,tn}\mid\theta_k\right)$ where ND is the underlying noise distribution and $\theta_k\sim\left\{w_{:,k},h_{k,:}\right\}$. For noise distributions which are closed under summation, this results in a composite generative noise model parameterized by the reconstruction $\sum_k w_{n,k}h_{k,t}$. When the appropriate dissimilarity measure is chosen, the maximum likelihood estimation of the composite model parameter $\sum_k w_{n,k}h_{k,t}$ under the noise model is equivalent to minimizing the respective NMF divergence, up to a multiplicative factor and an additive constant.

For example, performing NMF using the Euclidean distance and the generalized Kullback-Leibler divergence is equivalent to maximum likelihood estimation of the mean of Gaussian and Poisson distributions respectively (Cemgil, 2009; Févotte and Cemgil, 2009). Performing NMF using Itakura-Saito divergence corresponds to either the maximum likelihood estimation of the variance in a proper complex Gaussian distribution or the maximum

likelihood estimation of the intensity under a multiplicative Gamma distribution (Févotte et al., 2009). We extended space-by-time NMF to model neural variability by choosing appropriate noise distributions to model neural variability and choosing the optimal divergences for these noise distributions.

3.3.3 Noise models for neural variability

Neural variability is typically described as sub-Poisson, Poisson and supra-Poisson. A graphical illustration of these regimes is given in Figure 3.1.

When Fano factor is one, spike counts are often modeled using the Poisson distribution.

$$r_{t,n} \sim p\left(r_{t,n}; \lambda_{t,n}\right) = \frac{\lambda_{t,n}^{r_{t,n}} e^{-\lambda_{t,n}}}{r_{t,n}!} \tag{3.5}$$

where λ is the mean spike count.

We model supra-Poisson variability using the negative binomial distribution (Onken et al., 2009; Scott and Pillow, 2012) parameterized as a generalized Poisson distribution.

$$r_{t,n} \sim p\left(r_{t,n}; \lambda_{t,n}, v\right) = \binom{r_{t,n} + v - 1}{r_{t,n}} \frac{v^{v} \lambda_{t,n}^{r_{t,n}}}{\left(\lambda_{t,n} + v\right)^{r_{t,n}+v}} \tag{3.6}$$

v is an additional parameter of the model that controls the dispersion of the spike counts. The dispersion of the spike counts increases with the decrease in v. The dispersion decreases and approaches Poisson variability when $v \to \infty$.

We model sub-Poisson variability using the binomial distribution (DeWeese et al., 2003).

$$r_{t,n} \sim p\left(r_{t,n}; \lambda_{t,n}, n\right) = \binom{n}{r_{t,n}} \left(\frac{\lambda_{t,n}}{n}\right)^{r_{t,n}} \left(\frac{n - \lambda_{t,n}}{n}\right)^{n - r_{t,n}} \tag{3.7}$$

n is an additional parameter of the model that controls the regularity of the spike counts. Spike counts become more regular with the decrease in n. They lose regularity and approach Poisson variability when $n \to \infty$.

For negative binomial and binomial models, we assume that the model parameter is the same for all neurons and time bins.

Our choice of distributions stems from the following reasons. First, they have already been used to model spike count variability in the literature. Second, all are discrete distributions, which suit the discrete nature of spike counts. Finally, both binomial and negative binomial distributions converge to Poisson distribution as the model parameters approach infinity. Both the negative binomial and the binomial distribution can therefore be understood as generalizations of the Poisson distribution.

We proceeded to obtain the optimal update rules for each of the distributions using appropriate dissimilarity measures. We used dissimilarity measures based on the family of Bregman divergences.

3.3.4 Bregman divergences to model neural variability

Bregman divergences (Bregman, 1967) is a class of directed dissimilarity measures introduced in the context of convex optimization. Each Bregman divergence is associated with a continuously differentiable strictly convex function defined on a convex set. If $\varphi : S \subseteq \mathbb{R} \to \mathbb{R}$ is a strictly convex function that has a continuous first derivative, the corresponding Bregman divergence $D_\varphi : S \times \text{int}(S) \to \mathbb{R}_+$ is

$$D_\varphi \left(r_{t,n}, \hat{r}_{n,t} \right) \triangleq \varphi \left(r_{n,t} \right) - \varphi \left(\hat{r}_{n,t} \right) - \nabla \varphi \left(\hat{r}_{n,t} \right) \left(r_{n,t} - \hat{r}_{n,t} \right) \tag{3.8}$$

where $\text{int}(S)$ is the interior of the convex set S and $\nabla \varphi \left(\hat{r}_{nt} \right)$ represents the gradient of φ evaluated at $\hat{r}_{nt}$ (Dhillon and Sra, 2005; Sra and Dhillon, 2006). Bregman divergences are non-negative and are zero if and only if $r_{nt} = \hat{r}_{nt}$. There is a unique Bregman divergence for every regular exponential family (Banerjee et al., 2005). Both Euclidean distance and generalized Kullback-Leibler divergence could be obtained as special cases of Bregman divergences by selecting the appropriate convex functions. Furthermore, the sum of two Bregman divergences is also a Bregman divergence (Sra and Dhillon, 2006). This property is used to specify the total Bregman divergence of the dataset, i.e. the total dissimilarity, in terms of Bregman divergences of each data element.

$$D_\varphi \left(\mathbf{R}, \hat{\mathbf{R}} \right) = \sum_s D_\varphi \left(\mathbf{R}^s, \hat{\mathbf{R}}^s \right) = \sum_s \sum_{t,n} D_\varphi \left(\mathbf{r}^s_{t,n}, \hat{\mathbf{r}}^s_{t.n} \right) \tag{3.9}$$

We identified strictly convex functions corresponding to Poisson, binomial and negative binomial distributions which result in Bregman divergences such that minimizing the respective Bregman divergence is equivalent to the maximum likelihood estimation of the mean of the noise distribution.

For the Poisson distribution, $\varphi(z)$ is already known (Dhillon and Sra, 2005).

$$\varphi_{poiss} \left(z \right) = z \log z \tag{3.10}$$

For the binomial distribution,

$$\varphi_{bin} \left(z \right) = z \log z + \left(n - z \right) \log \left(n - z \right) \tag{3.11}$$

Here, n is the model parameter that describes the regularity of the spike counts and we assume that it is the same for all neurons and time bins. Spike counts become increasingly regular with the decrease in n.

We find that $f(x)=(n-z)\log(n-z)$ has a vertical asymptote at $z=n$ and is thus undefined for $z>n$. In other words, it is not possible to decrease n below the maximum spike count present in the data. This posed a practical constraint on increasing the regularity that the update rules could model. Thus to make $\varphi_{Bin}(z)$ continuous for $z\in\mathbb{R}_+$, we used an extension function $h(z)=\exp(z-n+1)-2z+2n-3$ for $z>n$. $h(z)$ was designed to ensure the continuity of Bregman divergence and update rules for binomial distribution at $z=n-1$. We note that there are other functions which could be used instead of $h(z)$ that would also satisfy these criteria. Thus the modified convex function for the binomial distribution is,

$$\varphi_{Bin}(z)=\begin{cases} z\log z+(n-z)\log(n-z) & z\le n-1 \\ z\log z+\exp(z-n+1)-2z+2n-3 & z>n-1 \end{cases} \tag{3.12}$$

Next, for the negative binomial distribution,

$$\varphi_{NBin}(z)=z\log z-(z+v)\log(z+v) \tag{3.13}$$

Using these functions, we obtained the following Bregman divergences corresponding to each noise model.

$$D_{\varphi,Poiss}(r_{t,n},\hat{r}_{t,n})=r_{t,n}\log\left(\frac{r_{t,n}}{\hat{r}_{t,n}}\right)-(r_{t,n}-\hat{r}_{t,n}) \tag{3.14}$$

$$D_{\varphi,Bin}(r_{t,n},\hat{r}_{t,n})=\begin{cases} r_{t,n}\log\left(\frac{r_{t,n}}{\hat{r}_{t,n}}\right)+(n-r_{t,n})\log\left(\frac{n-r_{t,n}}{n-\hat{r}_{t,n}}\right), r_{t,n}\le n-1,\hat{r}_{t,n}\le n-1 \\ r_{t,n}\log\left(\frac{r_{t,n}}{\hat{r}_{t,n}}\right)-(n-r_{t,n})\log(n-\hat{r}_{t,n})+\exp(r_{t,n}-n+1) \\ \qquad -2r_{t,n}+2n-3, r_{t,n}>n-1,\hat{r}_{t,n}\le n-1 \\ r_{t,n}\log\left(\frac{r_{t,n}}{\hat{r}_{t,n}}\right)+(n-r_{t,n})\log(n-r_{t,n})-(r_{t,n}-\hat{r}_{t,n}+1)\exp(\hat{r}_{t,n}-n+1) \\ \qquad +r_{t,n}+\hat{r}_{t,n}-2n+3, r_{t,n}\le n-1,\hat{r}_{t,n}>n-1 \\ r_{t,n}\log\left(\frac{r_{t,n}}{\hat{r}_{t,n}}\right)+\exp(r_{t,n}-n+1)-(r_{t,n}+\hat{r}_{t,n}+1)\exp(\hat{r}_{t,n}-n+1) \\ \qquad -(r_{t,n}-\hat{r}_{t,n}), r_{t,n}>n-1,\hat{r}_{t,n}>n-1 \end{cases} \tag{3.15}$$

$$D_{\varphi,NBin}(r_{t,n},\hat{r}_{t,n})=r_{t,n}\log\left(\frac{r_{t,n}}{\hat{r}_{t,n}}\right)-(r_{t,n}+v)\log\left(\frac{r_{t,n}+v}{\hat{r}_{t,n}+v}\right) \tag{3.16}$$

When noise in the data is Poisson distributed, minimizing the Bregman divergence in equation 3.15 is equivalent to the maximum likelihood estimation of the mean of the respective noise model. This has already been proven for the Poisson distribution (Cemgil, 2009). We proved that the same also hold for the original binomial distribution and the negative binomial distribution. These proofs are given in Appendix 1.

3.3.5 Update rules for space-by-time NMF that model neural variability

Space-by-time NMF updates the module matrices $\mathbf{W}_{spa}$ and $\mathbf{W}_{tem}$ using two-factor multiplicative update rules as described in Algorithm 1 of chapter 2. (Dhillon and Sra, 2005). derived the following generalization framework, which we use to derive the update rules for module matrices. For a two-factor decomposition $\mathbf{R} \approx \mathbf{XY}$, the gradients of the Bregman divergence $D_{\varphi}(\mathbf{R}, \mathbf{XY})$ with respect to $\mathbf{X}$ and $\mathbf{Y}$ are,

$$\nabla_{\mathbf{X}} D_{\varphi}(\mathbf{R}, \mathbf{XY}) = \left(\varsigma(\mathbf{XY}) \odot (\mathbf{XY} - \mathbf{R})\right)\mathbf{Y}^T \tag{3.17}$$

$$\nabla_{\mathbf{Y}} D_{\varphi}(\mathbf{R}, \mathbf{XY}) = \mathbf{X}^T\left(\varsigma(\mathbf{XY}) \odot (\mathbf{XY} - \mathbf{R})\right) \tag{3.18}$$

where $\zeta(z)$ is the second derivative of $\varphi(z)$ and $\odot$ denotes the Hadamard product (element-wise multiplication). From Karush-Khun-Tucker (KKT) conditions,

$$\nabla_{\mathbf{X}} D_{\varphi}(\mathbf{R}, \mathbf{XY}) = \mathbf{\Lambda} \tag{3.19}$$

$$\nabla_{\mathbf{Y}} D_{\varphi}(\mathbf{R}, \mathbf{XY}) = \mathbf{\Omega} \tag{3.20}$$

$$\lambda_{t,k} x_{t,k} = \omega_{k,n} y_{k,n} = 0 \tag{3.21}$$

where $\mathbf{\Lambda} \geq 0$ and $\mathbf{\Omega} \geq 0$ are Lagrange multiplier matrices. Then Equations 3.17, 3.19 and 3.21 give, $\left[\left(\varsigma(\mathbf{XY}) \odot (\mathbf{XY} - \mathbf{R})\right)\mathbf{Y}^T\right]_{t,k} x_{t,k} = \lambda_{t,k} x_{t,k} = 0$, leading to,

$$X_{n,t}^{k+1} \leftarrow X_{n,t}^{k} \frac{\left[\left(\varsigma\left(\mathbf{X}^k\mathbf{Y}^k\right) \odot \mathbf{R}\right)\left(\mathbf{Y}^k\right)^T\right]_{n,t}}{\left[\left(\varsigma\left(\mathbf{X}^k\mathbf{Y}^k\right) \odot \mathbf{X}^k\mathbf{Y}^k\right)\left(\mathbf{Y}^k\right)^T\right]_{n,t}} \tag{3.22}$$

Similarly, Equations 3.18, 3.20 and 3.21 give $\left[\mathbf{X}^T\left(\varsigma(\mathbf{XY}) \odot (\mathbf{XY} - \mathbf{R})\right)\right]_{k,n} y_{k,n} = \omega_{k,n} y_{k,n} = 0$, leading to,

$$Y_{n,t}^{k+1} \leftarrow Y_{n,t}^{k} \frac{\left[\left(\mathbf{X}^k\right)^T\left(\varsigma\left(\mathbf{X}^k\mathbf{Y}^k\right) \odot \mathbf{R}\right)\right]_{n,t}}{\left[\left(\mathbf{X}^k\right)^T\left(\varsigma\left(\mathbf{X}^k\mathbf{Y}^k\right) \odot \mathbf{X}^k\mathbf{Y}^k\right)\right]_{n,t}} \tag{3.23}$$

The two factor NMF update rule for Poisson distribution is already available (Lee and Seung, 1999), which we use for module matrices in space-by-time NMF update rules corresponding to the Poisson distribution. We derived the update rules for module matrices for binomial and negative binomial distributions using Equations 3.22 and 3.23 and the functions $\varphi_{Bin}(z)$ and $\varphi_{NBin}(z)$ given in Equations 3.12 and 3.13.

Space-by-time NMF updates the coefficient matrix $\mathbf{H}^s$ for each trial s using tri-factorization detailed in Algorithm 1 of chapter 2. We use the multi-factor update rules proposed by (Dhillon and Sra, 2005) to derive the update rules for coefficient matrix. For tri-factor decomposition $\mathbf{R}^s \approx \mathbf{W}_{tem}\mathbf{H}^s\mathbf{W}_{spa}$, the update rule for $\mathbf{H}^s$ is derived in a similar way as two factor decomposition by minimizing the Bregman divergence $D_{\varphi}\left(\mathbf{R}^s, \mathbf{W}_{tem}\mathbf{H}^s\mathbf{W}_{spa}\right)$ under the constraint $\mathbf{H}^s \geq 0$. This results in the following update rule for the coefficient matrix.

$$\left(\mathbf{H}^s_{i,j}\right)^{k+1} \leftarrow \left(\mathbf{H}^s_{i,j}\right)^{k} \frac{\left(\left(\mathbf{W}^{k+1}_{tem}\right)^T\left(\zeta\left(\mathbf{W}^{k+1}_{tem}\left(\mathbf{H}^s\right)^k\mathbf{W}^{k+1}_{spa}\right)\odot\left(\mathbf{R}^s\right)^k\right)\left(\mathbf{W}^{k+1}_{spa}\right)^T\right)_{i,j}}{\left(\left(\mathbf{W}^{k+1}_{tem}\right)^T\left(\zeta\left(\mathbf{W}^{k+1}_{tem}\left(\mathbf{H}^s\right)^k\mathbf{W}^{k+1}_{spa}\right)\odot\left(\mathbf{W}^{k+1}_{tem}\left(\mathbf{H}^s\right)^k\mathbf{W}^{k+1}_{spa}\right)\right)\left(\mathbf{W}^{k+1}_{spa}\right)^T\right)_{i,j}} \quad (3.24)$$

We uses this to derive update rules for the coefficient matrix update for Poisson, binomial and negative binomial distributions. All update rules for updating matrices in our neww space-by-time NMF algorithms are given in Table 3.1. A generalized version of the complete space-by-time NMF algorithm is detailed in Algorithm 2.

In the following sections we refer to the existing update rules based on the Gaussian noise model as Gaussian update rules. Similarly, we refer to the new update rules as Poisson update rules, binomial update rules and negative binomial update rules.

Algorithm 2: Extended space-by-time NMF algorithm

Initialize $\mathbf{W}_{tem}^{1} > \mathbf{0}$, $\left(\mathbf{H}^{'}\right)^{1} = \left[\left(\mathbf{H}^{1}\right)^{1}\left(\mathbf{H}^{2}\right)^{1}\ldots\left(\mathbf{H}^{S}\right)^{1}\right] > \mathbf{0}$, and $\mathbf{W}_{spa}^{1} > 0$.

Iterate

1. Given $\mathbf{H}$ and $\mathbf{W}_{tem}^{k}$, update $\mathbf{W}_{spa}^{k}$

   a. Formulate $\left(\mathbf{G}^{'}\right)^{k} = \mathbf{W}_{tem}^{k}\left(\mathbf{H}^{'}\right)^{k}$

   b. Obtain the block transpose $\mathbf{G}^{k}$

   c. Update $\mathbf{W}_{spa}^{k}$ to improve the approximation $\mathbf{R} \approx \mathbf{G}^{k}\mathbf{W}_{spa}^{k}$

$$\mathbf{W}_{spa,i,j}^{k+1} \leftarrow \mathbf{W}_{spa,i,j}^{k} \frac{\left(\left(\mathbf{G}^{k}\right)^{T}\left(\zeta\left(\mathbf{G}^{k}\mathbf{W}_{spa}^{k}\right)\odot\mathbf{R}\right)\right)_{i,j}}{\left(\left(\mathbf{G}^{k}\right)^{T}\left(\zeta\left(\mathbf{G}^{k}\mathbf{W}_{spa}^{k}\right)\odot\mathbf{G}^{k}\mathbf{W}_{spa}^{k}\right)\right)_{i,j}}, \forall i \in [1\ldots L], j \in [1\ldots N]$$

2. Given $\mathbf{H}'$ and $\mathbf{W}_{spa}^{k+1}$, update $\mathbf{W}_{tem}^{k}$

   a. Get the block transpose $\mathbf{H}^{k}$

   b. Calculate $\mathbf{P}^{k} = \mathbf{H}^{k}\mathbf{W}_{spa}^{k+1}$

   c. Obtain the block transpose $\left(\mathbf{P}^{k}\right)^{'}$

   d. Update $\mathbf{W}_{tem}^{k}$ to improve the approximation $\mathbf{R}^{'} \approx \mathbf{W}_{tem}^{k}\left(\mathbf{P}^{k}\right)^{'}$

$$\mathbf{W}_{tem,i,j}^{k+1} \leftarrow \mathbf{W}_{tem,i,j}^{k} \frac{\left(\left(\varsigma\left(\mathbf{W}_{tem}^{k}\left(\mathbf{P}^{k}\right)^{'}\right)\odot\mathbf{R}^{'}\right)\mathbf{P}^{k}\right)_{i,j}}{\left(\left(\varsigma\left(\mathbf{W}_{tem}^{k}\left(\mathbf{P}^{k}\right)^{'}\right)\odot\mathbf{W}_{tem}^{k}\left(\mathbf{P}^{k}\right)^{'}\right)\mathbf{P}^{k}\right)_{i,j}} \forall i \in [1\ldots T], j \in [1\ldots P]$$

3. Given $\mathbf{W}_{tem}^{k+1}$ and $\mathbf{W}_{spa}^{k+1}$, update $\mathbf{H}$

   Update $\left(\mathbf{H}^{s}\right)^{k}, \forall s \in [1\ldots S]$ to improve the approximation $\mathbf{R}^{s} \approx \mathbf{W}_{tem}^{k+1}\left(\mathbf{H}^{s}\right)^{k}\mathbf{W}_{spa}^{k+1}$

$$\left(\mathbf{H}_{i,j}^{s}\right)^{k+1} \leftarrow \left(\mathbf{H}_{i,j}^{s}\right)^{k} \frac{\left(\left(\mathbf{W}_{tem}^{k+1}\right)^{T}\left(\zeta\left(\mathbf{W}_{tem}^{k+1}\left(\mathbf{H}^{s}\right)^{k}\mathbf{W}_{spa}^{k+1}\right)\odot\mathbf{R}^{s}\right)\left(\mathbf{W}_{spa}^{k+1}\right)^{T}\right)_{i,j}}{\left(\left(\mathbf{W}_{tem}^{k+1}\right)^{T}\left(\zeta\left(\mathbf{W}_{tem}^{k+1}\left(\mathbf{H}^{s}\right)^{k}\mathbf{W}_{spa}^{k+1}\right)\odot\left(\mathbf{W}_{tem}^{k+1}\left(\mathbf{H}^{s}\right)^{k}\mathbf{W}_{spa}^{k+1}\right)\right)\left(\mathbf{W}_{spa}^{k+1}\right)^{T}\right)_{i,j}}$$

$$\forall i \in [1\ldots P], j \in [1\ldots L]$$

4. Estimate the dissimilarity E^k between dataset $\mathbf{R}$ and the total reconstruction $\sum_s \mathbf{W}_{tem}^{k+1}\left(\mathbf{H}^s\right)^k \mathbf{W}_{spa}^{k+1}$ using the respective Bregman divergence $\sum_s D_\varphi\left(\mathbf{R}^s, \mathbf{W}_{tem}\mathbf{H}^s\mathbf{W}_{spa}\right)$.

until $E^{k+1} - E^k < E_{tolarence}$ or $k+1 >$ max number of iterations

Here $\zeta(z)$ is the second derivative of $\varphi(z)$ and $\odot$ denotes the Hadamard product (element-wise multiplication).

Table 3.1: Update rules derived for Poisson, binomial and negative binomial distributions. Update rules for Gaussian distribution are also shown for comparison. Binomial and negative binomial update rules include model parameters n and v that specify the dispersion of the spike counts. Matrices use the same notation as in Algorithm 2. [1] is a matrix of ones with the respective dimensionality. $\odot$ denotes Hadamard product (element-wise multiplication). All divisions are element-wise divisions.

Data distribution	Spatial module update	Temporal module update	Coefficient matrix update
Gaussian	$\mathbf{W}_{spa}^{k+1} \leftarrow \mathbf{W}_{spa}^{k} \odot \frac{\left(\mathbf{G}^{k}\right)^{T}\mathbf{R}}{\left(\mathbf{G}^{k}\right)^{T}\left(\mathbf{G}^{k}\mathbf{W}_{spa}^{k}\right)}$	$\mathbf{W}_{tem}^{k+1} \leftarrow \mathbf{W}_{tem}^{k} \odot \frac{\mathbf{R}'\mathbf{P}^{k}}{\left(\mathbf{W}_{tem}^{k}\left(\mathbf{P}^{k}\right)'\right)\mathbf{P}^{k}}$	$\left(\mathbf{H}^{s}\right)^{k+1} \leftarrow \left(\mathbf{H}^{s}\right)^{k} \odot \frac{\left(\mathbf{W}_{tem}^{k+1}\right)^{T}\mathbf{R}^{s}\left(\mathbf{W}_{spa}^{k+1}\right)^{T}}{\left(\mathbf{W}_{tem}^{k+1}\right)^{T}\mathbf{W}_{tem}^{k+1}\left(\mathbf{H}^{s}\right)^{k}\mathbf{W}_{tem}^{k+1}\left(\mathbf{W}_{spa}^{k+1}\right)^{T}}$
Poisson	$\mathbf{W}_{spa}^{k+1} \leftarrow \mathbf{W}_{spa}^{k} \odot \frac{\left(\mathbf{G}^{k}\right)^{T}\left(\frac{\mathbf{R}}{\mathbf{G}^{k}\mathbf{W}_{spa}^{k}}\right)}{\left(\mathbf{G}^{k}\right)^{T}[\mathbf{1}]}$	$\mathbf{W}_{tem}^{k+1} \leftarrow \mathbf{W}_{tem}^{k} \odot \frac{\left(\frac{\mathbf{R}'}{\mathbf{W}_{tem}^{k}\left(\mathbf{P}^{k}\right)'}\right)\mathbf{P}^{k}}{[\mathbf{1}]\mathbf{P}^{k}}$	$\left(\mathbf{H}^{s}\right)^{k+1} \leftarrow \left(\mathbf{H}^{s}\right)^{k} \odot \frac{\left(\mathbf{W}_{tem}^{k+1}\right)^{T}\left[\left(\frac{1}{\mathbf{W}_{tem}^{k+1}\left(\mathbf{H}^{s}\right)^{k}\mathbf{W}_{spa}^{k+1}}\right)\odot\mathbf{R}^{s}\right]\left(\mathbf{W}_{spa}^{k+1}\right)^{T}}{\left(\mathbf{W}_{tem}^{k+1}\right)^{T}[\mathbf{1}]\left(\mathbf{W}_{spa}^{k+1}\right)^{T}}$
Binomial (original)	$\mathbf{W}_{spa}^{k+1} \leftarrow \mathbf{W}_{spa}^{k} \odot \frac{\left(\mathbf{G}^{k}\right)^{T}\left(\frac{n\mathbf{R}}{\mathbf{G}^{k}\mathbf{W}_{spa}^{k} \odot \left(n-\mathbf{G}^{k}\mathbf{W}_{spa}^{k}\right)}\right)}{\left(\mathbf{G}^{k}\right)^{T}\left(\frac{n}{\left(n-\mathbf{G}^{k}\mathbf{W}_{spa}^{k}\right)}\right)}$	$\mathbf{W}_{tem}^{k+1} \leftarrow \mathbf{W}_{tem}^{k} \odot \frac{\left(\frac{n\mathbf{R}'}{\mathbf{W}_{tem}^{k}\left(\mathbf{P}^{k}\right)' \odot \left(n-\mathbf{W}_{tem}^{k}\left(\mathbf{P}^{k}\right)'\right)}\right)\mathbf{P}^{k}}{\left(\frac{n}{\left(n-\mathbf{W}_{tem}^{k}\left(\mathbf{P}^{k}\right)'\right)}\right)\mathbf{P}^{k}}$	$\left(\mathbf{H}^{s}\right)^{k+1} \leftarrow \left(\mathbf{H}^{s}\right)^{k} \odot \frac{\left(\mathbf{W}_{tem}^{k+1}\right)^{T}\left[\left(\frac{n}{\mathbf{W}_{tem}^{k+1}\left(\mathbf{H}^{s}\right)^{k}\mathbf{W}_{spa}^{k+1} \odot \left(n-\mathbf{W}_{tem}^{k+1}\left(\mathbf{H}^{s}\right)^{k}\mathbf{W}_{spa}^{k+1}\right)}\right)\odot\mathbf{R}^{s}\right]\left(\mathbf{W}_{spa}^{k+1}\right)^{T}}{\left(\mathbf{W}_{tem}^{k+1}\right)^{T}\left(\frac{n}{\left(n-\mathbf{W}_{tem}^{k+1}\left(\mathbf{H}^{s}\right)^{k}\mathbf{W}_{spa}^{k+1}\right)}\right)\left(\mathbf{W}_{spa}^{k+1}\right)^{T}}$
Negative binomial	$\mathbf{W}_{spa}^{k+1} \leftarrow \mathbf{W}_{spa}^{k} \odot \frac{\left(\mathbf{G}^{k}\right)^{T}\left(\frac{v\mathbf{R}}{\mathbf{G}^{k}\mathbf{W}_{spa}^{k} \odot \left(v+\mathbf{G}^{k}\mathbf{W}_{spa}^{k}\right)}\right)}{\left(\mathbf{G}^{k}\right)^{T}\left(\frac{v}{\left(v+\mathbf{G}^{k}\mathbf{W}_{spa}^{k}\right)}\right)}$	$\mathbf{W}_{tem}^{k+1} \leftarrow \mathbf{W}_{tem}^{k} \odot \frac{\left(\frac{v\mathbf{R}'}{\mathbf{W}_{tem}^{k}\left(\mathbf{P}^{k}\right)' \odot \left(v+\mathbf{W}_{tem}^{k}\left(\mathbf{P}^{k}\right)'\right)}\right)\mathbf{P}^{k}}{\left(\frac{v}{\left(v+\mathbf{W}_{tem}^{k}\left(\mathbf{P}^{k}\right)'\right)}\right)\mathbf{P}^{k}}$	$\left(\mathbf{H}^{s}\right)^{k+1} \leftarrow \left(\mathbf{H}^{s}\right)^{k} \odot \frac{\left(\mathbf{W}_{tem}^{k+1}\right)^{T}\left[\left(\frac{v}{\mathbf{W}_{tem}^{k+1}\left(\mathbf{H}^{s}\right)^{k}\mathbf{W}_{spa}^{k+1} \odot \left(v+\mathbf{W}_{tem}^{k+1}\left(\mathbf{H}^{s}\right)^{k}\mathbf{W}_{spa}^{k+1}\right)}\right)\odot\mathbf{R}^{s}\right]\left(\mathbf{W}_{spa}^{k+1}\right)^{T}}{\left(\mathbf{W}_{tem}^{k+1}\right)^{T}\left(\frac{v}{\left(v+\mathbf{W}_{tem}^{k+1}\left(\mathbf{H}^{s}\right)^{k}\mathbf{W}_{spa}^{k+1}\right)}\right)\left(\mathbf{W}_{spa}^{k+1}\right)^{T}}$

3.4 Performance evaluation

We verified the performance of our new algorithms methodically. We started by generating spike counts from statistical models and generated idealistic datasets. Then we proceeded to generate more realistic simulations using a conductance based integrate-and-fire network. Finally, we applied the new algorithms to experimental data from the auditory cortex, recorded in response to long tones and click sequences, which were analyzed using Gaussian update rules in chapter 2 section 2.5.1 and chapter 2 section 2.5.2.

3.4.1 Statistical simulations

Now we explain in detail how we evaluated the performance of the new update rules using statistical simulations. First, we describe how we generated the dataset we used for the evaluation. Then we describe different ways we used to estimate the model parameters for binomial and negative binomial models. Finally we present the results of the statistical simulations.

Artificial dataset generation

We began evaluating our new algorithms using idealistic artificial datasets that had a predefined low dimensional modularity, but were representative of neural responses. We created these datasets using the low-dimensional structure present in the responses recorded from retinal ganglion cells to natural images. The details of the retinal ganglion cell dataset we used are fully described in section 2.5.3.

Figure 3.2A shows an illustration that describes the generation of the datasets. We binned the spikes into 100 ms time bins. We generated 25 artificial datasets from each recording by sampling at random responses of 30 neurons to 30 images over 30 trials and extracted the underlying module structure in the dataset using Gaussian update rules. Each dataset was reconstructed using the mean activation coefficient for the trials. Thus, this reconstruction is free from trial-to-trial variability. Trial-to-trial variability was added using two methods. In the first approach, we sampled spike counts directly from the respective probability distribution. In the second approach, we generated spike counts from spike trains that had inter-spike intervals (ISIs) distributed according to a gamma process. ISIs of cortical spike trains are often modeled using gamma processes (Dayan and Abbott, 2001; Maimon and Assad, 2009). An ISI x following a gamma process is given by, $p(x;k,\theta) = \frac{x^{k-1}}{\theta^k \Gamma(k)} e^{-x/\theta}$ where k is the shape parameter and θ is the scale parameter of the distribution. The shape parameter controls the level of dispersion in the spike counts. Poisson variability is modeled by setting k to one, while sub-Poisson and supra-Poisson variability are obtained using $k > 1$ and $k < 1$ respectively. This second series of datasets that had gamma distributed ISIs were closer to spike trains recorded from real cortical neurons, while offering the ability to control the dispersion in the spike train. The shape parameter was varied between 1.5 - 20 to generate sub-Poisson spike counts and in the range 0.1 - 0.9 for supra-Poisson spike counts.

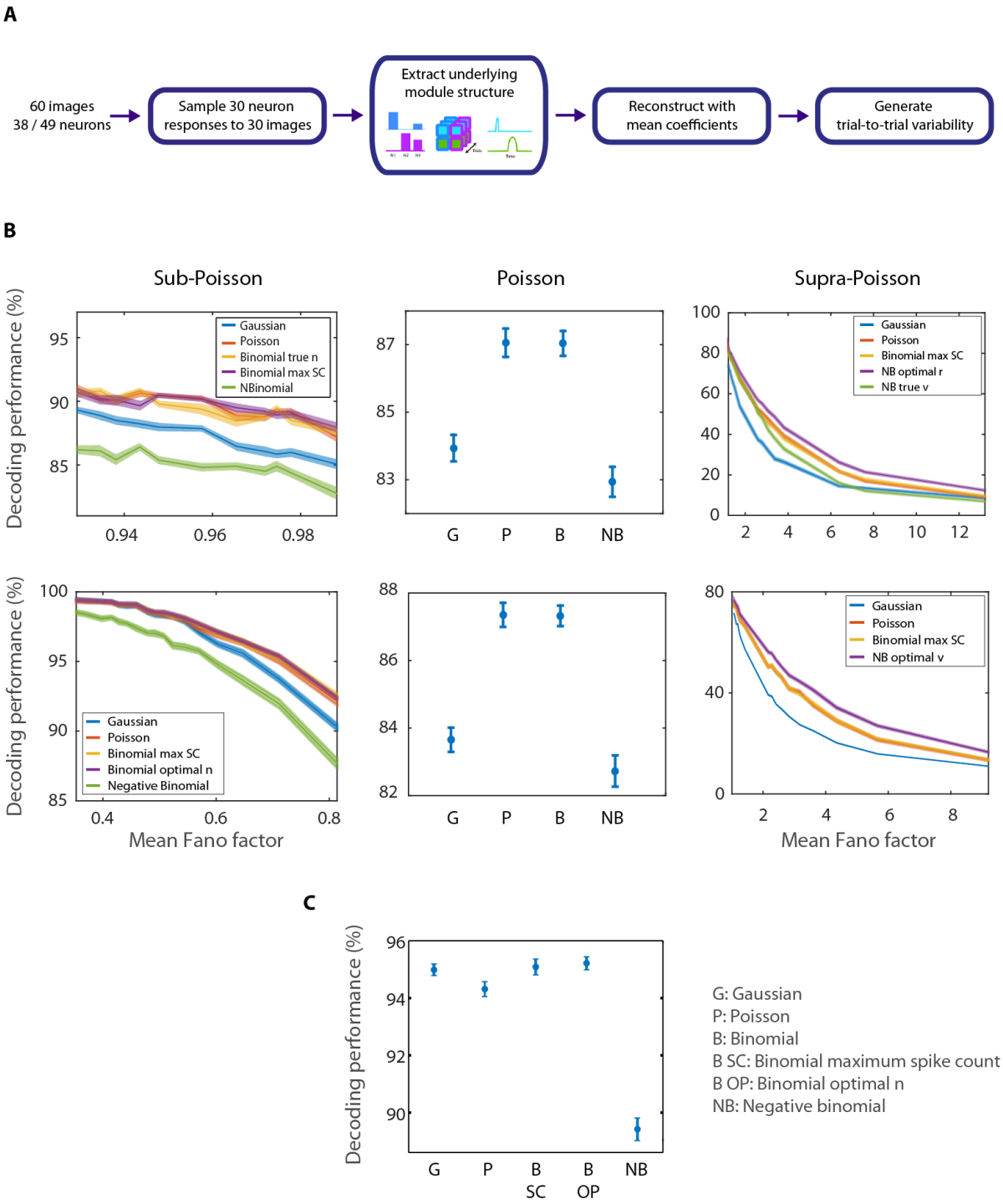

Figure 3.2: Performance evaluation of new update rules using statistical simulations. A: Generation of idealistic datasets using the neural responses of retinal ganglion cells to natural images. B: Decoding performance of space-by-time NMF update rules when the trial-to-trial variability is sub-Poisson (left), Poisson (middle) and supra-Poisson (right). The top row shows the decoding performance when the noise was generated by sampling from the respective probability distribution. The bottom row shows the decoding performance when spike counts were generated using spike trains that had gamma distributed inter-spike intervals and the variability of spike counts was changed using the shape parameter of the gamma distribution. (shadings and the error bars show s.e.m.)

Parameter estimation

When applying binomial and negative binomial rules, an additional model parameters n or ν needs to be specified as an input to the respective space-by-time NMF algorithm. When the data is directly sampled from the corresponding probability distribution, the ground truth value of the model parameter is known. However, when spike counts are generated from spike trains with Gamma distributed ISIs and for emperical data, we require a method to estimate them from the data. We investigated several methods of estimation and verified the performance of the respective update rule for each method.

In the first approach, we estimated n and ν based on the average Fano factor of the data matrix FF_{avg} and the average mean spike count of the full data matrix λ_{avg}.

$$n = \frac{\lambda_{avg}}{1 - FF_{avg}} \tag{3.25}$$

$$\nu = \frac{\lambda_{avg}}{FF_{avg} - 1} \tag{3.26}$$

We calculated the mean Fano factor for each dataset by first estimating the Fano factor for each neuron, time bin and stimulus and then calculating the average across them. The mean spike count was obtained in a similar way.

In the second approach, we ran a full parameter scan for each of the parameters in binomial and negative binomial rules. We selected the parameter value as the value that yielded the best decoding performance. Finally, for binomial rules, we used the maximum spike count in the data as the value of n (since theoretically the spike count cannot exceed n under the binomial model).

Decoding

We aimed to evaluate whether optimizing space-by-time NMF to model the underlying variability would yield a low-dimensional structure to improve stimulus encoding. To achieve this, we performed a cross-validated decoding analysis using linear discriminant analysis (LDA). The decoding procedure was implemented as described in section 2.5.1.

Results

We generated artificial datasets based on the low-dimensional structure in the responses of retinal ganglion cells for natural images. We varied the trial-to-trial variability in these datasets using two methods; first by directly sampling the variability from the underlying noise distribution and second using spike counts generated from spike trains that had gamma distributed ISIs. We used three methods to estimate the model parameters in binomial and negative binomial rules when spike counts were generated from spike trains; 1) by using the average Fano factor and the average spike count across each neuron, time bin and stimulus, 2) using a parameter scan across a range of parameters, and 3) using the maximum spike count in the dataset as the model parameter in the binomial distribution. We found that the first method yielded poor results (not shown). As shown in Figure 3.2, the performance of binomial rules when the maximum spike count was used as the model parameter was comparable to the performance obtained from the parameter scan.

We found that the decoding performance was qualitatively very similar when we sampled the variability in the datasets directly from the probability distribution as well as when we controlled it using ISIs (Figure 3.2B). In general, the performance decreased when the trial-to-trial variability changed from sub-Poisson to supra-Poisson. The update rule corresponding to the underlying data distribution gave a higher performance compared to the existing Gaussian update rules in all cases except when the spike counts were directly sampled from a Poisson distribution and when the data was sub-Poisson and had gamma distributed inter-spike-intervals (Wilcoxon rank sum test, $p<0.01$). This performance advantage was high when the data was supra-Poisson. This indicates that there is a gain in stimulus discriminating capability when using the optimal space-by-time NMF update rules.

As the Fano factor decreased and the data became reliable, the performance of the Gaussian rules approached and was equal to the performance of binomial and Poisson rules suggesting that when the variability in the spike counts is low, the exact noise model does not have a large impact on the performance of space-by-time rules. When the variability in the data was sub-Poisson, the performance of the negative binomial rules was always lower than that of the other rules (Wilcoxon rank sum test, $p<0.01$). This was expected because we set the model parameter ν in negative binomial rules to one (which would normally indicate high variability in the data) as a control condition. Finally, we applied the update rules on real datasets that we used to generate the artificial datasets (Figure 3.2C). In these datasets, the performance of Gaussian rules was similar to that of the binomial rules (Wilcoxon rank sum test, $p<0.01$). Since the mean Fano factor in these datasets were 0.78 ± 0.02 and 0.71 ± 0.03, this result confirms our findings from the simulations; when the data had higher reliability, the performance of the Gaussian rules was similar to Poisson update rules.

Interestingly, Poisson rules had similar performance for sub-Poisson variability as when using binomial rules. This is important for two reasons. First, Poisson rules have less computational complexity. Second, they do not require the evaluation of an extra model parameter. Both these reasons lead to less computational time for Poisson rules compared to binomial and negative binomial rules.

3.4.2 Network simulations

Next, we used a network of conductance-based integrate-and-fire neurons to investigate the performance difference between Gaussian and Poisson rules in detail. We modified the network model used in (Mazzoni et al., 2008; Cavallari et al., 2014). This network is able to generate a range of experimentally observed modulatory effects between spike rate, local field potentials and electroencephalographs (Mazzoni et al., 2008; Mazzoni et al., 2010). We used the modified network to simulate responses of auditory cortical neurons to tones.

A graphical representation of the network is given in Figure 3.3A. It contains 4000 excitatory neurons with AMPA like synapses and 1000 inhibitory neurons with GABA like synapses. Both types of neurons were modeled as leaky integrate and fire neurons (Tuckwell, 1988) similar to our previous work (Mazzoni et al., 2008; Mazzoni et al., 2010; Cavallari et al., 2014). The sub-threshold dynamics of the membrane potential of the i^{th} neuron $V_i(t)$ is modeled as,

$$\tau_m \frac{dV_i(t)}{dt} = -V_i(t) + V_{leak} - \frac{1}{g_{leak}}\left(I_{i,AMPA_{rec}}(t) + I_{i,GABA_{rec}}(t) + I_{i,AMPA_{ext}}(t)\right) \quad (3.27)$$

where τ_m is the membrane constant (20 ms and 10 ms for excitatory and inhibitory neurons respectively), V_{leak} is the leak membrane potential (set to -70 mV), g_{leak} is the leak membrane conductance (25 nS and 20 nS for excitatory and inhibitory neurons respectively) (Brunel and Wang, 2003) and $I_{i,AMPA_{rec}}(t)$, $I_{i,GABA_{rec}}(t)$ and $I_{i,AMPA_{ext}}(t)$ are the synaptic input currents from recurrent AMPA, recurrent GABA and external AMPA synapses respectively. The neuron discharges a spike when the membrane potential crosses a threshold $V_{threshold}$, -52 mV. The membrane voltage is then set to V_{reset}, -59 mV for an absolute refractory period of 2 ms for excitatory neurons and 1 ms for inhibitory neurons (Brunel and Wang, 2003). Each of the synaptic input currents $I_{i,AMPA_{rec}}(t)$, $I_{i,GABA_{rec}}(t)$ and $I_{i,AMPA_{ext}}(t)$ are modeled as,

$$I_{syn} = g_{syn} s_{syn}(t)\left(V(t) - V_{syn}\right) \quad (3.28)$$

g_{syn} is the conductance of the synapse while V_{syn} is the reversal potential of the synapse. Values of these parameters are given in Table 3.1.

Synaptic kinetics of the post-synaptic neuron receiving a spike discharged at time t^* are modeled by the function $s_{syn}(t)$ for both models using a delayed difference of exponentials given by (Brunel and Wang, 2003),

$$\Delta s_{syn}(t) = \frac{\tau_m}{\tau_d - \tau_r}\left[\exp\left(-\frac{t - \tau_l - t^*}{\tau_d}\right) - \exp\left(-\frac{t - \tau_l - t^*}{\tau_r}\right)\right] \quad (3.29)$$

τ_l is the latency of the post-synaptic currents, τ_r and τ_d are the rise time and decay times of the synapse. These parameters were set to the values given in Table 3.2.

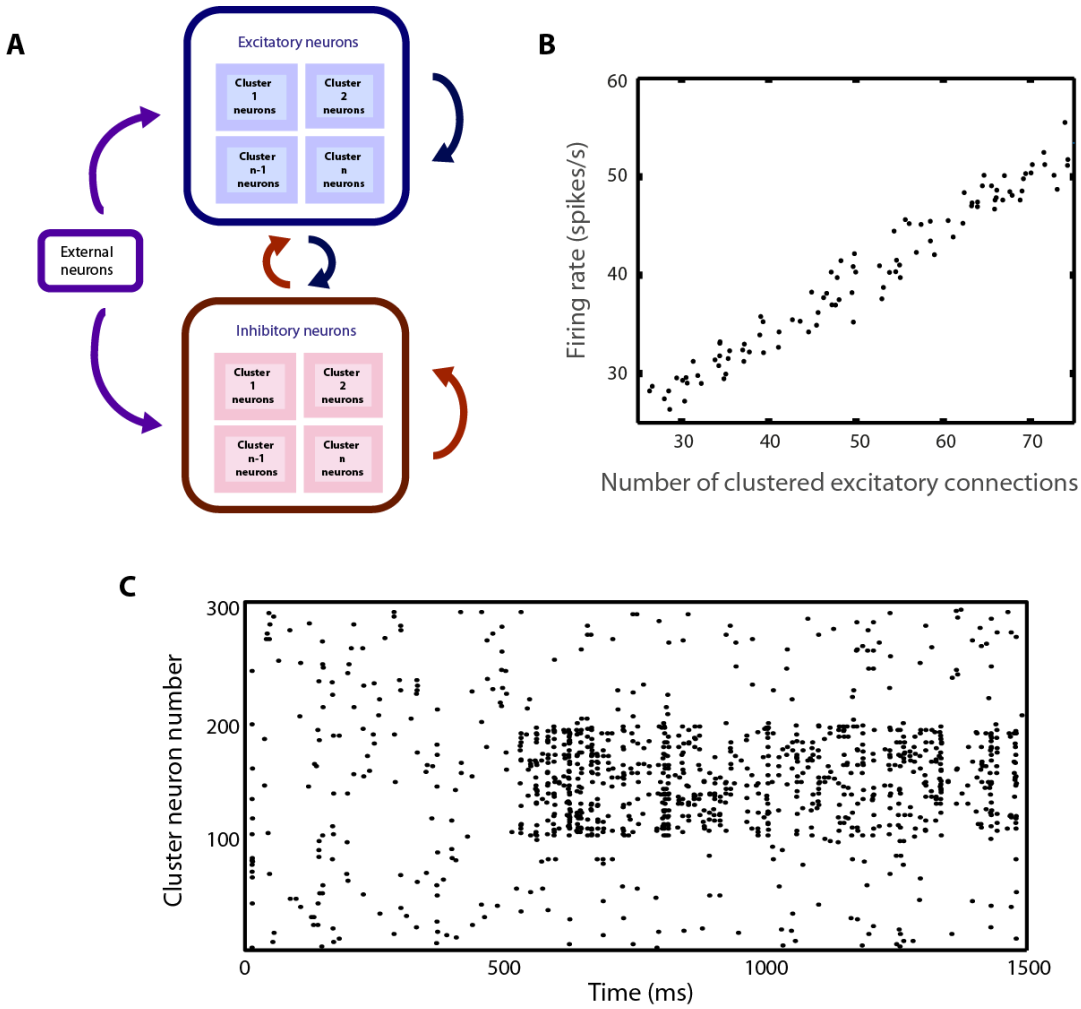

Figure 3.3: Conductance-based integrate and fire network simulations. A: An illustration of the network structure. B: Firing rate variation with the number of clustered excitatory inputs to a neuron C: An example response generated from the network.

Table 3. 2: Synaptic conductances and reversal potentials.

Synaptic conductances g_{syn} (nS)	
GABA on inhibitory	2.7
GABA on excitatory	2
$AMPA_{recurrent}$ on inhibitory	0.233
$AMPA_{recurrent}$ on excitatory	0.178
$AMPA_{external}$ on inhibitory	0.3
$AMPA_{external}$ on excitatory	0.234
Synaptic reversal potential V_{syn} (mV)	
V_{GABA}	-80
V_{AMPA}	0

Table 3. 3: Synaptic time constants.

	τ_r (ms)	τ_d (ms)	τ_l (ms)
GABA	0.25	5	1
AMPA on inhibitory	0.2	1	1
AMPA on excitatory	0.4	2	1

We formed 40 clusters of 100 excitatory neurons and 25 inhibitory neurons each in the original network. These clusters had the following properties: 1) a higher connection probability within two neurons in a cluster compared to the average connection probability between any two neurons in the network (Song et al., 2005), 2) a higher common external input to neurons in the same cluster compared to any neuron in the network (Yoshimura et al., 2005; Ko et al., 2011) 3) higher synaptic conductance between neurons in the same cluster (Song et al., 2005). The connection probability of neurons within a cluster is 0.4 while it is 0.2 on average between any two neurons in the network. We further generated firing rate variability between neurons in a cluster (Figure 3.3B) by varying the number of excitatory connections from one neuron of the cluster to the remaining neurons in the cluster while keeping the average connection probability between neurons in the cluster to 0.4. We obtained this variability by sampling the number of within cluster connections for each neuron from a truncated exponential distribution (truncated between 50% - 250% of the mean number of connections). The conductance of the clustered excitatory-excitatory synapses was 2.5 times that of unclustered excitatory-excitatory synapses and the conductance of the remaining clustered synapses was twice that of the respective unclustered synapses.

Feed-forward input to the network is provided by 40 external clusters. Each of these clusters has connectivity to one of the 40 network clusters. Each clustered neuron receive input from 50 external neurons randomly selected from a pool of 200 neurons that provide input to that particular cluster. All external neurons shared the same firing rate, but generated independent Poisson distributed spikes..

We used this network to simulate responses of A1 neurons to long tones. The spikes of individual neurons in the network were asynchronous, but the population firing displayed synchronous events (Figure 3.3C). This simulation did not model up-down activity states in the dataset we analyzed in section 2.5.1. We generated the stimulus using feed-forward thalamic inputs modeled through external excitatory cells to the network without implementing a detailed model of the network mechanisms such as co-tuning and lateral inhibition (de la Rocha et al., 2008; Levy and Reyes, 2011). Stimuli consisted of 30 long tones of 500 ms duration that were equally distributed along the log frequency axis. We modeled frequency selectivity of the clusters using Gaussian tuning curves. We varied the width of the tuning curves between 0.5 - 1.5 octaves. 24 clusters of the network formed eight groups, each with a different best frequency. These eight best frequencies were equally distributed along the log frequency axis spanning four octaves as shown in the top panel of Figure 3.4A. The three clusters with the same best frequency responded with different

temporal profiles shown in Figure 3.4A bottom panel; either phasic, phasic-tonic or tonic responses (Chimoto et al., 2002). Further two groups, each containing three clusters having phasic, phasic-tonic and tonic response profiles, fired for all stimuli and thus carried no information about tones. These have a role of outliers in the data. We modeled the phasic and phasic-tonic temporal profiles using a difference of exponentials of the form,

$$R(t) = \exp\left(-\frac{t-a}{2\tau_1}\right) - \exp\left(-\frac{t}{2\tau_1}\right) + b \qquad (3.30)$$

We set parameters a and b to 0.154 and -0.002 for phasic firing rate and to -9.974 and 0.049 for phasic-tonic firing rate. We set τ_1 to 500 ms for both cases. These firing rates had a maximum amplitude of 3 spikes/ms/cell. The tonic firing rate was set to 1.5 spikes/ms/cell and the background spontaneous firing rate was 4 spikes/ms/cell.

We generated the trial-to-trial variability in the network from two sources. The first is due to the independent realization of Poisson spike trains of the external neurons. The effect of this was small compared to the second source of variability modeled using an Ornstein-Uhlenbeck (OU) process $n(t)$.

$$\tau_n \frac{dn(t)}{dt} = -n(t) + \sigma_n\left(\sqrt{2\tau_n}\right)\eta(t) \qquad (3.31)$$

The time constant τ_n controls the cut-off frequency of the noise spectrum while σ_n is the standard deviation of the noise. τ_n was set to 16 ms while σ_n^2 was set to 0.16 spikes/ms as used in our previous studies.

The network was simulated using a finite difference integration scheme based on the second order Runge Kutta algorithm (Press et al., 1996) with a time step of 0.05 ms.

Analysis of data generated by the network simulations

We generated 40 samples of 85 neurons each by randomly sampling neurons from the 30 stimulated clusters and binned the responses into 100 ms time bins. The mean Fano factor of the dataset averaged over neurons and time bins was 1.006 ± 0.002. This dataset was analyzed using Gaussian and Poisson update rules. The number of temporal modules was set to two, the ground truth value, and the number of spatial modules was varied between 1 - 20.

Evaluation of clustering quality

We evaluated the clustering quality of the spatial modules with the sparseness measure used in (Hoyer, 2004).

$$sparseness(\mathbf{x}) = \frac{\sqrt{n} - \frac{\sum |\mathbf{x}_i|}{\sqrt{\sum \mathbf{x}_i^2}}}{\sqrt{n} - 1} \tag{3.32}$$

where $\mathbf{x}$ is the vector of NMF coefficients and n is the number of coefficients. This measure evaluates to one if and only if $\mathbf{x}$ consists of only one non-zero component and is zero if and only if all components are equal. It interpolates smoothly in between zero and one for other values.

Results

Figure 3.4B-D summarize the results from Gaussian and Poisson space-by-time NMF rules. When the tuning curves have a standard deviation of 0.5 octaves, the decoding performance reaches a plateau after seven spatial modules (Figure 3.4B). This occurs because there are only eight clusters with different tuning profiles although they differ in their temporal profiles. As the tuning curves become broader and the frequency bandwidth increases, the highest achievable decoding performance decreases. Compared to Gaussian rules, Poisson rules give a higher decoding performance of up to 5% when the standard deviation is 0.5 or 0.75 octaves and up to 3.2% when the standard deviation is 1 octave (Wilcoxon rank sum test, $p < 0.05$). This increase in performance occurs when the number of spatial modules increases beyond five clusters when the tuning curve standard deviation is 0.75 or 1 octaves and when the number of spatial modules is higher than 7 when the tuning curve standard deviation is 0.5 octaves. This decrease in the number of spatial modules required to obtain the highest performance could be attributed to the broadening of the tuning.

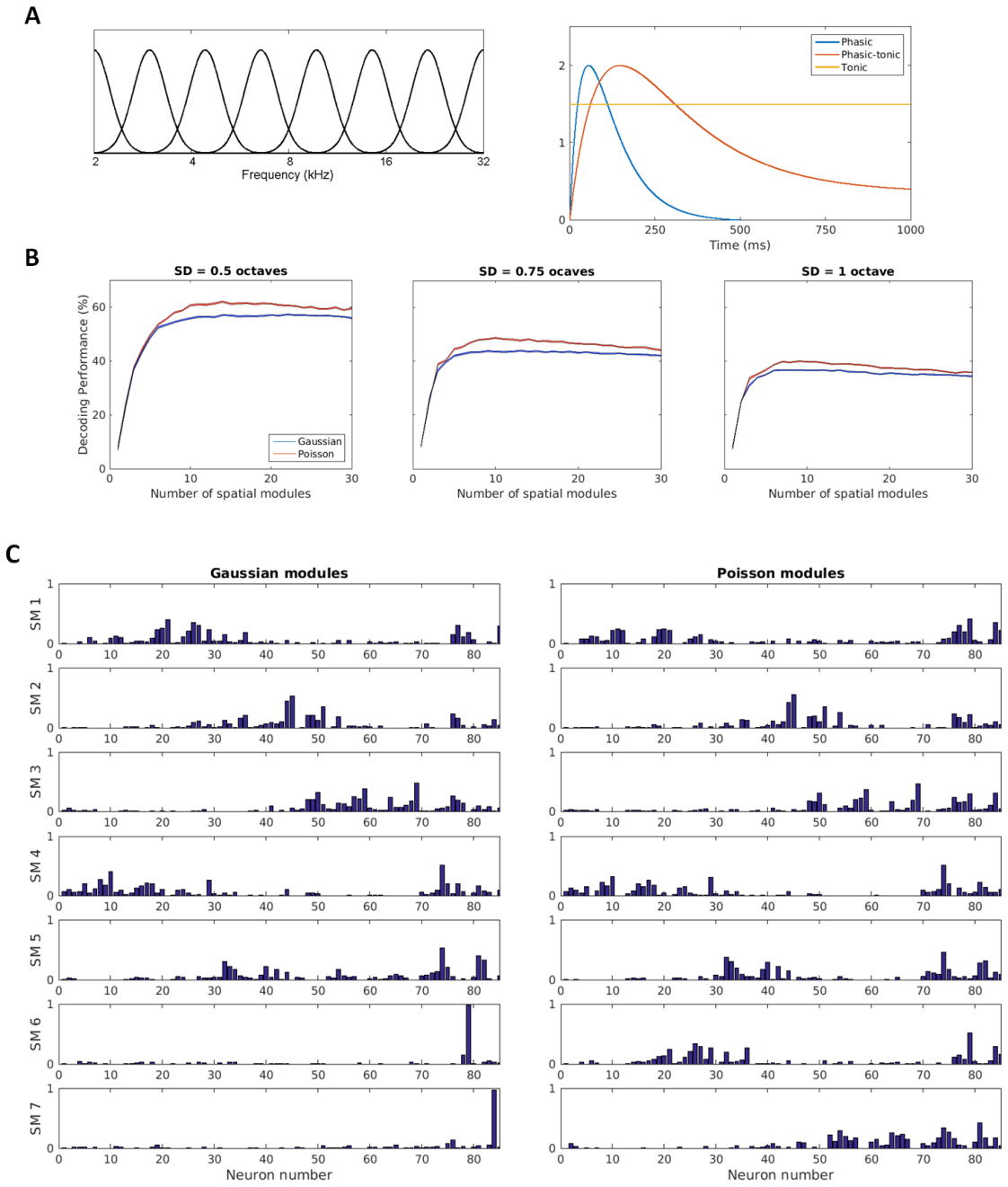

Figure 3.4: Simulation of network responses to long tones. A: Three clusters in the network shared the same frequency selectivity (left), but had different firing profiles (right). B: Decoding performance when the width of the tuning curves change. C: Comparison between the spatial modules (SMs) derived from Gaussian rules and Poisson rules.

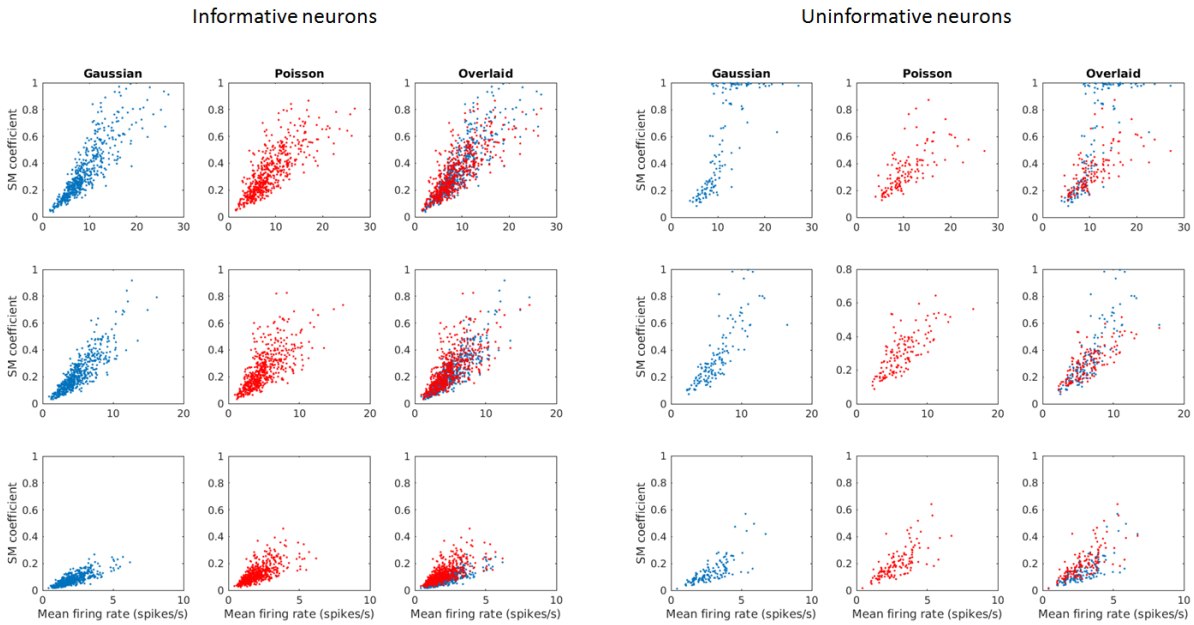

Figure 3.5: Variation of the weights of neurons in the spatial modules (SM) with the mean firing rates. Neurons were separated into informative (those that show frequency selectivity) and uninformative (those that fire to all tones). The first row shows how the weights allocated to each neuron in the spatial modules vary with the firing rate of the neuron for neurons that show tonic firing profiles. the second and the third rows are similar as the first row, but for neurons having phasic-tonic and phasic firing profiles.

There is a qualitative difference in the spatial modules derived from the two rules. As the number of spatial modules increases, Gaussian modules tend to assign a single spatial module with high weight to an uninformative neuron that responds to all tones. These neurons have a role of outliers in the data. An illustrative example of such decomposition is given in Figure 3.4C. In the example, neurons are sorted such that their best frequencies span the frequency axis from low frequencies to high frequencies. Neurons 70 - 85 respond to all stimuli. Modules derived using Gaussian rules have assigned the last two spatial modules for uninformative neurons that respond to all stimuli while Poisson modules have allocated the uninformative neurons to modules that modulate the activation coefficients for different tones. To quantify whether this biasness would lead to a quantifiable difference in the quality of spatial modules, we quantified the clustering quality using the sparseness measure proposed in (Hoyer, 2004). We found that the sparseness of the Gaussian spatial modules (0.69 ± 0.03) was slightly, but significantly higher than the sparseness in Poisson modules (0.63 ± 0.02, Wilcoxon rank sum test, $p < 0.01$).

In order to identify whether there is a relationship in the weights of neurons in spatial modules and their firing rate, we plot the weight against their maximum firing rate. We separated neurons in each sample into informative neurons and uninformative neurons (neurons that fire for all stimuli) and further classified them into three groups based on their

temporal firing profiles. The results in Figure 3.4D show that the relationship between the weight and the firing rate is qualitatively similar between the two rules for informative neurons. The weight assigned to the neuron increases linearly with an increase in the spread when the firing rate increases. However, for uninformative neurons, when the response pattern is either tonic or phasic-tonic, there is a supra-linear increase in the weights assigned by the Gaussian modules while the weights of the Poisson modules increase linearly. This suggests that Gaussian modules could be more biased towards selecting high firing neurons into individual spatial modules as compared with modules derived from Poisson rules.

3.4.3 Using new update rules to study population coding of sounds in rat auditory cortex

Data analysis

We compared the performance of new space-by-time NMF algorithms on the auditory neural responses to long tones and clicks. These datasets are fully described in section 2.5.1. We constructed the data matrix, performed space-by-time NMF analysis and decoding similar to the procedure described in that section. We evaluated the sparseness of the spatial modules using the measure from (Hoyer, 2004) that is described in section 3.4.2.

We performed an evaluation of the impact of the first spike information by removing the first spike from each trial in each neuron prior to applying space-by-time NMF update rules. In another analysis, noise correlations were removed from the data by shuffling trials of each neuron independently for each stimulus.

We tested two preprocessing methods on the performance of update rules. In the first method, we normalized the spike counts in each time bin with the logarithm of the average spike count of the neurons across all tones. In the second method, we computed the cross correlation of the spike counts of each neuron in each trial with the trial-and-tone-averaged population spike count and used this measure as the data value.

Quantification of the similarity between modules

We quantified the similarity between two modules $\mathbf{w}_i$ and $\mathbf{w}_j$ (the modules are normalized by the algorithm) using cosine similarity (Steinbach et al., 2000) as follows.

$$similarity = \left\langle \mathbf{w}_i, \mathbf{w}_j \right\rangle \tag{3.33}$$

The similarity is a value between 0 and 1. When quantifying the similarity between the modules first we compute the similarity between each pair of modules. Then we match the modules based on this similarity measure and compute the similarity between the two decompositions as the mean similarity between the matched modules.

Results

We analyzed the neural responses of 85 auditory neurons recorded for long tones and click sequences using the new space-by-time NMF algorithms. The same data was previously analyzed for Gaussian update rules in section 2.5.1. The dataset containing the responses to long tones had a Fano factor of 0.98 when averaged over neurons, time bins and stimuli. The corresponding value for the responses to click dataset was 0.99.

Gaussian, Poisson, binomial (with n set to the maximum value in the dataset) and negative binomial (with $r = 1$) update rules for space-by-time NMF were applied for each dataset. We found that binomial and negative binomial results were very similar to those of Poisson rules (Figure 3.8). Since the Fano factors of the datasets were also close to one, we only used Gaussian and Poisson update rules for the subsequent analysis.

First, we will compare the quality of the modules derived from Gaussian and Poisson rules. The temporal modules derived from space-by-time NMF for Gaussian and Poisson rules for the responses to long tones were highly similar (Figure 3.6A). There are two differences in the module that describes the phasic response between the two rules: the module from Gaussian rules has a sharper peak, sharper fall off and a shorter duration (0 - 120 ms Vs 0 - 150 ms). The spatial modules from the two decompositions are shown in Figure 3.6B. The similarity between the spatial modules of the two rules was 0.61 ± 0.03 indicating there were similarities between modules derived from two rules. We could visually observe that the Poisson derived modules have better clustering in terms of the weight distribution across neurons in the modules. They are also better localized in closeby tetrodes – a result we expected from the tonotopical organization in A1. Some modules derived from Gaussian rules were almost entirely allocated to one neuron while Poisson rules derived modules had more equalized weight distributions across neurons. To quantify this observation, we evaluated the sparseness of the modules derived from the two update rules using the measure from (Hoyer, 2004). We found that the Gaussian modules were more sparse compared to Poisson modules (sparseness of 0.8±0.01 compared to 0.71 ±0.01, Wilcoxon rank sum test, $p < 0.001$).

Next we compared the quality of the modules extracted by the two rules for the responses recorded from same neurons to click sequences (Figure 3.7A). The temporal modules derived from both rules clustered the repetitive clicks of 4 and 8 Hz sequences into separate single modules. To compare the spatial modules from the two rules we computed the similarity between the two sets of spatial modules using the best two matching pairs. We found the distance to be 0.86 ± 0.03, indicating that the two best matching pairs of modules for the click dataset had a higher similarity than the modules derived for long tones.

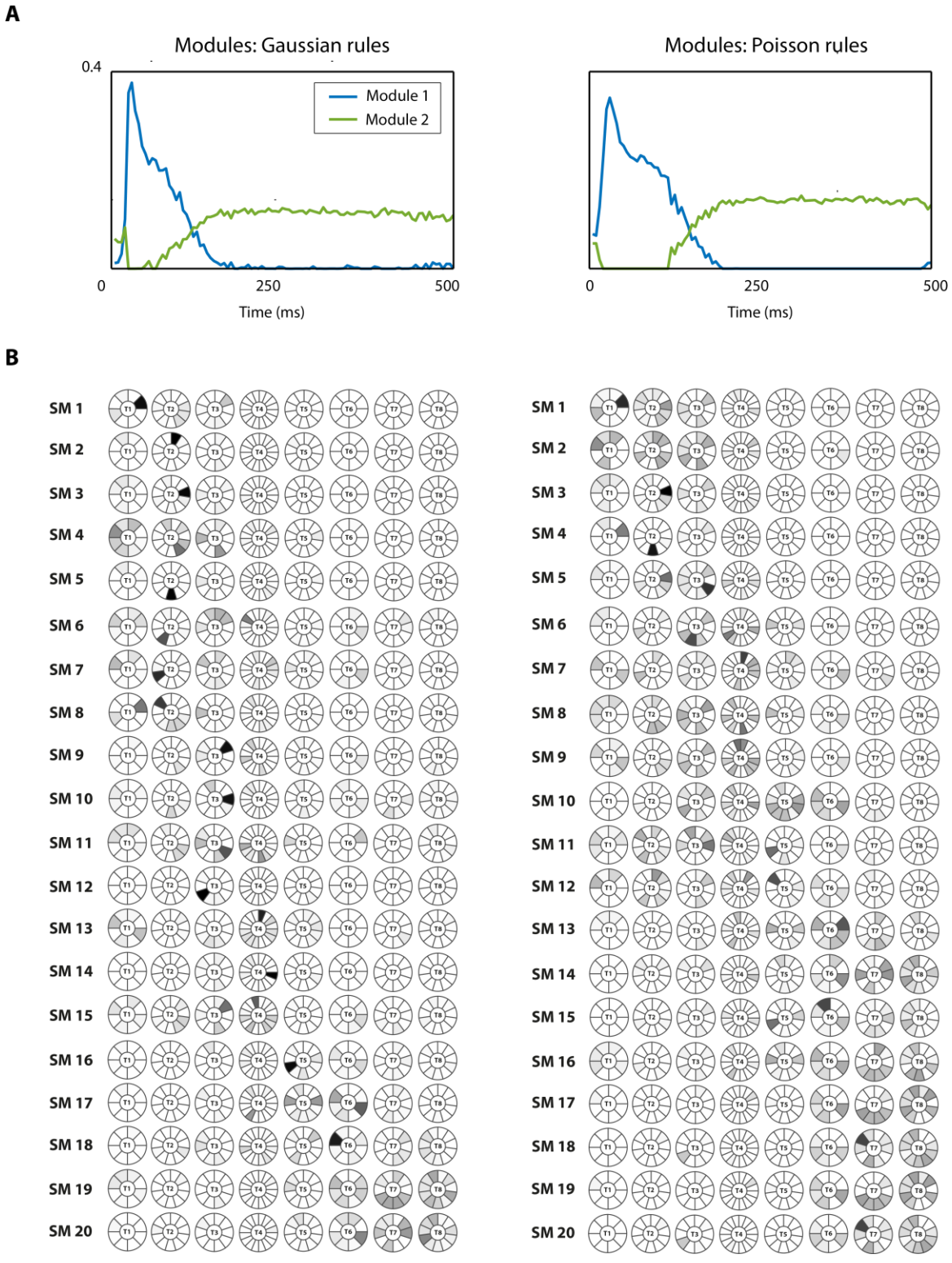

Figure 3.6: The module structure derived from Gaussian and Poisson space-by-time NMF rules for the neural responses recorded from 85 A1 neurons to long tones. A: Temporal modules. B: Spatial modules mapped on to a schematic figure that represents the eight tetrode configuration used for the recording (T1: tetrode1, ..., T8: tetrode 8). Each row shows the weights of the neurons in one spatial module (SM). The weight of each neuron in spatial module takes a value between 0 and 1. All neurons recorded from one tetrode are shown around the respective tetrode.

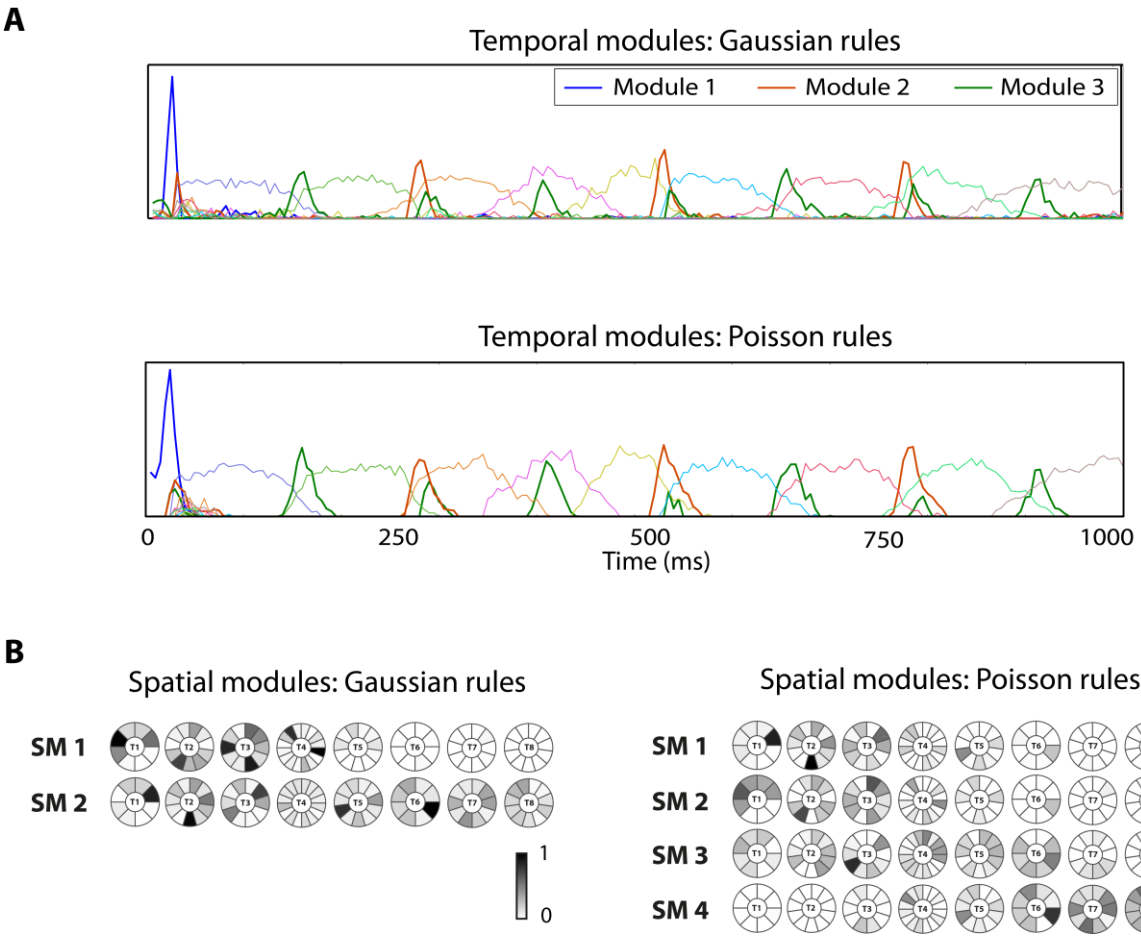

Figure 3.7: The module structure derived from Gaussian and Poisson space-by-time NMF rules for the neural responses recorded from 85 A1 neurons to click sequences. A: Temporal modules. B: Spatial modules mapped on to a schematic figure that represents the tetrode configuration used for the recording (T1: tetrode1, ..., T8: tetrode 8).

Then we compared the decoding performance of the new rules with that from Gaussian update rules. For long tones, Gaussian update rules showed a small performance advantage of 3.76 ± 3.45 % (Wilcoxon rank sum test, $p < 0.001$). There was no significant difference between the ability to discriminate stimuli by other rules. This result was reversed in the click dataset. Poisson update rules had a small gain in decoding performance of 5.24 ± 4.29 % compared to Gaussian rules (Wilcoxon rank sum test, $p < 0.001$). Again, there was no performance difference between Poisson, binomial and negative binomial rules. This similarity might partly be due to both binomial and negative binomial distributions converging to Poisson distribution when the Fano factor approaches one. It could also be due to the heterogeneity in the datasets; they contain supra-Poisson distributed spike counts, but near the stimulus onset, the spike counts become sub-Poisson (Churchland et al., 2010). We note that although there was a performance difference between Gaussian and Poisson rules for the two datasets, the difference was small, confirming our observations from simulated data.

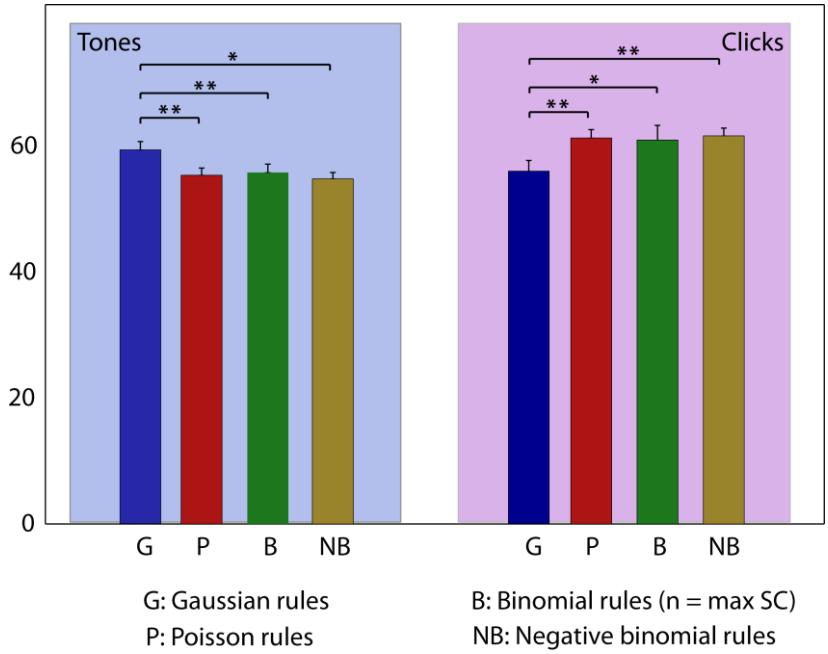

Figure 3.8: Decoding performance of auditory responses using space-by-time NMF rules. (*: $p < 0.05$, **: $p < 0.001$; Wilcoxon rank sum test, error bars show s.d.)

Many neurons in the dataset display a decrease in the Fano factor below one at the response onset indicating that they fire more reliably during the initial response (Figure 3.9A). From our previous observations on temporal modules of the two rules, we found that the temporal module which describes the phasic response had a sharper peak, sharper fall off and a shorter duration for Gaussian rules. In chapter 2, we found that sharpening of temporal modules could lead to enhanced extraction of precise spike timing information from the response onset. We performed an analysis using the first spike latencies to identify how the two rules would perform when extracting the first spike latency information using two datasets. The first dataset contained only the first spike of each neuron in each trial. The second dataset contained all spikes except the first spike of each neuron in each trial. In order to compare the qualitative differences between the decompositions derived from the two rules we describe the decomposition that gave the optimal performance, not the one that corresponds to the optimal decomposition obtained using the full information space and time.

As shown in Figure 3.9B. we found a qualitative difference in the temporal modules extracted from the two rules when only the first spike of each neuron in each trial was available. All five temporal modules of the Gaussian rules were concentrated at the response onset with peaks at the time bins consisting of 15 - 20 ms, 20 - 25 ms, 25 - 30 ms, 35 - 40 ms and 40 - 45 ms respectively indicating that Gaussian rules identified which tone was played using fine temporal differences between the response onset of neurons. In contrast, temporal modules extracted by Poisson rules spanned longer time windows with a decline in the value of the temporal module with time indicating that the Poisson rules gave more emphasis for spike count information in longer time windows than fine differences in spikes at the onset to identify which tone was played. Poisson spatial modules were again more dense compared to

Gaussian spatial modules (Figure 3.9Figure 3.10A). Interestingly, there was no significant difference between the decoding performance of the two rules (Figure 3.9C). Thus, tones could be discriminated at a same level using different spatial and temporal patterns from the first spike latencies of the auditory cortical neurons. We further used a rank order decoder to identify the decoding performance when only the order in which the first spikes arrive is available. The performance of the rank order decoder was only 42 % compared to that from the Gaussian rules using first spike latency information. (46 % compared to the performance of the Poisson rules, Wilcoxon rank sum test, $p < 0.001$). Thus, about 2.2 - 2.4 times higher decoding performance could be obtained when absolute difference in the first spike time latencies are used than when only the order in which the neruons send their first spikes is available.

When using only the second and the subsequent spikes, the temporal module describing the phasic response was broader in the Poisson module than in the Gaussian module (Figure 3.9B). The decoding performance shown in Figure 3.9C indicated that the Gaussian rules were able to identify which tone was played slightly better than Poisson rules (54.78 ± 2.86 % compared to 52.04 ± 2.39 %, Wilcoxon rank sum test, $p < 0.05$). For both rules, the performance was slightly below the performance when using all spikes (4.27 ± 3.87 % for Gaussian rules and 2.99 ± 3.28 % for Poisson rules, Wilcoxon rank sum test, $p < 0.05$). Thus, the first spike latency information for long tones is highly redundantly coded in the subsequent spikes at the population level in our data.

Next, we applied two preprocessing methods to identify the robustness of the update rules. First we normalized the spike counts in each time bin with the logarithm of the average spike count of the neurons across all tones. This increases the value of the spike counts for low firing neurons (Figure 3.11A). Second, we used computed the cross correlation of the spike counts of each neuron in each trial with the trial and tone averaged population spike count. We used these cross correlations as our data matrix. Thus, this is also an evaluation of the information content in the cross correlation as well. This has the effect of increasing the value of spike counts in the onset response compared to the sustained response due to the highly reliable responses at the responses compared to sustained responses (Figure 3.11A, (Curto et al., 2009)).

When the spike counts of each time bin were normalized using the average spike count across tones, space-by-time NMF rules gave high weights in the spatial modules to low firing neurons that had larger amplitude in the time bins after normalization (the Spearman's rank correlation between the maximum weight in the spatial module and the averaged firing rate of the neuron was -0.66 for Poisson rules derived after preprocessing while the same for data without preprocessing was 0.47. Furthermore, Gaussian modules derived from the raw spike counts and preprocessed spike counts had a similarity of 0.45 ± 0.04 while the same for Poisson modules was 0.55 ± 0.04 indicating that the spatial modules of the two methods did not have high similarity). As shown in Figure 3.11B, both space-by-time NMF rules had a lower decoding performance after the preprocessing than when unprocessed spike counts were used (53.22 ± 1.31 % compared to 59.01 ± 2.61 % for Gaussian rules and 47.1 ± 1.86 % compared to 55.03 ± 2.25 % for Poisson rules, Wilcoxon rank sum test, $p < 0.001$). At the

population level, the responses of the low firing auditory neurons could also be informative about tone frequencies and intensities.

When using cross correlation values, the spatial modules derived from both update rules were qualitatively similar to the spatial modules derived from raw spike counts (Gaussian modules derived from the raw spike counts and preprocessed spike counts had a similarity of 0.65 ± 0.03 while the same for Poisson modules was 0.78 ± 0.03). The decoding performance of only Gaussian rules decreased (51.38 ± 0.91 % compared to 59.01 ± 2.61 %, $p < 0.001$). The modules derived from Poisson rules became slightly sparser (0.02 ± 0.02, $p < 0.001$) while those from Gaussian rules increased the sparseness by a larger amount (0.11 ± 0.009, $p < 0.001$). Thus, Poisson space-by-time NMF rules are able to extract information in the cross correlations robustly.

Auditory responses were recorded when the anesthetized cortex was in a synchronized state. Noise correlations in the data could be influenced by the up-down activity states as well as changes in the level of the synchronization (Curto et al., 2009; Pachitariu et al., 2015). Thus, we analyzed the effect of removing noise correlations in the data on the performance of space-by-time NMF rules. We removed the noise correlations in the dataset by randomly permuting trials of each stimulus independently for each neuron. Figure 3.12A shows the decoding results when applying Gaussian and Poisson rules on the responses to click sequences and long tones. Removal of noise correlations resulted in an increase in the optimal decoding performance (an increase of 6.47 ± 4.86 % for Gaussian rules, $p < 0.05$, and 13.9 ± 4.06 % for Poisson rules, $p < 0.001$, for tone data while an increase of 24.77 ± 3.83 % for Gaussian rules, $p < 0.05$, and 24.78 ± 4.39 % for Poisson rules, $p < 0.001$) indicating that the noise correlations reduce the ability to correctly identify stimuli in the recorded population responses. With the trial shuffling, the number of spatial modules required to obtain the optimal performance to identify long tones reduced from 20 modules to 15 modules for Gaussian rule and from 20 modules to 12 modules for Poisson rules respectively. For responses to clicks, both rules gave the optimal performance for only one spatial module while previously Gaussian rules required two spatial modules and Poisson rules required four spatial modules. Obtaining optimal performance using only one spatial module to identify clicks is as expected since many neurons across the tonotopic axis would be stimulated from a click which has a large frequency bandwidth. When removing the noise correlations by trial-shuffling, the spike counts become independent, thus becoming closer to a Poisson distribution. Representative temporal modules for a “three temporal modules and one spatial module”-decomposition are shown in Figure 3.12B for Gaussian rules and Poisson rules. The increase in the decoding performance for lower number of temporal modules is likely to have resulted because the structured trial to trial variability that the NMF update rules has to account due to up-down activity state changes is reduced when the trials are permuted across neurons.

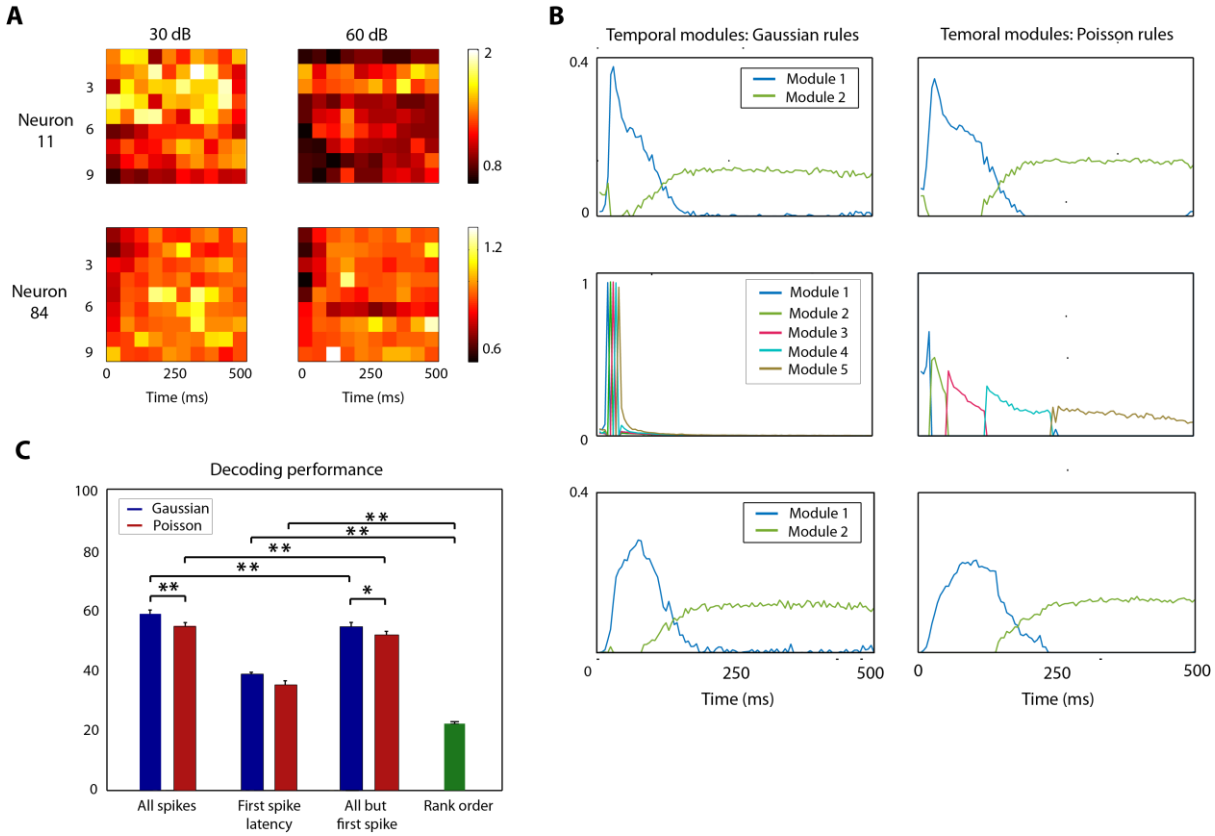

Figure 3.9: Analysis of first spike latency. A: Fano factor variation across the trial in two representative neurons in the dataset. Each row of the figure correspond to one neuron and each column corresponds to one of the two tone intensities of the nine frequencies used (30 dB and 60 dB). Decoding performance of space-by-time NMF for Gaussian and Poisson update rules for 50 ms bins. B: Temporal modules identified by Gaussian (left) and Poisson (right) space-byt-ime NMF update rules. The first row corresponds to modules identified for the full dataset containing all spikes. The second and the third rows correspond to the conditions where only the first spike of each neuron was used and all spikes except the first spike was used respectively. C: Decoding performance for the three conditions. (*: $p < 0.05$, **: $p < 0.001$, Wilcoxon rank sum test, error bars show s.d.)

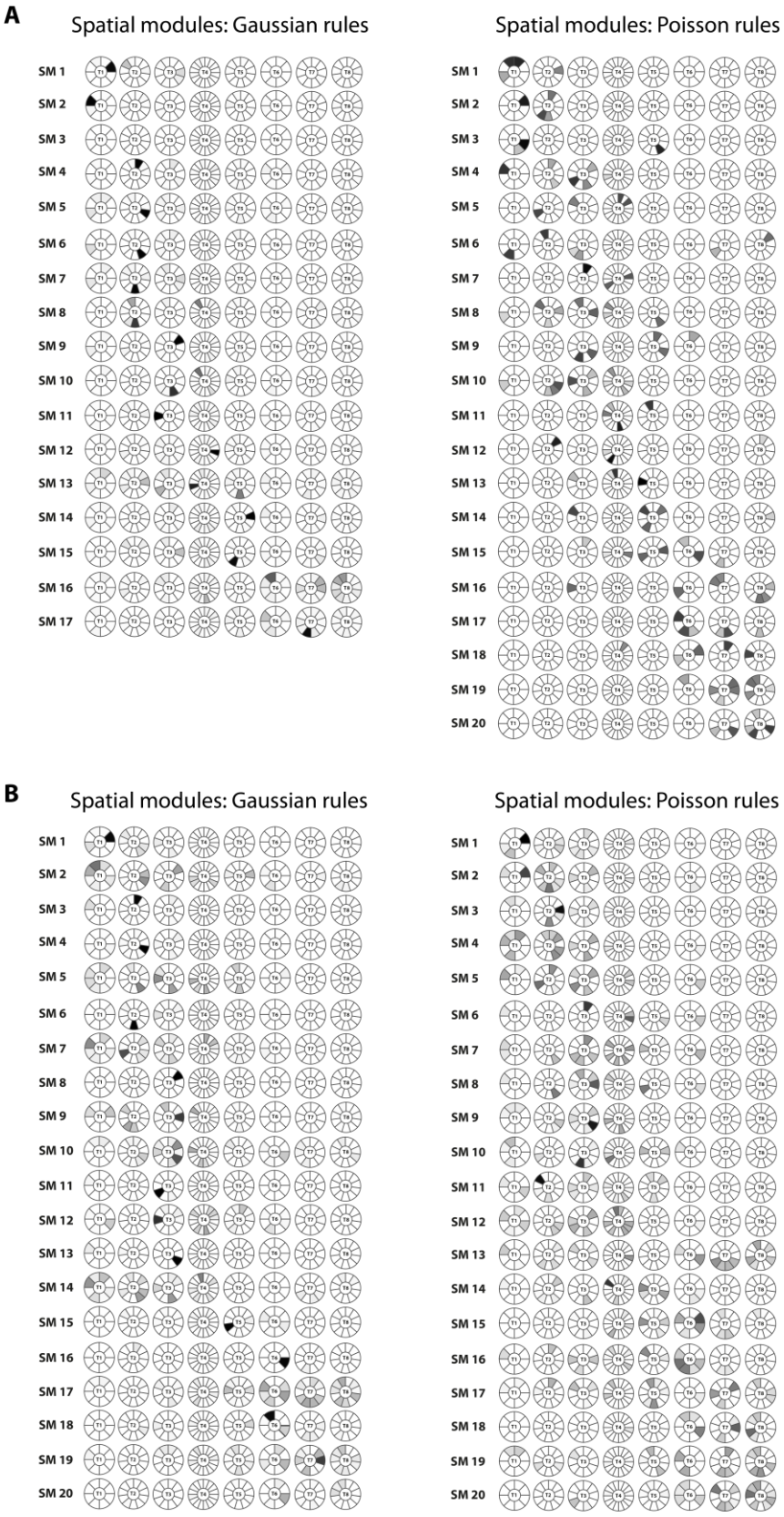

Figure 3.10: Spatial modules identified by space-by-time NMF for spike latency analysis. A: Spatial modules when the dataset contained only the first spike of each neuron in each trial. B: Spatial modules when the dataset contained only the second and subsequent spikes.

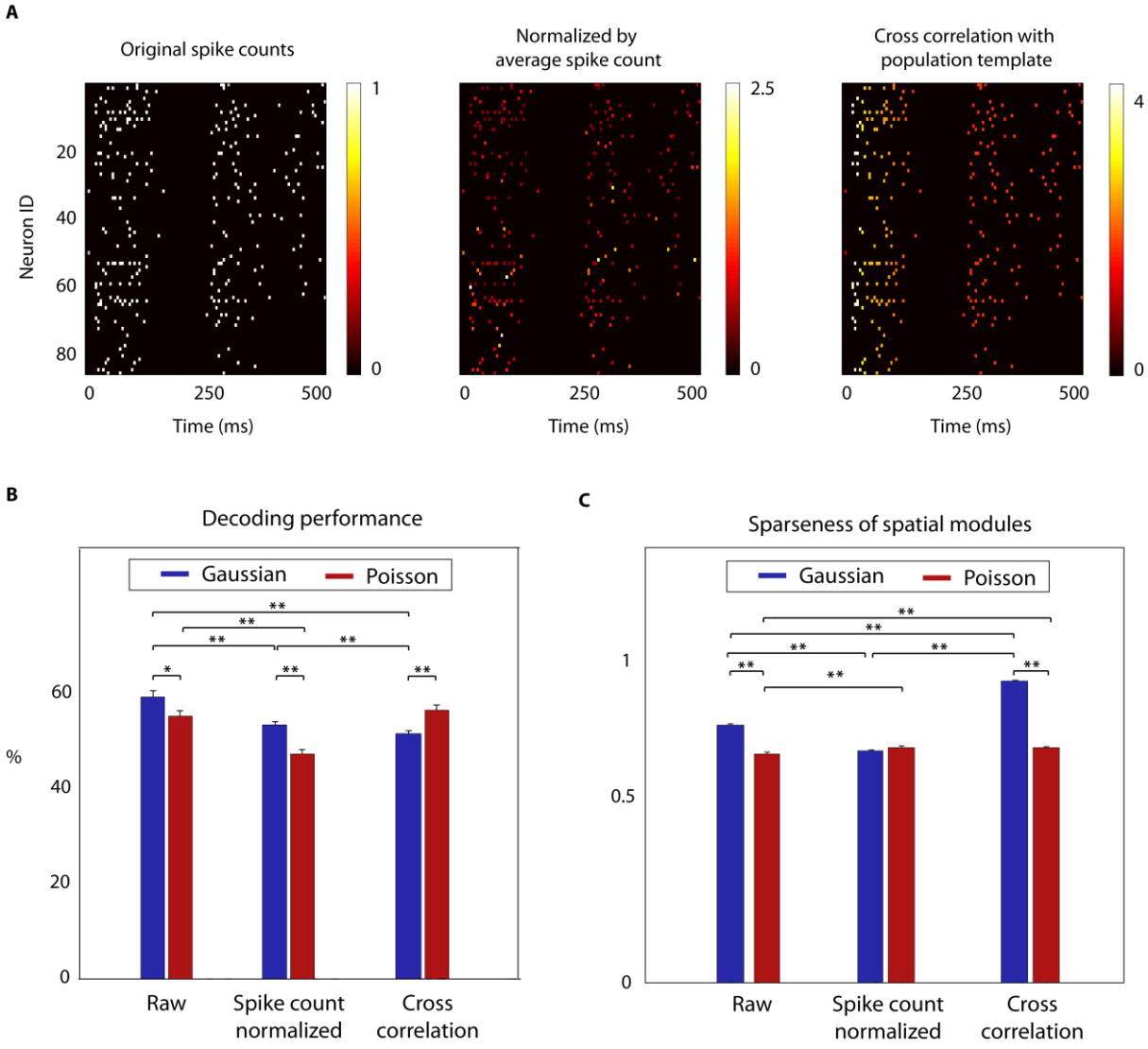

Figure 3.11: The performance of space-by-time NMF update rules after preprocessing data. A: An example of how the two methods affected the spike counts. Normalizing the spike count of each time bin by the average spike count of the neuron across all trials and tones (middle figure) decreased the value in the time bin for high firing neurons while increasing it for low firing neurons. The cross correlation with the population template increased the value of the spike counts at the trial onset compared to the sustained period. B: The decoding performance of the two update rules for the two methods compared to the values obtained using unprocessed spike counts. C: The sparsity of the spatial modules from each preprocessing method compared to those for unprocessed data. (**: $p < 0.001$ and *: $p<0.05$; Wilcoxon rank sum test, erorr bars show s.d.)

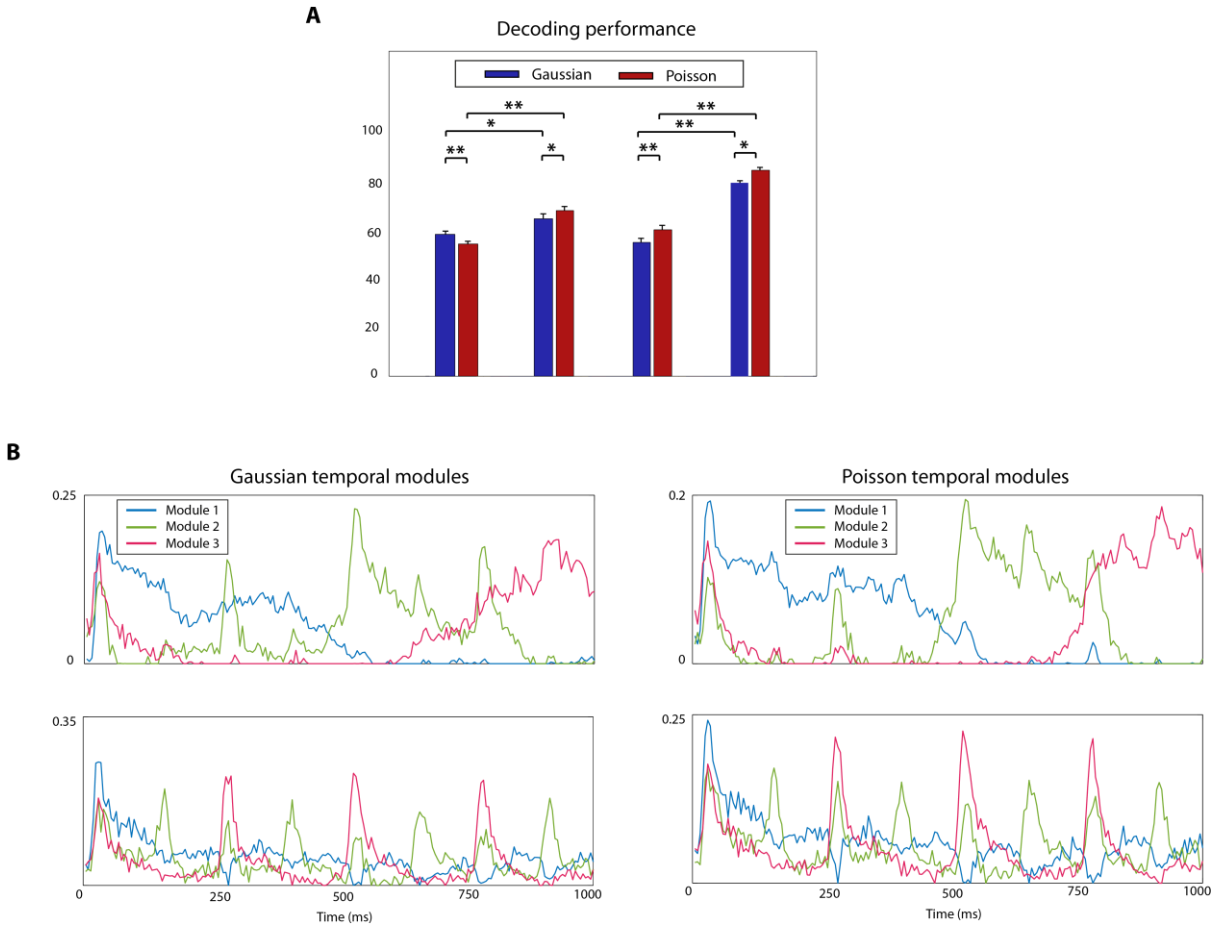

Figure 3.12: Application of space-by-time update rules on trial-shuffled responses to long tones and click sequences. A: Decoding performance for click sequences and long tones (**: $p < 0.001$ and *: $p<0.05$; Wilcoxon rank sum test, erorr bars show s.d.) B: Temporal modules of the 'three temporal modules and one spatial module' decompositions extracted by Gaussian rules (left) and trial-shuffled data (right). The top row shows the spatial modules of the correlated data while bottom row shows the spatial modules of the independent data.

3.4.4 Using new update rules to study population coding of stimulus location in rat somatosensory cortex

Next we evaluated the performance of the Poisson update rules on the neural responses recorded from somatosensory cortical neurons for whisker deflections that we analyzed in chapter 2 using Gaussian update rules.

Data analysis

The details of the dataset are fully described in section 2.5.2. We ran the same analysis for two mode datasets; 1) a dataset which contained the first spike of each neuron in each trial and 2) a dataset that had all spikes except the first spike of each neuron in each trial. We applied the Poisson space-by-time NMF update rules for each dataset in the same way as described in section 2.5.2 and used the same decoding procedure to evaluate the stimulus discrimination capability. To evaluate the decoding performance for first spike latency information, we used a rank order decoder described in the section 2.5.3. We evaluated the sparsity of the modules using the sparsity measure from (Hoyer, 2004) that is described in section 3.4.2. We compared the similarity of the spatial modules through the procedure described in section 3.4.3.

Results

First, we will look at the module structure derived from the two update rules for the full dataset. This is shown in Figure 3.13A. Both update rules gave the best stimulus discrimination for five spatial modules. However, Gaussian rules required only one temporal module while Poisson rules required two temporal modules. The optimal decomposition for Gaussian rules is the same as we discussed in chapter 2 section 2.5.2. Here we will recapture the main points when comparing the two rules. Spatial modules have a clear composition with respect to neuron responses as could be observed from the coefficient structure. The first three modules correspond to neurons responding mainly to D1 - D3 stimulations. Last two modules group neurons responding to non-principle whisker deflections in addition to principle whisker deflections. The spatial modules of the Poisson rules is broadly similar to the spatial modules from Gaussian rules with a similarity between the two sets of spatial modules being 0.87 ± 0.03 .

The temporal module of the Gaussian rules has an elevated activation level at the start of the trial between 5 - 15 ms of the response during the highly reliable onset response of the neurons. The temporal module decays during the trial. We found that this effect gave rise to a temporal weighing of the spikes in the coefficient matrix that code the temporal dimension into the low dimensional representation. Poisson rules use two temporal modules to represent the temporal dimension. The first module mainly encodes the activity between 0 - 25 ms of

the response while the second module mainly encodes the activity between 25 - 100 ms of the response. As could be seen from the coefficients, this aids to segregate the early response into the principle whisker into the first temporal module and the response to the non-principle whisker mainly of the second temporal module. It further shows that neurons in the first spatial module (which are mainly from the D1 barrel column) responding mainly to D1 principle whisker also responds to E1 non-principle whisker. Furthermore, neurons in the third spatial module (which are mainly from the D3 barrel column) responding mainly to D3 principle whisker also responds to E2 and E3 whiskers. In both cases, the coefficients for the non-principle whiskers are smaller because the number of spikes for the non-principle whisker is lower compared to the principle whisker and the latency of the response is higher for the non-principle whisker than for the principle whisker. The fourth spatial module of Poisson rules describes responses to C1 - C3 and E3 whiskers. These responses have an early as well as a sustained component. The last spatial module contains sustained responses. A neuron such as neuron 263 shown in Figure 2.7B, could belong to a spatial module describing onset responses as well as spatial modules that describe sustained responses.

In terms of the ability to decode whiskers deflections (Figure 3.14), Poisson rules were slightly, but significantly better than Gaussian rules (84.72 $\pm$ 3.41 % compared to 82.92 $\pm$ 3.82 %, Wilcoxon rank sum test, $p<0.05$). This could be mainly because the dataset had a mean Fano factor of 0.99 averaged over all neurons and time bins which indicates that the variability is Poisson distributed. Moreover, the dataset was made by pooling the activity recorded over multiple recording sessions and the noise correlations are very small as we reported in chapter 2. This would have further contributed to the better performance of Poisson rules over Gaussian rules.

Next we compared the capability of the update rules to represent the first-spike latency information. The module structure derived when the data matrix contained only the first spike of each neuron in each trial is given in Figure 3.13B. The optimal decompositions that represented this information only required one temporal module for both Gaussian and Poisson rules. However, the Gaussian module had a narrower peak centered at 5 - 10 ms while the Poisson rules extracted a wider module with a peak between 5 - 15 ms. In terms of spatial modules, the best performance for Gaussian rules was obtained for six spatial modules and that for Poisson rules required five spatial modules. The Gaussian spatial modules were sparser than Poisson modules (sparseness of 0.31 $\pm$ 0.004 compared to 0.51 $\pm$ 0.01, Wilcoxon rank sum test, $p<0.001$). This could be visually observed in Figure 3.13B.

Although these differences in spatial and temporal modules are present between the two rules, both rules had similar capacity to discriminate whiskers using the first spike information. Interestingly, the first spike latency has the same or slightly higher ability to decode stimuli as when using full information in space and time. It is very slightly but significantly (88.06 $\pm$ 4.86 % compared to 85.56 $\pm$ 4.38 %, Wilcoxon rank sum test, $p<0.05$) higher for Gaussian rules, but there is no difference for Poisson rules. This result has to be considered keeping in mind that the effect of noise correlations on the coding is minimal because the population has little noise correlations since it is a pooling across multiple recording sessions. As we discussed in chapter 2 in detail, the only way that the

space-by-time NMF temporal module code temporal information is by modulating the value of the temporal module across time giving a higher value to the spikes that occur at the response onset and decreasing the importance of spikes as the response progresses. Under such conditions, this indicates that if the first spikes of a large population of neurons (300 neurons in this dataset) are available, the deflected whisker could be identified at the same accuracy as when using information about neuron identity and temporal information from all spikes if there is a method to weigh the first spikes appropriately. This can be intuitively understood from Figure 3.13D. The trial averaged population firing rate when only the first spike is available has a similar temporal profile to that of when all spikes are available.

To compare the effect of the temporal weighting on the performance with the performance when we only use the spike arriving order, which is used in many studies to identify the coding capability of the first spike latencies (Johansson and Birznieks, 2004; Panzeri and Diamond, 2010), we used a rank-order decoder to decode the whiskers. We found that when only the first spike order was available to a linear decoder, the ability to discriminate the whiskers is 83.8 $\pm$ 0.02 %. This is only significantly different from the first spike decoding performance of the Gaussian rules (Wilcoxon rank sum test, $p < 0.001$). Thus, this indicates that the temporal weighting method used by space-by-time NMF performs comparable to a rank order decoder.

Since the first spike latency of the population is sufficient to obtain the full decoding of the whiskers, we tested the contribution from the remaining spikes on whisker identity. In order to do so, we removed the first spike from all neurons in all trials and performed space-by-time NMF. In order to compare the qualitative differences between the decompositions derived from the two rules we describe the decomposition that gave the optimal performance, not the one that corresponds to the optimal decomposition obtained using the full information carried by space and time. As shown in Figure 3.13C, we found that Gaussian rules required much more modules (two temporal modules and seven spatial modules) compared to Poisson rules (one temporal module and five spatial modules) to give the best performance. Temporal modules derived without using the first spike were broader than when using all spikes. The first temporal module extracted by the Gaussian rules ranged between 0 - 35 ms while the second temporal module spanned the remaining response duration. The Gaussian derived spatial modules were sparser than the modules derived using Poisson rules (sparseness of 0.79 $\pm$ 0.01 compared to a sparseness of 0.64 $\pm$ 0.01, Wilcoxon rank sum test, $p < 0.001$). This is mainly because of the dominance of single neurons in spatial modules 5 - 7. All three of these neurons have sustained responses. Raster plot and the tuning curve of one of these neurons, neuron 278 is shown in Figure 2.7B. It shows that this neuron responds for many whiskers. Thus the biasness that we observed in Gaussian rules for long firing neurons when performing network simulations in section 3.4.2 could occur in experimental data as well.

When we compared the decoding performance of the two rules, we found that the loss of the first spike information caused relatively large loss of the decoding performance for both rules (23.61 $\pm$ 4.27 for Gaussian rules and 18.05 $\pm$ 4.97 for Poisson rules, Wilcoxon rank sum test, $p < 0.001$). This indicates that while the first spike latency information is redundantly coded in the remaining spikes across space and time, it is still an important component in

encoding the whisker identity. The loss in the whisker identification capability in the Poisson rules was less than in the Gaussian rules (Wilcoxon rank sum test, $p < 0.001$). This can be explained by the results of our network simulations in section 3.4.2 where we observed that Gaussian rules could be sparser due to the long firing neurons in the following way. According to Figure 3.13 E, the trial-averaged population firing rate to the second and subsequent spikes is less precise at the onset and has less variability between different whiskers. This resulted in broader temporal modules that suggest that the coefficients give more emphasis for counting spikes rather than indicating the precise spike timing of spikes by using a time dependent weight. In this case, when the spatial modules are dense, meaning that similar weight is given to spikes of many neurons, the coefficient values would be more sensitive to small variations in the population spike counts. This would give an advantage in whisker discrimination to Poisson rules over Gaussian rules. Furthermore, Figure 3.13E suggests that since the temporal profiles of the trial-averaged responses for different whisker deflections are highly similar, the spatial dimension has to contain information that is variable across different whiskers. This could be achieved if it is possible to efficiently extract differences in the firing rates between different subpopulations of neurons. Thus more homogenous weight distributions across neurons would likely to be useful towards this.

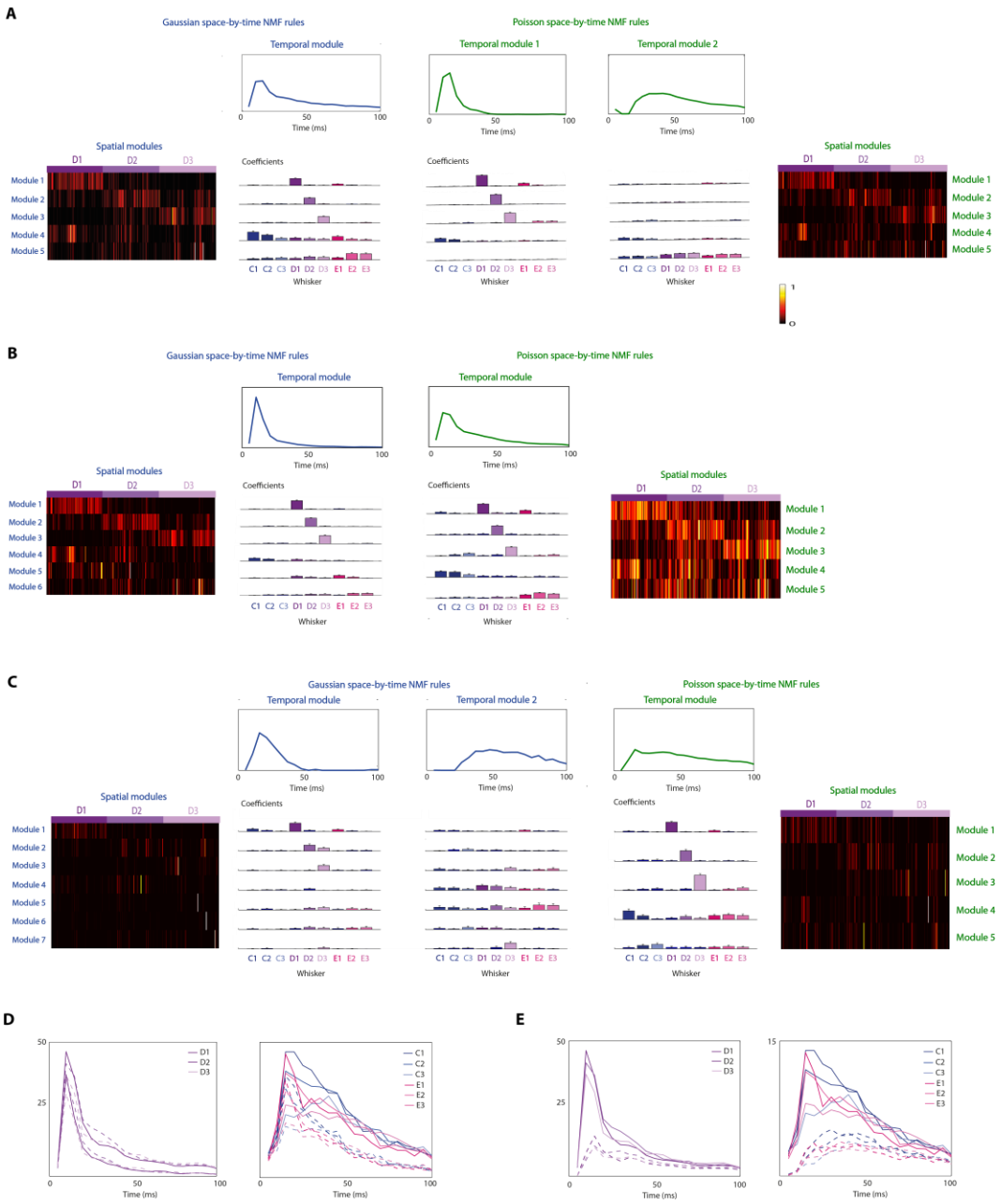

Figure 3.13: Module structure and the coefficients extracted by Gaussian and Poisson update rules from the neural responses of 300 neurons to whisker deflections A: Modules and the coefficients for the full dataset (D_F). B and C: same as in A but for the responses consisting of only the first spike of each neuron in each trial (D_{FS}) and when the responses contained all spikes except the first spike of each neuron in each trial (D_{SS}) respectively. In all cases, coefficients are grouped according on whiskers. Each row of coefficients are related to the spatial module on the side. Each column of coefficients are related to the temporal module shown above them. D: Population average responses for the full dataset and for D_F (thick lines) and D_{FS} (dashed lines). E: Population average responses for the full dataset and for D_F (thick lines) and D_{SS} (dashed lines).

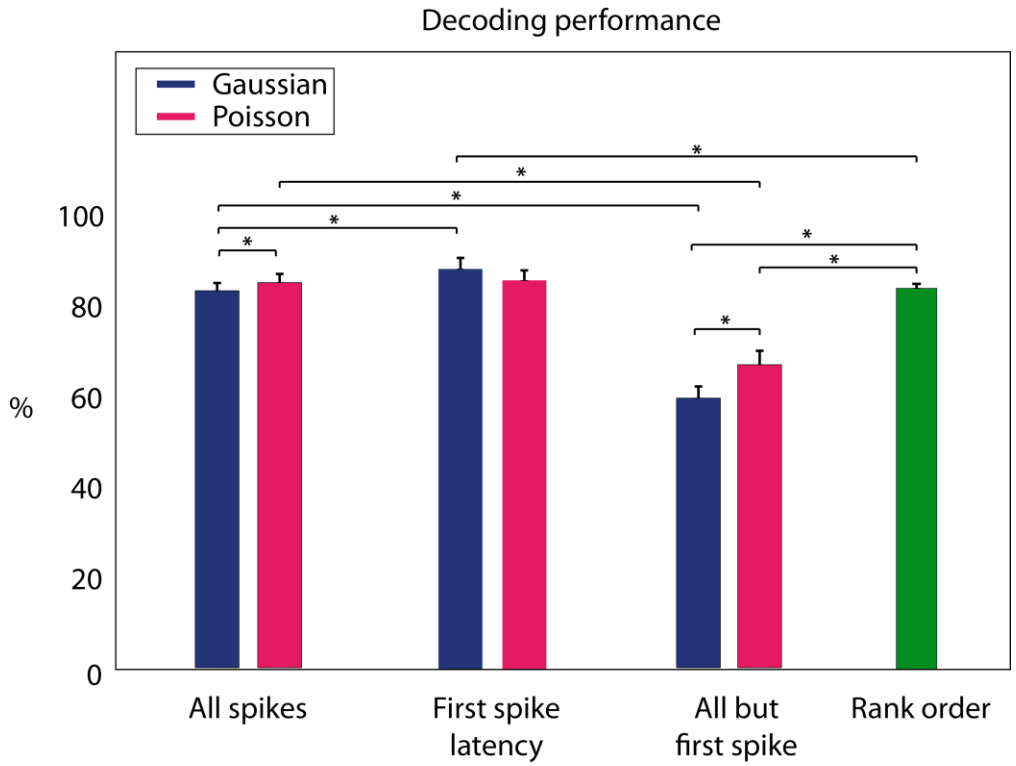

Figure 3.14: Decoding performance of Gaussian and Poisson space-by-NMF update rules for somatosensory neural responses. Decoding performances were evaluated for the following four conditions: when using all spikes in the dataset, using only the first spike of each neuron in each trial, using information in all spikes except the first spike in each neuron in each trial and when using the information present in the order in which the first spike is discharges by neurons. (*: $p<0.05$; Wilcoxon rank sum test, erorr bars show s.d.)

3.5 Discussion

We extended space-by-time NMF to model neural variability. We modeled sub-Poisson, Poisson and supra-Poisson variability using binomial, Poisson and negative binomial noise models. We identified Bregman divergences corresponding to these distributions, which enable space-by-time NMF to be formulated in a probabilistic setting where each observed spike count is a sample from the respective noise distribution fit with the reconstructed spike count as its mean. Then, the minimization of these Bregman divergences through NMF corresponded to maximum likelihood estimation of the mean of the distribution. We used the generalization framework proposed by (Dhillon and Sra, 2005) and derived new update rules for space-by-time NMF that minimize these optimal Bregman divergences. We evaluated the new update rules methodically using statistical simulations, network simulations using a conductance-based integrate-and-fire network and investigated the performance of the new rules in neural responses recorded from the auditory cortex to long tones and clicks. We now discuss the implications of our findings when using space-by-time NMF to analyze neural data.

Choosing appropriate update rules

The choice of the appropriate update rule to use for space-by-time NMF depends on several criteria such as the required accuracy, the computation time and the properties of the neural responses.

Our findings indicate that there is a performance advantage when using the update rule corresponding to the underlying noise distribution. This finding is a general observation across other types of data corrupted by different noise distributions (Banerjee et al., 2005; Cheung and Tresch, 2006; Févotte et al., 2009). The performance advantage over existing Gaussian rules was highest when the trial-to-trial variability was supra-Poisson. When the data became increasingly regular, the performance difference decreased. Thus, if the data had high variability (Oram et al., 1999), a higher performance could be obtained using negative binomial rules. However, when the data had low variability (Kara et al., 2000), the Gaussian update rules performed similarly to the optimal binomial update rule as we observed for responses of retinal ganglion cells to natural images. Thus, if the variability in the data is low, Gaussian rules are well suited for space-by-time NMF.

The second main factor to consider is the computation time. This is an important aspect because NMF algorithms are iterative. In terms of algorithmic complexity, Gaussian update rules are the simplest. Poisson update rules come next in terms of complexity. Binomial and negative binomial update rules have the highest algorithmic complexity and require more matrix multiplications per iteration, thus increasing the overall computing time. Due to the complexity of using the extended update rules Binomial rules take the longest to run. Furthermore, both binomial and negative binomial rules have an extra model parameter to be evaluated. While we

found that the maximum spike count in the data could be used as the model parameter for binomial rules, the best performance for negative binomial rules was obtained when the model parameter was selected using parameter exploration. The requirement to run parameter exploration can increase the computational time considerably. Thus, we suggest that if the data have supra-Poisson variability and the purpose of the analysis is exploratory, it would be faster to use Gaussian rules as a first pass and then optimize using negative binomial rules.

Poisson update rules gave comparable performance to binomial rules for sub-Poisson noise distributions. From the analysis using the auditory responses and network simulations, we found that the difference between Gaussian and Poisson rules resides mainly in the equality of the spatial modules and the robustness to outliers. In the datasets we examined, Poisson rules showed a lesser propensity to be biased from outliers and led to denser clustering compared to Gaussian rules. However, the slightly lower performance of Poisson rules compared to Gaussian rules in the tone dataset suggests that the denser clustering in the spatial modules may decrease the level of fine detail that the lower-dimensional structure can represent. Thus, the choice between Gaussian and Poisson update rules should take into account the statistics of the data to be analyzed and the required level of accuracy and detail.

Finally, if there is structured variability present in the data, similar to up-down state variability, in general, more modules will be required to obtain the optimal performance. Since the update rules in NMF algorithms are designed to minimize the dissimilarity between the original dataset and the reconstructed dataset, they are likely to capture large-scale stimulus independent variability when only a small number of modules are used. This will need to be considered when deciding on the number of modules to be extracted.

Convergence of update rules

We did not prove the convergence of our new space-by-time update rules to a stationary solution. However, on all datasets we analyzed, the update rules did converge. We also found that results obtained with space-by-time NMF were more stable across multiple runs with different random initializations compared to two factor NMF. The stability of the modules could be improved further by initializing several times using different random values, running the algorithm for some iterations and then starting the full iteration process using the initialization that gave the lowest Bregman divergence.

Space-by-time NMF to study population coding in space and time

We gained further insight into the population code used by the somatosensory and auditory neurons using the new Poisson space-by-time NMF update rules. To summarize, spike timing,

particularly the first spike time latency was found to be an important component for coding whisker location by somatosensory neurons at the population level. Spike timing also carried information in auditory cortical neurons, but it was found to be redundently coded in the information present in the firing rate. We will discuss these in detain in chapter 4.

Extraction of temporal information

The first-spike latency in somatosensory neurons gives similar whisker discrimination capability as when using information in all spikes. The method by which space-by-time NMF extracts this temporal information is by weighing the spikes using a weight that varies across time. This weight, that is the shape of the temporal module, is determined from the summed population activity. This could be observed by comparing the temporal modules in Figure 3.13 A - C with the trial-averaged population responses in Figure 3.13 D and E. Deriving the temporal modules in this manner has the advantage that they are data driven. In other words, they can dynamically change when the stimulus paradigm is changed. An interesting question is whether such a mechanism is actually biologically realistic. Here, we speculate on this idea. The dependence of the temporal modules on the summed population response is in favor for a possibility of a biological implementation since this information is readily available to a downstream neuron. Then the question would be whether it would be possible to implement a temporal weighing mechanism based on the summed population response using a biologically realistic method. One possible way that this may be achieved could be using an appropriate short-term spike timing-dependent synaptic plasticity (STDP) mechanism (Abbott and Regehr, 2004). According to many experimental and modeling studies, the efficiency of synaptic transmission is activity and history dependent. When neurons communicate through chemical synapses, pre-synaptic neurons release neurotransmitters that get attached to receptors of the post-synaptic neurons. This changes the activity of the post-synaptic neurons. This process can change on rapid time-scales from milliseconds and can last minutes (Klug et al., 2012). For example, prolonged exposure to neurotransmitter could lower the sensitivity of the neuron to the neurotransmitter and can decrease the response of the cell to the neurotransmitter (Trussell and Fischbach, 1989; Xu-Friedman and Regehr, 2004). The decrease of the synaptic efficacy is known as synaptic depression and the increase is known as synaptic facilitation. STDP has been suggested as a mechanism through which temporal information could be decoded (Buonomano, 2000; Goel and Buonomano, 2014), to have a role in processing natural stimuli (Klyachko and Stevens, 2006; Kandaswamy et al., 2010), to be a plausible explanation to experimentally observed effects (Chance et al., 1998; Carandini et al., 2002) and could change the task performance levels (Bourjaily and Miller, 2012). (Masquelier et al., 2008) have shown that STDP could detect early patterns from input spike trains. A time dependent learning mechanism through the changes in synaptic weights has been proposed in (Rao and Sejnowski, 2003). Thus it may be possible that cells could exhibit time dependent changes in synaptic efficacy that may result in a computation similar to that of temporal modules used in space-by-time NMF. A related question is whether

such a mechanism could be plausible for downstream neurons accessing whisker related information from the layer 4 of the barrel cortex. Layer 2/3 neurons in the barrel cortex receive afferent input from the layer 4 neurons (Petersen, 2007). Layer 2/3 neurons have been shown to display synaptic plasticity (Kim et al., 2016). Further experimental and modeling work is needed to verify whether a mechanism such as that used by space-by-time NMF may be used in these neurons to decode input from the layer 4.

.

Chapter 4: Discussion

In this thesis, we explored new methods to identify biologically meaningful response characteristics on a single trial basis from simultaneously recorded large scale neural datasets. Specifically, we introduced space-by-time NMF as a new method that can extract meaningful patterns concurrently from spatial and temporal dimensions. Space-by-time NMF identifies salient spatial and temporal patterns separately into two modules matrices. A third coefficient matrix describes how the population response of each trial is generated through the recruitment of the identified spatial and temporal modules. Then secondly, we extended the algorithm to model sub-Poisson, Poisson and supra-Poisson variability in spike counts.

At the end of each chapter we discussed our findings from each analysis. Here we summarize and extend those discussions into a broader context elaborating on using NMF as a method for spatiotemporal population analysis in a broader context.

4.1 NMF with respect to other methods that can be used to study spatiotemporal activity

Many methods have been proposed as dimension reduction methods to extract representative and informative detail from large scale datasets. These methods were summarized in chapter 1 of this thesis. Broadly these methods are classified as static dimension reduction methods and dynamic dimension reduction methods (Roweis and Ghahramani, 1999; Churchland et al., 2007). Commonly used static dimension reduction methods include principal component analysis (PCA) (Chapin and Nicolelis, 1999; Hu et al., 2005; Mazor and Laurent, 2005; Peyrache et al., 2010; Lopes-dos-Santos et al., 2011; Ahrens et al., 2012; Churchland et al., 2012), independent component analysis (ICA) (Comon, 1994; Laubach et al., 1999a; Hyvärinen and Oja, 2000; Laubach et al., 2000; Brown et al., 2001) and factor analysis (FA) (Byron et al., 2009; Santhanam et al., 2009; Yu et al., 2009; Cunningham and Byron, 2014). Commonly used linear dynamical methods include hidden Markov models (HMM) (Radons et al., 1994; Abeles et al., 1995; Seidemann et al., 1996; Gat et al., 1997; Kemere et al., 2008; Escola et al., 2011; Bollimunta et al., 2012; Ponce-Alvarez et al., 2012), Kalman filters (Wu et al., 2004; Wu et al., 2006; Wei et al., 2015; Wu and Liu, 2015) and autoregressive models (Kulkarni and Paninski, 2007; Lawhern et al., 2010; Buesing et al., 2012a, b). We now discuss space-by-time NMF as an analysis method in relation to these methods.

Similar to NMF, PCA, ICA and FA are data-driven dimension reduction methods. Often these methods are used only along the spatial dimension to describe the population activity using neuronal groupings that show simultaneous firing. Variants of PCA, ICA and FA that are similar to spatiotemporal NMF can be formulated (Onken et al., In preparation). In these variations, the

extracted features are patterns of the population activity over time. Thus, the spatial and temporal dimensions are not separated in these methods. Currently, it is unknown how to obtain space and time separated representations from these methods. On the other hand, spatial and temporal dimensions are the two dimensions that describe the low dimensional representation. This naturally gives rise to the extraction of low-dimensional features describing spatial and temporal patterns in the population code. It describes neuronal groupings with simultaneous firing as well as the patterns of firing rate modulations displayed by these neuronal groups. Furthermore, using the permutation based procedure that we used in our work, space-by-time NMF could be used to analyze neural responses in a single dimension. We found that space-by-time NMF is able to perform equally or with greater performance compared to spatiotemporal versions of PCA, ICA and FA (Onken et al., In preparation). Thus, we find that space-by-time NMF is a competitive method compared to other dimension reduction methods such as PCA, ICA and FA while offering the capability of separating space and time.

Static dimension reduction methods such as PCA formulate the time progression of neural activity by describing how the identified latent components vary across time. Dynamical methods such as HMM, Kalman filters and autoregressive models on the other hand explicitly model the time progression of activity using a low dimensional dynamical process that moves through a state space. Many of these dynamical approaches have modeled neural activity that varies on slow time scales. Space-by-time NMF is able to describe the temporal dimension using firing patterns that vary either in very small time scales on the order of a few milliseconds or on a longer time scales in the order of hundreds of milliseconds or using a combination of fast and slow varying patterns. As we saw in our three datasets, the method can specify how these fast and slow components are recruited in individual trials. We found that such identified temporal patterns can be crucial components of the population code in the retina. Sensory neurons can code natural stimuli in very short time scales in the order of milliseconds (Kayser et al., 2010; Panzeri et al., 2010b; Luczak et al., 2015) and can use neural codes that multiplex different temporal scales (Panzeri et al., 2010b; Zuo et al., 2015). Thus space-by-time NMF is a promising method to describe temporal activity over multiple time scales in sensory populations.

4.2 Background of NMF as a method to study population activity

NMF has been previously applied as spatiotemporal NMF in a few studies (Kim et al., 2005; Overduin et al., 2015; Wei et al., 2015). However, it has not yet been widely used as a method to analyze large scale neural datasets. From our work reported in this thesis and from (Onken et al., In preparation), we found that spatiotemporal NMF is less data robust than space-by-time NMF, spatiotemporal versions of NMF, PCA, ICA and FA. Our contribution to this work is to introduce space-by-time NMF as a method that can separate space and time. We found that the introduction of the separability in space and time made a great improvement in the ability of using NMF as a method to extract salient single-trial information from large scale datasets which

may establish NMF as a general method for neural analysis. NMF is a well established method in the machine learning community (Cichocki et al., 2009) and our algorithms could be improved further using optimization methods available from these studies.

4.3 Modeling neural variability using space-by-time NMF

The initial space-by-time NMF model we introduced in chapter 2 assumes that the variability in the spike counts are Gaussian distributed. Spike counts are often modeled as Gaussian distributed in analytical studies mainly for mathematical tractability. In chapter 3, we extended the space-by-time NMF update rules to model sub-Poisson, Poisson and supra-Poisson variability (Tolhurst et al., 1983; van Steveninck et al., 1997; Shadlen and Newsome, 1998; Oram et al., 1999; Kara et al., 2000; Maimon and Assad, 2009). This is an important step when introducing a new method to analyze population responses. On the one hand a primary requirement of a model is to be able to account for the statistical properties of the dataset. On the other hand, the informativeness of the low-dimensional representation extracted by a model does not necessarily improve because the model has a better capacity to describe the data (Santhanam et al., 2009).

Our new update rules model spike counts that are in sub-Poisson, Poisson and supra-Poisson regimes. However, the neuronal variability is more heterogeneous such that the same neuron could change its spiking statistics over a single trial, often decreasing the variability shortly after the stimulus onset (Churchland et al., 2010). A single space-by-time NMF update rule may not be able to model these changes across a single trial. However, the results from our empirical data indicate that the performance difference between the Gaussian update rules and the Poisson update rule was small in general, in the range of 5 - 6%. The neural variability in our datasets decreased to sub-Poisson range at the stimulus onset (Figure 3.9A), as observed across brain regions (Churchland et al., 2010). According to our statistical simulations in chapter 3, the performance of both Gaussian and Poisson space-by-time NMF update rules become similar as the spike counts become more reliable in the sub-Poisson region. Thus, Poisson rules are expected to have comparable or better performance to Gaussian rules when spike counts have both Poisson and sub-Poisson variability. This was the case for most of our analyses on auditory cortical neurons to click sequences and somatosensory neurons to whisker deflections. However, when we analyzed auditory responses to long tones, Gaussian rules performed better than Poisson rules. We think that this could be because of the module structure extracted by the two rules. Poisson modules had a higher clustering level in the spatial modules and a slightly broader peak in the temporal module describing the onset response compared to Gaussian rules. The first spike latencies of auditory neurons vary depending on the frequency and the intensity of the tone (Schreiner et al., 2011). Our analysis on somatosensory neural responses suggests that the temporal module describing the onset response of the Poisson rules for long tones would be less sensitive to the precise timing of the first-spike latencies of the neurons grouped into the spatial module and would be more sensitive to computing the spike counts of the neurons. Gaussian

spatial modules that have a single neuron as a spatial module could be expected to better account for the fine differences in first spike latencies to different tones at different intensities. This could give rise to a higher performance of Gaussian rules if the spike onsets carry information that is not present in the remaining response.

We further note that by using the decoding performance as the performance measure to compare Gaussian and Poisson rules, we do not directly measure how well the update rules model the data. Rather, we measure the ability of the extracted low dimensional representations to represent stimuli and implicitly assume that by improving the ability of the update rules to model the underlying spike count variability, the extracted modules would have a better representation of stimuli. Since the representation is low dimensional, this may not necessarily be the case if the optimal update rules extract modules that can better account for the global variability in the dataset compared to Gaussian rules, while the variability carries less information about the stimuli. Thus, whether or not the optimal update rule would perform better than Gaussian update rules would depend on the firing characteristics of the neurons and how many of them carry information about the stimuli. Our results indicate that both Gaussian and the optimal rule are able to extract informative low-dimensional representations and the optimal rule performs slightly better in general. As we discussed in chapter 3, considering the performance improvement that can be achieved when using the optimal update rule with other factors such as computational cost and the purpose of the analysis, we could recommend to first use either Gaussian or Poisson rules that have lower computational cost. Decision of whether or not to further optimize using binomial and negative binomial rules could be taken based on the outcome of the analysis.

4.4 Effect of outliers

Using the beta family of divergences (Mihoko and Eguchi, 2002; Cichocki et al., 2011) showed that KL divergence has a lower dissimilarity for the reconstructed data points with high values compared to the Euclidean distance and thus it could be more prone to outliers compared to the Euclidean distance. This means that Poisson rules would be more susceptible to be biased by outliers compared to Gaussian rules. This is an effect that is opposite to our findings and interpretations which indicate that Poisson rules are more robust in terms of the level of clustering in the spatial modules compared to Gaussian rules. However, in general NMF literature, outliers are defined to be data points that have a low probability of occurrence (Fujisawa and Eguchi, 2008; Eguchi and Kato, 2010). In our work we defined outliers as neurons that have high firing rates, not isolated time bins that have unusually high spike counts. Furthermore, when spike counts are binned into small time bins, the value of the spike counts of a high firing neuron in individual time bins can be comparable to those for a low firing neuron. High firing neurons tend to bias the spatial modules because they have many time bins with low spike counts, not because they have one or more time bins with spuriously high spike counts.

4.5 Space-by-time NMF to study population coding in space and time

We used space-by-time NMF to analyze three datasets containing neurons encoding vision, audition and touch. We found that in all three sensory modalities neural responses in spatial and temporal dimensions together carried more information than using either of the dimensions separately. Moreover, the temporal dimension played a role in the population code. Now we will discuss these results in detail.

When analyzing responses to flashed images (natural images and gratings differing by spatial phase, we found that, while differences in image features on coarser spatial scales could be discriminated based on spike rates defined on long windows, both for artificial and natural images the information about image details on a finer spatial scale could only be recovered from the precise spike times of RGCs on as 10 ms scale, but not from spike rates. Given that the flashed images were held fixed for the stimulus presentation time, this represents evidence for temporal encoding (Theunissen et al., 1996; Panzeri et al., 2010b), that is, for the conversion of non-temporal (spatial) information into a sequence of spike times. This finding is consistent with an earlier theory that spike timing information of RGCs reflects local differences in stimulus intensities (Rullen and Thorpe, 2001; VanRullen and Thorpe, 2002) and that this could be a primary cause of temporal encoding of visual information, complementing other sources such as fixational eye movements (Rucci and Victor, 2015).

Notably, first-spike latencies turned out to be a key part of spike timing information. Although spike rates were sufficient to encode well coarse image information, this coarse image information was also almost entirely available in first-spike latencies. Moreover, latencies carried more information than spike rates about fine image features, always carried a large proportion of the total information contained in population spike trains, and were decodable without needing an external stimulus time reference. These results corroborate the idea that latencies form a dominant part of the retinal neural code for visual images allowing for rapid encoding of large amounts of visual information (Rullen and Thorpe, 2001; VanRullen and Thorpe, 2002; Gollisch and Meister, 2008).

First-spike latencies had been shown to be a key component of the retinal code for fine details of artificial stimuli in small populations of few cells (Gollisch and Meister, 2008; Gütig et al., 2013). Our results extended this previous work in two important ways. First, we showed that spike timing and latencies are important also for coding fine spatial information in natural images. Second, the demonstration that latencies can be read out from larger populations of tens of cells shows that fluctuations of latencies are sufficiently robustly coordinated across tens of cells to underlie robust image coding (something that was shown to hold only for cell pairs in (Gollisch and Meister, 2008)), and that the information carried by latency of one neuron is not redundant with the information carried by rate of another neuron. These results, which are important for establishing a key role of latencies in population activity, could be achieved because of the data robustness and the effectiveness of space-by-time NMF to capture

information from limited datasets. This property was key to accurately compare first-spike latency information with the total information carried by this population both along the space and time dimensions.

Spike trains of somatosensory neurons we analyzed consisted of combining recordings carried out in different experimental sessions. The same dataset was previously analyzed in several studies (Panzeri et al., 2001; Petersen et al., 2001; Panzeri et al., 2003; Panzeri and Diamond, 2010). (Panzeri et al., 2001) studied responses of single neurons from of neurons recorded from the barrel column D2 using an information theoretic framework. They found that the precise timing of spikes adds 44% extra information on which whisker was stimulated to the information content in the spike counts. The first spike latency information contributed 83% of the total information in the spike train. (Petersen et al., 2001) analyzed the how pairs of neurons in all three barrel-columns encoded whisker location. They found that 82 - 85 % of the total information was coded using spike timing. The first spike latency in pairs of neurons from D2 was 91 ± 7% from the total information while that of pairs of neurons taken across barrel columns was 89 ± 15 %. Neurons within the same barrel column were found to code redundantly while neurons in neighboring barrel columns coded independently. Consistent with these findings, we found in chapter 3 that at the population level, the first spike-time latency has the same or slightly higher level of decoding performance about whisker location as when considering the full information available in both spatial and temporal dimensions in the full spike trains. Our results further indicated that the information in the first spike lies in relative time differences between first spikes of different neurons. This observation has been made using the same dataset as in (Panzeri and Diamond, 2010). Our analysis is limited in the sense that the neural population was non-simultaneously recorded. The average pair-wise noise correlations between any two neurons in the population were very small. We could not estimate the correlations between neurons from the same recording session since we did not have the details of the recording sessions. However, according to (Petersen et al., 2001), the average noise correlation between neurons in the same column is 0.25 and (Panzeri et al., 2002) indicate that only less than 2 % of the total information is lost when decoding stimuli using the responses recorded from D3 barrel column with a decoder that has no knowledge of the correlations. Thus this suggests that the first-spike is a very important component of the neural population code in the barrel cortex coding whisker location. Our analysis using the second and subsequent spikes indicated that the firing rate of these remaining spikes contain 79% of the full discrimination capacity. This suggests that the neural code in the barrel cortex first code whisker locations using relative latencies in the first spikes and then redundantly code the stimuli using a rate code.

When analyzing auditory responses, we found that the spatial dimension is the key contributor to code tone frequencies. This is as expected given the tonotopical organization of the auditory cortex (Schreiner et al., 2011). Our results also indicated that the first spike latency was informative about long tones having 65% as the decoding performance obtained using all spikes. We further identified that the absolute times of the first spikes was 2.3 times more informative

about tones than the temporal order in which neurons discharge the first spike. Comparing this result with the reports from previous studies that have investigated first spike latency is complicated because of the differences in the species, anesthetic state and the type of stimuli used and these parameters can affect the first spike time latency (Camalier et al., 2012). Our finding that the order in which neurons discharge the first spike carries 17% decoding performance and thus has a small amount of information about tone frequency and intensity is consistent with the frequency selectivity of the first spike latency reported by (Imaizumi et al., 2004; Carrasco and Lomber, 2011; Scott et al., 2011) where mainly the first spike latency increases when moving anterior to posterior along the rostro-caudal axis of the auditory cortex. It is also corroborated by (Schreiner et al., 1997; Schreiner et al., 2011) which report that the first spike latency decreases when the sound pressure level increases. However, the first spike latency also displays variability within a caudo-rostral level. This could because it decreases with the firing rate (Kilgard and Merzenich, 1999; Kajikawa et al., 2005; Pienkowski and Harrison, 2005). Furthermore, the first latency can vary depending on the response duration (whether the response is phasic or sustained) as well as the position of the tone frequency in the frequency response area of the neuron (Schreiner et al., 2011). These could be the reasons for our finding that the absolute first spike latency differences have nearly 2.3 times the ability to discriminate long tones compared to just considering the order in which neurons discharge their first spikes. We further found that the first spike time information is largely redundantly coded in the later spikes. (Kajikawa et al., 2005) makes the same observation about the first spike latency and the firing rate in A1 of the Marmoset monkey. Thus our results are consistent with those that have been made previously mainly using single neuron analysis. Our findings using space-by-time NMF indicate that these modulations of the first spike time also exist in simultaneously recorded population of tens of neurons and consistently between two different brain states in rat A1. One limitation of our findings could be that they are from only one recording session. Further experimental work could verify these results using multiple recordings.

Finally, we find that the first spike latency is recruited to different extents to code different stimuli in different sensory system consistent with previous studies (VanRullen et al., 2005). This information is typically redundantly coded in firing rates. Due to the ability of space-by-time NMF we were able to quantify the information encoded in the full spike trains at the level of large populations in order to verify the differential contribution of first spike latencies to the neural population code.

To conclude, we introduced and refined a new method, space-by-time NMF to analyze large neural recordings. This method is able to identify biologically meaningful and informative patterns from spatial and temporal dimensions. We demonstrated that the method was able to give insight into how neurons in the populations we analyzed coded information in space and time concurrently as well as separately. These results suggest that this technique may become a very useful methodology to analyze the neural population code.

4.6 Future directions

NMF is a well established data analysis method in the machine learning community. Thus space-by-time NMF could be further improved by integrating knowledge available from NMF analysis done on other large scale datasets. For example, the sparseness of the derived matrices is typically controlled by adding a regularization term to the dissimilarity measure (Cichocki et al., 2009). The L1 matrix norm is commonly used to enforce regularization. The regularization term is weighted by a regularization coefficient that controls the degree of sparseness. The framework defined using Bregman divergences supports the inclusion of this term.

Space-by-time NMF is based on the assumption that the activity of a single trial is a linear combination of a set of space-time modules. As we discussed in chapter 2, this is a similar concept to that of packet-based communication in cortex where a set of stereotyped spike patterns a recruited in the responses to different stimuli (Luczak et al., 2009; Luczak et al., 2013; Luczak et al., 2015). If this is the case, and if the patterns are separable in space and time (Onken et al., 2009), then this suggests that space-by-time NMF can be applied to identify the functional organization of neurons by means of spatial modules.

Appendix 1: Proof of the equivalence of minimization of the Bregman divergence and the maximum likelihood estimation of the mean for binomial and negative binomial update rules

Binomial rules

If the spike count r_{tn} of neuron n in time bin t, r_{tn}, is a binomially distributed random variable, its probability mass function is given by (Equation 3.7),

$$p\left(r_{tn};\lambda_{tn},n\right)=\binom{n}{r_{tn}}\left(\frac{\lambda_{tn}}{n}\right)^{r_{tn}}\left(\frac{n-\lambda_{tn}}{n}\right)^{n-r_{tn}} \tag{1.1}$$

where λ_{tn} is the mean spike count and the parameter n specifies the regularity of the spike counts under the model. We assume that n is a constant for all N neurons during all T time bins.

Assuming that the spike count of each time bin is independent, the negative log likelihood L of the data matrix $\mathbf{R}$ is given by,

$$-\log L=-\log\prod_{NT}p\left(r_{tn};\lambda_{tn},n\right)=-\sum_{NT}\log p\left(r_{tn};\lambda_{tn},n\right)$$

$$-\log p\left(r_{tn};\lambda_{tn},n\right)=-\log\left[\binom{n}{r_{tn}}\left(\frac{\lambda_{tn}}{n}\right)^{r_{tn}}\left(\frac{n-\lambda_{tn}}{n}\right)^{n-r_{tn}}\right]$$

It can be shown that,

$$-\log p\left(r_{tn};\lambda_{tn},n\right)=r_{tn}\log\left(\frac{r_{tn}}{\lambda_{tn}}\right)+\left(n-r_{tn}\right)\log\left(\frac{n-r_{tn}}{n-\lambda_{tn}}\right)+K'$$

where K' contains terms independent of λ_{tn}.

Therefore,

$$-\log L=\sum_{NT}r_{tn}\log\left(\frac{r_{tn}}{\lambda_{tn}}\right)+\left(n-r_{tn}\right)\log\left(\frac{n-r_{tn}}{n-\lambda_{tn}}\right)+K \tag{1.2}$$

where K contains terms independent of λ_{tn}.

The Bregman divergence $D_{\varphi,Bin}\left(r_{tn},\hat{r}_{tn}\right)$ between r_{tn} and its reconstruction $\hat{r}_{tn}$ (when $r_{tn} \le n-1$ and $\hat{r}_{tn} \le n-1$) is (Equation 3.15),

$$D_{\varphi,Bin}\left(r_{tn},\hat{r}_{tn}\right) = r_{tn}\log\left(\frac{r_{tn}}{\hat{r}_{tn}}\right) + \left(n-r_{tn}\right)\log\left(\frac{n-r_{tn}}{n-\hat{r}_{tn}}\right) + K \qquad (1.3)$$

When $\hat{r}_{tn} = \lambda_{tn}$, the maximum likelihood estimation of λ_{tn} using Equation 1.2 is equivalent to minimizing the Bregman divergence between r_{tn} and λ_{tn} using Equation 1.3.

Negative binomial rules

If the spike count r_{tn} of neuron in time bin t, r_{tn}, is a binomially distributed random variable, its probability mass function is given by (from Equation 3.6),

$$p\left(r_{t,n};\lambda_{t,n},v\right) = \binom{r_{t,n}+v-1}{r_{t,n}} \frac{v^{v}\lambda_{t,n}^{r_{t,n}}}{\left(\lambda_{t,n}+v\right)^{r_{t,n}+v}} \qquad (1.4)$$

where λ_{tn} is the mean spike count and parameter v specifies the dispersion of the spike counts under the model. We assume that this parameter v is a constant for all N neurons during all T time bins.

Assuming that the spike count of each time bin is independent, the negative log likelihood L of the data matrix $\mathbf{R}$ is given by,

$$-\log L = -\log \prod_{NT} p\left(r_{tn};\lambda_{tn},v\right) = -\sum_{NT} \log p\left(r_{tn};\lambda_{tn},v\right)$$

$$-\log p\left(r_{tn};\lambda_{tn},v\right) = -\log\left[\binom{r_{tn}+v-1}{r_{tn}} \frac{v^{v}\lambda_{tn}^{r_{tn}}}{\left(\lambda_{tn}+v\right)^{v+r_{tn}}}\right]$$

It can be shown that,

$$-\log L = \sum_{NT} r_{tn}\log\left(\frac{r_{tn}}{\lambda_{tn}}\right) + \left(v+r_{tn}\right)\log\left(\frac{v+\lambda_{tn}}{v+r_{tn}}\right) + K' \qquad (1.5)$$

where K' contains terms independent of λ_{tn}.

The Bregman divergence $D_{\varphi,NBin}\left(r_{tn},\hat{r}_{tn}\right)$ between r_{tn} and its reconstruction $\hat{r}_{tn}$ is (Equation 3.16),

$$D_{\varphi,NBin}\left(r_{t,n},\hat{r}_{t,n}\right)=r_{t,n}\log\left(\frac{r_{t,n}}{\hat{r}_{t,n}}\right)+\left(v+r_{t,n}\right)\log\left(\frac{v+\hat{r}_{t,n}}{v+r_{t,n}}\right)+K \qquad (1.6)$$

When $\hat{r}_{tn}=\lambda_{tn}$, the maximum likelihood estimation of λ_{tn} using Equation 1.5 is equivalent to minimizing the Bregman divergence between r_{tn} and λ_{tn} using Equation 1.6.

Appendix 2: Changes in response properties of auditory cortical neurons to state change

We identified a change in the synchronization state while auditory cortical neurons responded to long tones. These neurons were stimulated with long tones that were of 500 ms duration at 1200 ms intervals. We examined the spontaneous activity of 100 ms period starting 200 ms after the previous tone was removed. We first computed the power in the multi unit activity (MUA) in these periods using multitaper power spectral analysis. Figure A1.A shows the ratio of the MUA power between 0 - 5 Hz to MUA power between 0 - 50 Hz. The mean ratio for trials 1 - 40 was significantly higher compared to the ratio for trials 61 - 100 (two tailed t-test, $p<0.05$). The state of the cortex is more synchronized when this ratio is high (Curto et al., 2009), which indicates that the early trials were in a higher synchronized state compared to the latter trials in the experiment. Next we formed a data matrix that contained the total spike counts of each neuron during the 1 s period. Then we applied two factor NMF to extract two spatial modules from the spike count matrix (Figure A1.2). The coefficients corresponding to the two spatial modules showed a clear change in their activation level between trials 40 - 60. The first spatial module declined its activation level while the second module increased its activation level . Spiking response of the population during each state is shown in Figure A1.C. The activity in the trial from the synchronized state shows activity structured with up-down states while the trial from the desynchronized state contains unstructured activity.

Since we found that the neurons grouped together into a spatial module have higher signal and noise correlations compared to randomly selected pairs of neurons, we investigated the signal and noise correlation structure across states. We computed signal correlation as the Pearson correlation coefficient between the mean spike counts (using the total spike count) of neurons across tones. Noise correlations were computed as the Pearson correlation coefficient between the differences in the spike counts from the mean spike counts for each neuron pair. The variation of noise correlations between the two states are shown in Figure A2. Correlation values were summarized for pairs of neurons recorded from each tetrode. Signal correlations in pairs of neurons recorded from tetrodes one and two decreased when the state became more synchronized while they increased slightly in pairs of neurons recorded from tetrodes seven and eight (two tailed t-test, $p < 0.05$). Noise correlations in pairs of neurons recorded from tetrodes one and two also decreased after desynchronization while they were unchanged in tetrode seven and increased slightly in tetrode eight. Noise correlations between pairs of neurons recorded from tetrodes 1 - 4 increased with the geometric mean firing rate during the synchronized state. However, it reversed and the noise correlations decreased with the geometric mean firing rate when the state was desynchronized. On the other hand, pairs of neurons recorded from tetrodes 6 - 8 mainly decreased noise correlations with the increase in the geometric mean firing rate. This indicate that signal and noise correlations changed across the state change as have been

reported previously (Pachitariu et al., 2015). Further, their structure changed differential across the cortex when moving from anterior to posterior.

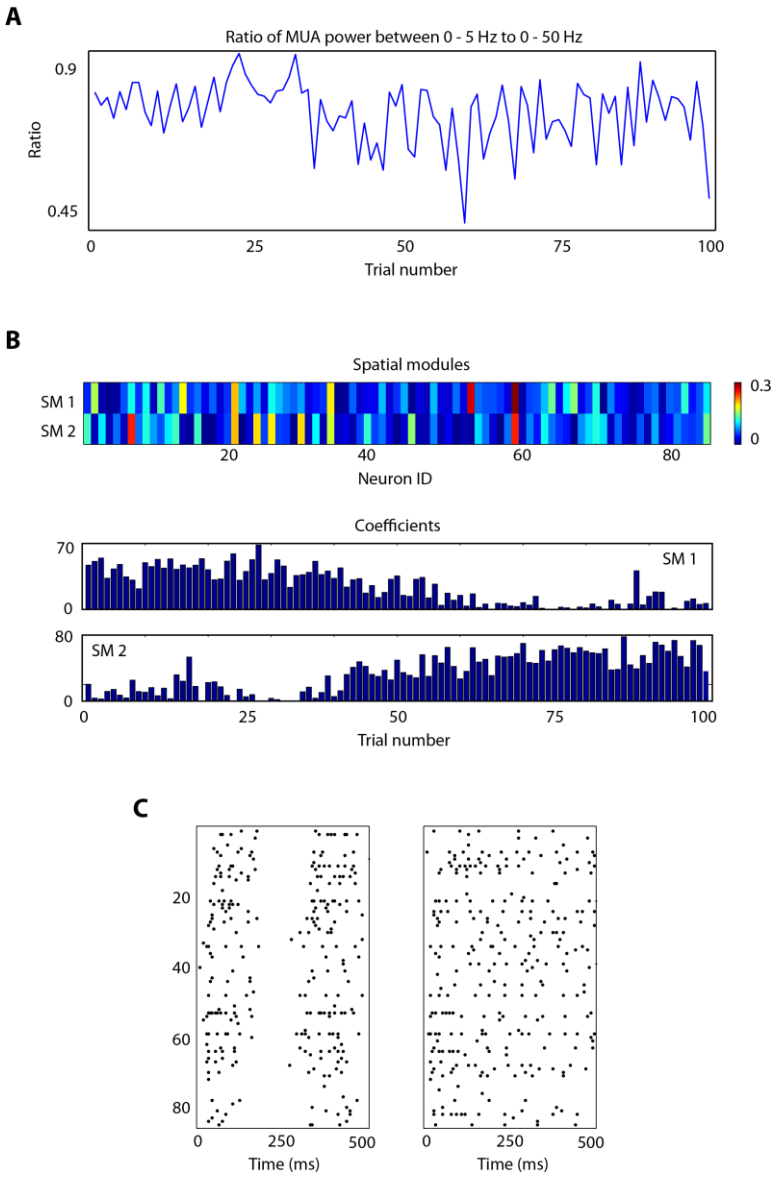

Figure A1. 1: Identification of the change in the synchronization state. A: the ratio of the power in the multi unit activity between 0 - 5 Hz to that of 0 - 50 Hz across the experiment. The trials are ordered according to the performed sequence. B: Two factor NMF is able to identify the change in the synchronization state using two spatial modules. Spatial modules (SM) of the decomposition are shown in the top figure and their activation levels are shown in by the coefficients in the bottom figure. C: Raster responses of spiking activity in a representative trial from the higher synchronized state (left) and the lower synchronized state (right).

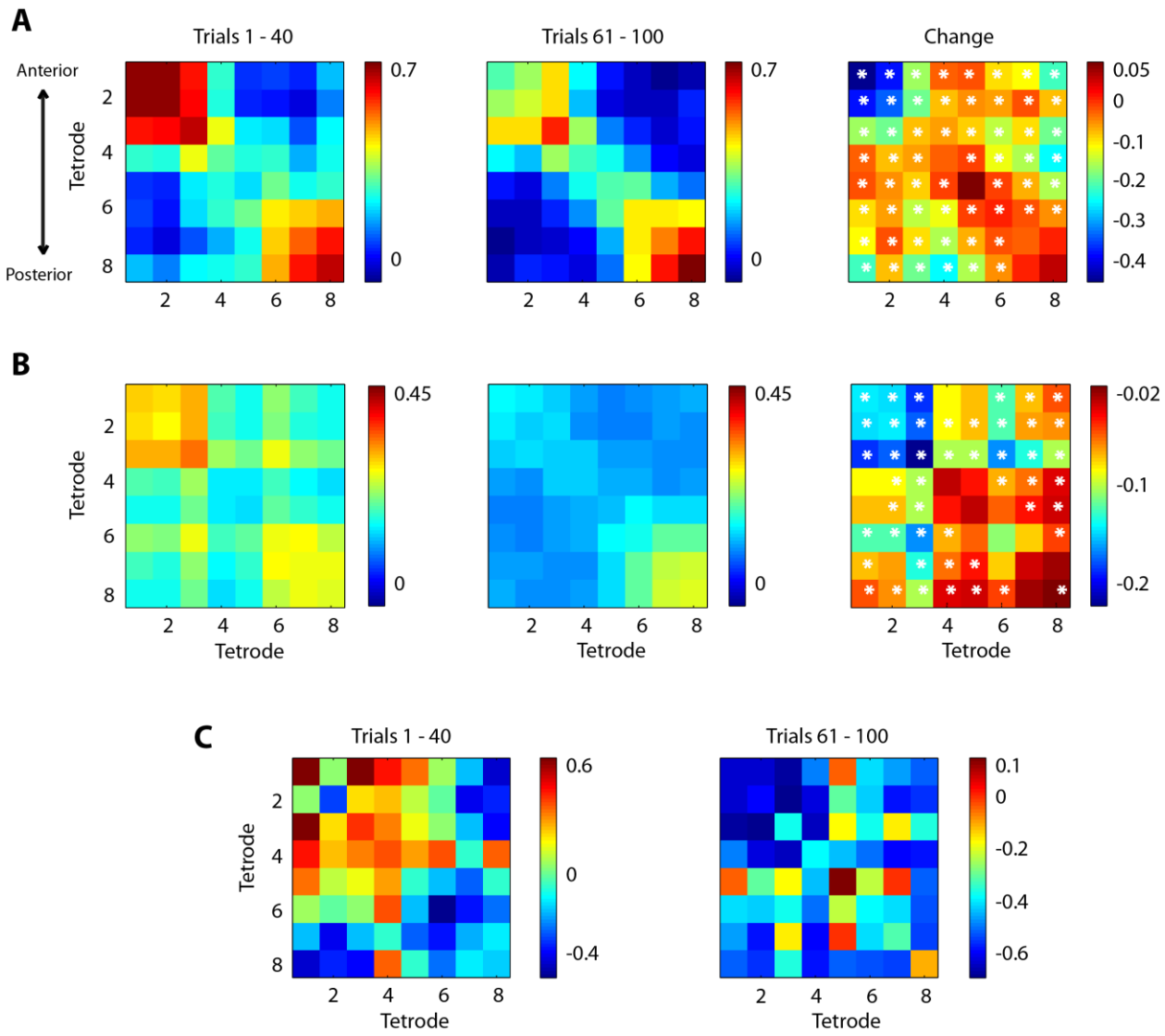

Figure A1. 2: Changes in the correlation structure with the state change. A: Signal correlations between pairs of neurons recorded from each tetrode. The first figure shows signal correlations between neurons in trials 1 - 40 while the middle figure shows the same for trials 61 - 100. The figure on the right indicates the change in signal correlation when the state became desynchronized. * show significant changes (two-tailed t-test, $p < 0.05$). B: Noise correlations averaged across tones in pairs of neurons recorded from each tetrode in the same layout as in A. C: Mean Pearson correlation coefficient between noise correlations averaged between pairs and their geometric firing rate.

References

Abbott L, Dayan P (1999) The effect of correlated variability on the accuracy of a population code. Neural computation 11:91-101.

Abbott L, Regehr WG (2004) Synaptic computation. Nature 431:796-803.

Abeles M, Bergman H, Gat I, Meilijson I, Seidemann E, Tishby N, Vaadia E (1995) Cortical activity flips among quasi-stationary states. Proceedings of the National Academy of Sciences 92:8616-8620.

Afshar A, Santhanam G, Byron MY, Ryu SI, Sahani M, Shenoy KV (2011) Single-trial neural correlates of arm movement preparation. Neuron 71:555-564.

Ahrens MB, Li JM, Orger MB, Robson DN, Schier AF, Engert F, Portugues R (2012) Brain-wide neuronal dynamics during motor adaptation in zebrafish. Nature 485:471-477.

Arabzadeh E, Panzeri S, Diamond ME (2004) Whisker vibration information carried by rat barrel cortex neurons. J Neurosci 24:6011-6020.

Armstrong-James M, Fox K (1987) Spatiotemporal convergence and divergence in the rat S1 “barrel” cortex. Journal of Comparative Neurology 263:265-281.

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

Babadi B, Casti A, Xiao Y, Kaplan E, Paninski L (2010) A generalized linear model of the impact of direct and indirect inputs to the lateral geniculate nucleus. J Vis 10:22.

Banerjee A, Merugu S, Dhillon IS, Ghosh J (2005) Clustering with Bregman divergences. The Journal of Machine Learning Research 6:1705-1749.

Baum LE, Petrie T, Soules G, Weiss N (1970) A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. The annals of mathematical statistics 41:164-171.

Bell AJ, Sejnowski TJ (1997) The “independent components” of natural scenes are edge filters. Vision research 37:3327-3338.

Berry MW, Browne M (2005) Email surveillance using non-negative matrix factorization. Computational & Mathematical Organization Theory 11:249-264.

Bizley JK, Nodal FR, Nelken I, King AJ (2005) Functional organization of ferret auditory cortex. Cerebral Cortex 15:1637-1653.

Bollimunta A, Totten D, Ditterich J (2012) Neural dynamics of choice: single-trial analysis of decision-related activity in parietal cortex. The Journal of neuroscience 32:12684-12701.

Bourjaily MA, Miller P (2012) Dynamic afferent synapses to decision-making networks improve performance in tasks requiring stimulus associations and discriminations. Journal of neurophysiology 108:513-527.

Boutsidis C, Gallopoulos E (2008) SVD based initialization: A head start for nonnegative matrix factorization. Pattern Recognition 41:1350-1362.

Brasselet R, Panzeri S, Logothetis NK, Kayser C (2012) Neurons with stereotyped and rapid responses provide a reference frame for relative temporal coding in primate auditory cortex. The Journal of neuroscience 32:2998-3008.

Bregman LM (1967) The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. USSR computational mathematics and mathematical physics 7:200-217.

Brown EN, Kass RE, Mitra PP (2004) Multiple neural spike train data analysis: state-of-the-art and future challenges. Nat Neurosci 7:456-461.

Brown EN, Frank LM, Tang D, Quirk MC, Wilson MA (1998) A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. J Neurosci 18:7411-7425.

Brown GD, Yamada S, Sejnowski TJ (2001) Independent component analysis at the neural cocktail party. Trends in neurosciences 24:54-63.

Brunel N, Wang X-J (2003) What determines the frequency of fast network oscillations with irregular neural discharges? I. Synaptic dynamics and excitation-inhibition balance. Journal of neurophysiology 90:415-430.

Brunet J-P, Tamayo P, Golub TR, Mesirov JP (2004) Metagenes and molecular pattern discovery using matrix factorization. Proceedings of the National Academy of Sciences 101:4164-4169.

Buesing L, Macke JH, Sahani M (2012a) Spectral learning of linear dynamics from generalised-linear observations with application to neural population data. In: Advances in neural information processing systems, pp 1682-1690.

Buesing L, Macke JH, Sahani M (2012b) Learning stable, regularised latent models of neural population dynamics. Network: Computation in Neural Systems 23:24-47.

Buonomano DV (2000) Decoding temporal information: a model based on short-term synaptic plasticity. The Journal of neuroscience 20:1129-1141.

Buzsaki G (2004) Large-scale recording of neuronal ensembles. Nat Neurosci 7:446-451.

Buzsáki G, Stark E, Berényi A, Khodagholy D, Kipke DR, Yoon E, Wise KD (2015) Tools for probing local circuits: high-density silicon probes combined with optogenetics. Neuron 86:92-105.

Byron MY, Afshar A, Santhanam G, Ryu SI, Shenoy KV (2005) Extracting dynamical structure embedded in neural activity. In: Advances in neural information processing systems, pp 1545-1552.

Byron MY, Cunningham JP, Santhanam G, Ryu SI, Shenoy KV, Sahani M (2009) Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. In: Advances in neural information processing systems, pp 1881-1888.

Calabrese A, Schumacher JW, Schneider DM, Paninski L, Woolley SM (2011) A generalized linear model for estimating spectrotemporal receptive fields from responses to natural sounds. PLoS One 6:e16104.

Camalier CR, D'Angelo WR, Sterbing-D'Angelo SJ, Lisa A, Hackett TA (2012) Neural latencies across auditory cortex of macaque support a dorsal stream supramodal timing advantage in primates. Proceedings of the National Academy of Sciences 109:18168-18173.

Carandini M, Heeger DJ, Senn W (2002) A synaptic explanation of suppression in visual cortex. The Journal of neuroscience 22:10053-10065.

Carmona-Saez P, Pascual-Marqui RD, Tirado F, Carazo JM, Pascual-Montano A (2006) Biclustering of gene expression data by non-smooth non-negative matrix factorization. BMC bioinformatics 7:78.

Carrasco A, Lomber SG (2011) Neuronal activation times to simple, complex, and natural sounds in cat primary and nonprimary auditory cortex. Journal of neurophysiology 106:1166-1178.

Cavallari S, Panzeri S, Mazzoni A (2014) Comparison of the dynamics of neural interactions between current-based and conductance-based integrate-and-fire recurrent networks. Frontiers in neural circuits 8.

Cemgil AT (2009) Bayesian inference for nonnegative matrix factorisation models. Computational Intelligence and Neuroscience 2009.

Chance FS, Nelson SB, Abbott LF (1998) Synaptic depression and the temporal response characteristics of V1 cells. The Journal of neuroscience 18:4785-4799.

Chapin JK, Nicolelis MA (1999) Principal component analysis of neuronal ensemble activity reveals multidimensional somatosensory representations. Journal of Neuroscience Methods 94:121-140.

Chase SM, Young ED (2007) First-spike latency information in single neurons increases when referenced to population onset. Proceedings of the National Academy of Sciences 104:5175-5180.
Cheung VC, Tresch MC (2006) Non-negative matrix factorization algorithms modeling noise distributions within the exponential family. In: Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the, pp 4990-4993: IEEE.
Cheung VC, d'Avella A, Tresch MC, Bizzi E (2005) Central and sensory contributions to the activation and organization of muscle synergies during natural motor behaviors. The Journal of neuroscience 25:6419-6434.
Chichilnisky E (2001) A simple white noise analysis of neuronal light responses. Network: Computation in Neural Systems 12:199-213.
Chimoto S, Kitama T, Qin L, Sakayori S, Sato Y (2002) Tonal response patterns of primary auditory cortex neurons in alert cats. Brain research 934:34-42.
Churchland MM, Byron MY, Sahani M, Shenoy KV (2007) Techniques for extracting single-trial activity patterns from large-scale neural recordings. Current opinion in neurobiology 17:609-618.
Churchland MM, Cunningham JP, Kaufman MT, Foster JD, Nuyujukian P, Ryu SI, Shenoy KV (2012) Neural population dynamics during reaching. Nature 487:51-56.
Churchland MM, Byron MY, Cunningham JP, Sugrue LP, Cohen MR, Corrado GS, Newsome WT, Clark AM, Hosseini P, Scott BB (2010) Stimulus onset quenches neural variability: a widespread cortical phenomenon. Nature neuroscience 13:369-378.
Cichocki A, Cruces S, Amari S-i (2011) Generalized alpha-beta divergences and their application to robust nonnegative matrix factorization. Entropy 13:134-170.
Cichocki A, Zdunek R, Phan AH, Amari S-i (2009) Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation: John Wiley & Sons.
Clement EA, Richard A, Thwaites M, Ailon J, Peters S, Dickson CT (2008) Cyclic and sleep-like spontaneous alternations of brain state under urethane anaesthesia. PLoS One 3:e2004.
Clopath C, Büsing L, Vasilaki E, Gerstner W (2009) Connectivity reflects coding: A model of voltage-based spike-timing-dependent-plasticity with homeostasis. Nature.
Cohen MR, Maunsell JH (2009) Attention improves performance primarily by reducing interneuronal correlations. Nature neuroscience 12:1594-1600.
Comon P (1994) Independent component analysis, a new concept? Signal processing 36:287-314.
Cover TM, Thomas JA (2006) Elements of information theory: Wiley-interscience.
Crumiller M, Knight B, Yu Y, Kaplan E (2011) Estimating the amount of information conveyed by a population of neurons. Front Neurosci 5:90.
Cunningham JP, Byron MY (2014) Dimensionality reduction for large-scale neural recordings. Nature neuroscience.
Curto C, Sakata S, Marguet S, Itskov V, Harris KD (2009) A simple model of cortical dynamics explains variability and state dependence of sensory responses in urethane-anesthetized auditory cortex. The Journal of neuroscience 29:10600-10612.
d'Avella A, Saltiel P, Bizzi E (2003) Combinations of muscle synergies in the construction of a natural motor behavior. Nature neuroscience 6:300-308.
d Avella A, Tresch MC (2002) Modularity in the motor system: decomposition of muscle patterns as combinations of time-varying synergies. Advances in neural information processing systems 1:141-148.
Danóczy M, Hahnloser R (2006) Efficient estimation of hidden state dynamics from spike trains. Advances in neural information processing systems 18:227.
Dayan P, Abbott LF (2001) Theoretical neuroscience: Cambridge, MA: MIT Press.

de la Rocha J, Marchetti C, Schiff M, Reyes AD (2008) Linking the response properties of cells in auditory cortex with network architecture: cotuning versus lateral inhibition. The Journal of neuroscience 28:9151-9163.

Decharms RC, Zador A (2000) Neural representation and the cortical code. Annu Rev Neurosci 23:613-647.

Delis I, Berret B, Pozzo T, Panzeri S (2013) Quantitative evaluation of muscle synergy models: a single-trial task decoding approach. Frontiers in computational neuroscience 7.

Delis I, Panzeri S, Pozzo T, Berret B (2014) A unifying model of concurrent spatial and temporal modularity in muscle activity. Journal of neurophysiology 111:675-693.

Demirci O, Stevens MC, Andreasen NC, Michael A, Liu J, White T, Pearlson GD, Clark VP, Calhoun VD (2009) Investigation of relationships between fMRI brain networks in the spectral domain using ICA and Granger causality reveals distinct differences between schizophrenia patients and healthy controls. Neuroimage 46:419-431.

DeWeese MR, Wehr M, Zador AM (2003) Binary spiking in auditory cortex. The Journal of neuroscience 23:7940-7949.

Dhillon IS, Sra S (2005) Generalized nonnegative matrix approximations with Bregman divergences. In: Advances in neural information processing systems, pp 283-290.

Ding C, Li T, Peng W, Park H (2006) Orthogonal nonnegative matrix t-factorizations for clustering. In: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, pp 126-135: ACM.

Durstewitz D, Vittoz NM, Floresco SB, Seamans JK (2010) Abrupt transitions between prefrontal neural ensemble states accompany behavioral transitions during rule learning. Neuron 66:438-448.

Ecker AS, Berens P, Tolias AS, Bethge M (2011) The effect of noise correlations in populations of diversely tuned neurons. The Journal of neuroscience 31:14272-14283.

Ecker AS, Berens P, Cotton RJ, Subramaniyan M, Denfield GH, Cadwell CR, Smirnakis SM, Bethge M, Tolias AS (2014) State dependence of noise correlations in macaque primary visual cortex. Neuron 82:235-248.

Edelman GM, Gall WE, Cowan WM (1984) Dynamic aspects of neocortical function: John Wiley & Sons.

Eguchi S, Kato S (2010) Entropy and divergence associated with power function and the statistical application. Entropy 12:262-274.

Escola S, Fontanini A, Katz D, Paninski L (2011) Hidden markov models for the stimulus-response relationships of multistate neural systems. Neural computation 23:1071-1132.

Févotte C, Cemgil AT (2009) Nonnegative matrix factorizations as probabilistic inference in composite models. In: Signal Processing Conference, 2009 17th European, pp 1913-1917: IEEE.

Févotte C, Bertin N, Durrieu J-L (2009) Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis. Neural computation 21:793-830.

Franke F, Fiscella M, Sevelev M, Roska B, Hierlemann A, da Silveira RA (2016) Structures of neural correlation and how they favor coding. Neuron 89:409-422.

Fujisawa H, Eguchi S (2008) Robust parameter estimation with a small bias against heavy contamination. Journal of Multivariate Analysis 99:2053-2081.

Ganmor E, Segev R, Schneidman E (2011) Sparse low-order interaction network underlies a highly correlated and learnable neural population code. Proc Natl Acad Sci U S A 108:9679-9684.

Gat I, Tishby N, Abeles M (1997) Hidden Markov modelling of simultaneously recorded cells in the associative cortex of behaving monkeys. Network: Computation in Neural Systems 8:297-322.

Gauthier JL, Field GD, Sher A, Greschner M, Shlens J, Litke AM, Chichilnisky E (2009) Receptive fields in primate retina are coordinated to sample visual space more uniformly. PLoS Biol 7:e1000063.

Gawne TJ, Kjaer TW, Richmond BJ (1996) Latency: another potential code for feature binding in striate cortex. Journal of neurophysiology 76:1356-1360.

Georgopoulos AP, Kalaska JF, Caminiti R, Massey JT (1982) On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. The Journal of neuroscience 2:1527-1537.

Georgopoulos AP, Caminiti R, Kalaska JF, Massey JT (1983) Spatial coding of movement: a hypothesis concerning the coding of movement direction by motor cortical populations. Exp Brain Res Suppl 7:336.

Goel A, Buonomano DV (2014) Timing as an intrinsic property of neural networks: evidence from in vivo and in vitro experiments. Phil Trans R Soc B 369:20120460.

Gollisch T, Meister M (2008) Rapid neural coding in the retina with relative spike latencies. Science 319:1108-1111.

Granot-Atedgi E, Tkacik G, Segev R, Schneidman E (2013) Stimulus-dependent maximum entropy models of neural population codes. PLoS Comput Biol 9:e1002922.

Grun S (2009) Data-driven significance estimation for precise spike correlation. J Neurophysiol 101:1126-1140.

Grun S, Diesmann M, Aertsen A (2002) Unitary events in multiple single-neuron spiking activity: II. Nonstationary data. Neural Comput 14:81-119.

Grün S, Diesmann M, Aertsen A (2010) Unitary Event Analysis. In: Analysis of Parallel Spike Trains (Grün S, Rotter S, eds), pp 191-220: Springer US.

Guillamet D, Vitria J (2002) Classifying faces with nonnegative matrix factorization. In: Proc. 5th Catalan conference for artificial intelligence, pp 24-31.

Gütig R, Gollisch T, Sompolinsky H, Meister M (2013) Computing complex visual features with retinal spike times. PLoS One 8:e53063.

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

Haddad R, Lanjuin A, Madisen L, Zeng H, Murthy VN, Uchida N (2013) Olfactory cortical neurons read out a relative time code in the olfactory bulb. Nature neuroscience 16:949-957.

Harris KD (2013) Information Coding by Cortical Populations. Principles of Neural Coding:391.

Harris KD, Thiele A (2011) Cortical state and attention. Nat Rev Neurosci 12:509-523.

Heil P (2004) First-spike latency of auditory neurons revisited. Current opinion in neurobiology 14:461-467.

Hoffman K, McNaughton B (2002) Coordinated reactivation of distributed memory traces in primate neocortex. Science 297:2070-2073.

Horton JC, Adams DL (2005) The cortical column: a structure without a function. Philosophical Transactions of the Royal Society of London B: Biological Sciences 360:837-862.

Hoyer PO (2004) Non-negative matrix factorization with sparseness constraints. The Journal of Machine Learning Research 5:1457-1469.

Hu J, Si J, Olson BP, He J (2005) Feature detection in motor cortical spikes by principal component analysis. Neural Systems and Rehabilitation Engineering, IEEE Transactions on 13:256-262.

Hubel DH, Wiesel TN (1959) Receptive fields of single neurones in the cat's striate cortex. The Journal of physiology 148:574-591.

Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. The Journal of physiology 160:106.

Hyvärinen A, Oja E (2000) Independent component analysis: algorithms and applications. Neural networks 13:411-430.

Imaizumi K, Priebe NJ, Crum PA, Bedenbaugh PH, Cheung SW, Schreiner CE (2004) Modular functional organization of cat anterior auditory field. Journal of neurophysiology 92:444-457.

Ince RA, Panzeri S, Kayser C (2013) Neural codes formed by small and temporally precise populations in auditory cortex. J Neurosci 33:18277-18287.

Ince RA, Senatore R, Arabzadeh E, Montani F, Diamond ME, Panzeri S (2010) Information-theoretic methods for studying population codes. Neural Netw 23:713-727.
Janecek A, Tan Y (2011) Using population based algorithms for initializing nonnegative matrix factorization. In: Advances in Swarm Intelligence, pp 307-316: Springer.
Jaynes ET (1957) Information theory and statistical mechanics. Physical review 106:620.
Jazayeri M, Movshon JA (2006) Optimal representation of sensory information by neural populations. Nat Neurosci 9:690-696.
Johansson RS, Birznieks I (2004) First spikes in ensembles of human tactile afferents code complex spatial fingertip events. Nature neuroscience 7:170-177.
Jolliffe I (2002) Principal component analysis: Wiley Online Library.
Jones LM, Fontanini A, Sadacca BF, Miller P, Katz DB (2007) Natural stimuli evoke dynamic sequences of states in sensory cortical ensembles. Proceedings of the National Academy of Sciences 104:18772-18777.
Josic K, Shea-Brown E, Doiron B, de la Rocha J (2009) Stimulus-dependent correlations and population codes. Neural computation 21:2774-2804.
Kajikawa Y, de La Mothe L, Blumell S, Hackett TA (2005) A comparison of neuron response properties in areas A1 and CM of the marmoset monkey auditory cortex: tones and broadband noise. Journal of neurophysiology 93:22-34.
Kandaswamy U, Deng P-Y, Stevens CF, Klyachko VA (2010) The role of presynaptic dynamics in processing of natural spike trains in hippocampal synapses. The Journal of neuroscience 30:15904-15914.
Kara P, Reinagel P, Reid RC (2000) Low response variability in simultaneously recorded retinal, thalamic, and cortical neurons. Neuron 27:635-646.
Kayser C, Logothetis NK, Panzeri S (2010) Millisecond encoding precision of auditory cortex neurons. Proceedings of the National Academy of Sciences 107:16976-16981.
Kayser C, Montemurro MA, Logothetis NK, Panzeri S (2009) Spike-phase coding boosts and stabilizes information carried by spatial and temporal spike patterns. Neuron 61:597-608.
Kayser C, Wilson C, Safaai H, Sakata S, Panzeri S (2015) Rhythmic Auditory Cortex Activity at Multiple Timescales Shapes Stimulus–Response Gain and Background Firing. The Journal of neuroscience 35:7750-7762.
Kemere C, Santhanam G, Byron MY, Afshar A, Ryu SI, Meng TH, Shenoy KV (2008) Detecting neural-state transitions using hidden Markov models for motor cortical prostheses. Journal of neurophysiology 100:2441-2452.
Kilavik BE, Roux S, Ponce-Alvarez A, Confais J, Grun S, Riehle A (2009) Long-term modifications in motor cortical dynamics induced by intensive practice. J Neurosci 29:12653-12663.
Kilgard MP, Merzenich MM (1999) Distributed representation of spectral and temporal information in rat primary auditory cortex. Hearing research 134:16-28.
Kim S-P, Rao YN, Erdogmus D, Sanchez JC, Nicolelis MA, Principe JC (2005) Determining patterns in neural activity for reaching movements using nonnegative matrix factorization. EURASIP Journal on Applied Signal Processing 2005:3113-3121.
Kim T, Oh WC, Choi JH, Kwon H-B (2016) Emergence of functional subnetworks in layer 2/3 cortex induced by sequential spikes in vivo. Proceedings of the National Academy of Sciences:201513410.
Klug A, Borst JGG, Carlson BA, Kopp-Scheinpflug C, Klyachko VA, Xu-Friedman MA (2012) How do short-term changes at synapses fine-tune information processing? The Journal of neuroscience 32:14058-14063.
Klyachko VA, Stevens CF (2006) Excitatory and feed-forward inhibitory hippocampal synapses work synergistically as an adaptive filter of natural spike trains. PLoS Biol 4:e207.

Ko H, Hofer SB, Pichler B, Buchanan KA, Sjöström PJ, Mrsic-Flogel TD (2011) Functional specificity of local synaptic connections in neocortical networks. Nature 473:87-91.
Kobak D, Brendel W, Constantinidis C, Feierstein CE, Kepecs A, Mainen ZF, Romo R, Qi X-L, Uchida N, Machens CK (2014) Demixed principal component analysis of population activity in higher cortical areas reveals independent representation of task parameters. arXiv preprint arXiv:14106031.
Koch C (2004) Biophysics of computation: information processing in single neurons: Oxford university press.
Kulkarni JE, Paninski L (2007) Common-input models for multiple neural spike-train data. Network: Computation in Neural Systems 18:375-407.
Latham PE, Roudi Y, eds (2013) Role of correlations in population coding: CRC Press.
Laubach M, Shuler M, Nicolelis MA (1999a) Independent component analyses for quantifying neuronal ensemble interactions. Journal of Neuroscience Methods 94:141-154.
Laubach M, Shuler M, Nicolelis MAL (1999b) Independent component analyses for quantifying neuronal ensemble interactions. Journal of Neuroscience Methods 94:141-154.
Laubach M, Wessberg J, Nicolelis MA (2000) Cortical ensemble activity increasingly predicts behaviour outcomes during learning of a motor task. Nature 405:567-571.
Laurent G (1999) A systems perspective on early olfactory coding. Science 286:723-728.
Lawhern V, Wu W, Hatsopoulos N, Paninski L (2010) Population decoding of motor cortical activity using a generalized linear model with hidden states. Journal of Neuroscience Methods 189:267-280.
Lebedev MA, Mirabella G, Erchova I, Diamond ME (2000) Experience-dependent plasticity of rat barrel cortex: redistribution of activity across barrel-columns. Cerebral Cortex 10:23-31.
Lee DD, Seung HS (1999) Learning the parts of objects by non-negative matrix factorization. Nature 401:788-791.
LeVan P, Tyvaert L, Moeller F, Gotman J (2010) Independent component analysis reveals dynamic ictal BOLD responses in EEG-fMRI data from focal epilepsy patients. Neuroimage 49:366-378.
Levy RB, Reyes AD (2011) Coexistence of lateral and co-tuned inhibitory configurations in cortical networks.
Lewis JE, Kristan WB, Jr. (1998) Representation of touch location by a population of leech sensory neurons. J Neurophysiol 80:2584-2592.
Lopes-dos-Santos V, Ribeiro S, Tort AB (2013) Detecting cell assemblies in large neuronal populations. Journal of Neuroscience Methods 220:149-166.
Lopes-dos-Santos V, Conde-Ocazionez S, Nicolelis MA, Ribeiro ST, Tort AB (2011) Neuronal assembly detection and cell membership specification by principal component analysis. PLoS One 6:e20996.
Luczak A, Barthó P, Harris KD (2009) Spontaneous events outline the realm of possible sensory responses in neocortical populations. Neuron 62:413-425.
Luczak A, Bartho P, Harris KD (2013) Gating of sensory input by spontaneous cortical activity. The Journal of neuroscience 33:1684-1695.
Luczak A, McNaughton BL, Harris KD (2015) Packet-based communication in the cortex. Nature Reviews Neuroscience.
Maimon G, Assad JA (2009) Beyond Poisson: increased spike-time regularity across primate parietal cortex. Neuron 62:426-440.
Makarov VA, Makarova J, Herreras O (2010) Disentanglement of local field potential sources by independent component analysis. Journal of computational neuroscience 29:445-457.
Makeig S, Jung T-P, Bell AJ, Ghahremani D, Sejnowski TJ (1997) Blind separation of auditory event-related brain responses into independent components. Proceedings of the National Academy of Sciences 94:10979-10984.

Maldonado P, Babul C, Singer W, Rodriguez E, Berger D, Grun S (2008) Synchronization of neuronal responses in primary visual cortex of monkeys viewing natural images. J Neurophysiol 100:1523-1532.

Mante V, Sussillo D, Shenoy KV, Newsome WT (2013) Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature 503:78-84.

Marguet SL, Harris KD (2011) State-dependent representation of amplitude-modulated noise stimuli in rat auditory cortex. The Journal of neuroscience 31:6414-6420.

Masland RH (2012) The neuronal organization of the retina. Neuron 76:266-280.

Masquelier T, Guyonneau R, Thorpe SJ (2008) Spike timing dependent plasticity finds the start of repeating patterns in continuous spike trains. PLoS One 3:e1377.

Maunsell JH, Van Essen DC (1983) Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. J Neurophysiol 49:1127-1147.

Mayhew T (1991) Anatomy of the Cortex: Statistics and Geometry. Journal of anatomy 179:203.

Mazor O, Laurent G (2005) Transient dynamics versus fixed points in odor representations by locust antennal lobe projection neurons. Neuron 48:661-673.

Mazzoni A, Panzeri S, Logothetis NK, Brunel N (2008) Encoding of naturalistic stimuli by local field potential spectra in networks of excitatory and inhibitory neurons. PLoS Comput Biol 4:e1000239-e1000239.

Mazzoni A, Whittingstall K, Brunel N, Logothetis NK, Panzeri S (2010) Understanding the relationships between spike rate and delta/gamma frequency bands of LFPs and EEGs using a local cortical network model. Neuroimage 52:956-972.

Mazzoni A, Yanfang Z, Notaro G, Panzeri S, Diamond ME (2013) Spike timing in rat somatosensory cortex contributes to behavior. BMC Neuroscience 14:P109.

Meyer HS, Egger R, Guest JM, Foerster R, Reissl S, Oberlaender M (2013) Cellular organization of cortical barrel columns is whisker-specific. Proceedings of the National Academy of Sciences 110:19113-19118.

Mihoko M, Eguchi S (2002) Robust blind source separation by beta divergence. Neural computation 14:1859-1886.

Miller JP, Jacobs GA, Theunissen FE (1991) Representation of sensory information in the cricket cercal sensory system. I. Response properties of the primary interneurons. J Neurophysiol 66:1680-1689.

Moreno-Bote R, Beck J, Kanitscheider I, Pitkow X, Latham P, Pouget A (2014) Information-limiting correlations. Nature neuroscience 17:1410-1417.

Mountcastle VB (1997) The columnar organization of the neocortex. Brain 120:701-722.

Mountcastle VB, Davies PW, Berman AL (1957) Response properties of neurons of cat's somatic sensory cortex to peripheral stimuli. Journal of neurophysiology.

Nádasdy Z, Hirase H, Czurkó A, Csicsvari J, Buzsáki G (1999) Replay and time compression of recurring spike sequences in the hippocampus. The Journal of neuroscience 19:9497-9507.

Nasser H, Marre O, Cessac B (2013) Spatio-temporal spike train analysis for large scale networks using the maximum entropy principle and Monte Carlo method. J Statist Mech 2013:P03006.

Obermayer K, Blasdel GG (1993) Geometry of orientation and ocular dominance columns in monkey striate cortex. The Journal of neuroscience 13:4114-4129.

Ohiorhenuan IE, Mechler F, Purpura KP, Schmid AM, Hu Q, Victor JD (2010) Sparse coding and high-order correlations in fine-scale cortical networks. Nature 466:617-621.

Oja E (1982) Simplified neuron model as a principal component analyzer. Journal of mathematical biology 15:267-273.

Okun M, Lampl I (2008) Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. Nature neuroscience 11:535-537.
Olmos A (2004) A biologically inspired algorithm for the recovery of shading and reflectance images. Perception 33:1463-1473.
Onken A, Grünewälder S, Munk MH, Obermayer K (2009) Analyzing short-term noise dependencies of spike-counts in macaque prefrontal cortex using copulas and the flashlight transformation.
Onken A, Karunasekara P, Kayser C, Panzeri S (2014) Understanding neural population coding: information theoretic insights from the auditory system. Advances in Neuroscience 2014.
Onken A, Liu JK, Karunasekara PPCR, Delis I, Gollisch T, Panzeri S (In preparation) Using Non-Negative Matrix Factorization for the Analysis of Large Scale Population Spike Trains.
Optican LM, Richmond BJ (1987) Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. III. Information theoretic analysis. J Neurophysiol 57:162-178.
Oram M, Wiener M, Lestienne R, Richmond B (1999) Stochastic nature of precisely timed spike patterns in visual system neuronal responses. Journal of neurophysiology 81:3021-3033.
Overduin SA, d'Avella A, Roh J, Carmena JM, Bizzi E (2015) Representation of Muscle Synergies in the Primate Brain. The Journal of neuroscience 35:12615-12624.
Pachitariu M, Lyamzin DR, Sahani M, Lesica NA (2015) State-dependent population coding in primary auditory cortex. The Journal of neuroscience 35:2058-2073.
Paninski L (2004) Maximum likelihood estimation of cascade point-process neural encoding models. Network 15:243-262.
Panzeri S, Diamond ME (2010) Information carried by population spike times in the whisker sensory cortex can be decoded without knowledge of stimulus time.
Panzeri S, Petroni F, Petersen RS, Diamond ME (2003) Decoding neuronal population activity in rat somatosensory cortex: role of columnar organization. Cerebral Cortex 13:45-52.
Panzeri S, Senatore R, Montemurro MA, Petersen RS (2007) Correcting for the sampling bias problem in spike train information measures. J Neurophysiol 98:1064-1072.
Panzeri S, Brunel N, Logothetis NK, Kayser C (2010a) Sensory neural codes using multiplexed temporal scales. Trends in neurosciences 33:111-120.
Panzeri S, Brunel N, Logothetis NK, Kayser C (2010b) Sensory neural codes using multiplexed temporal scales. Trends Neurosci 33:111-120.
Panzeri S, Ince RA, Diamond ME, Kayser C (2014) Reading spike timing without a clock: intrinsic decoding of spike trains. Philosophical Transactions of the Royal Society of London B: Biological Sciences 369:20120467.
Panzeri S, Macke JH, Gross J, Kayser C (2015) Neural population coding: combining insights from microscopic and mass signals. Trends in cognitive sciences 19:162-172.
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 29:769-777.
Panzeri S, Pola G, Petroni F, Young MP, Petersen RS (2002) A critical assessment of different measures of the information carried by correlated neuronal firing. Biosystems 67:177-185.
Perin R, Berger TK, Markram H (2011) A synaptic organizing principle for cortical neuronal groups. Proceedings of the National Academy of Sciences 108:5419-5424.
Petersen CC (2007) The functional organization of the barrel cortex. Neuron 56:339-355.
Petersen CC, Hahn TT, Mehta M, Grinvald A, Sakmann B (2003) Interaction of sensory responses with spontaneous depolarization in layer 2/3 barrel cortex. Proceedings of the National Academy of Sciences 100:13638-13643.
Petersen RS, Diamond ME (2000) Spatial–temporal distribution of whisker-evoked activity in rat somatosensory cortex and the coding of stimulus location. The Journal of neuroscience 20:6135-6143.

Petersen RS, Panzeri S, Diamond ME (2001) Population coding of stimulus location in rat somatosensory cortex. Neuron 32:503-514.
Petersen RS, Panzeri S, Diamond ME (2002) The role of individual spikes and spike patterns in population coding of stimulus location in rat somatosensory cortex. Biosystems 67:187-193.
Peyrache A, Benchenane K, Khamassi M, Wiener SI, Battaglia FP (2010) Principal component analysis of ensemble recordings reveals cell assemblies at high temporal resolution. Journal of computational neuroscience 29:309-325.
Pfau D, Pnevmatikakis EA, Paninski L (2013) Robust learning of low-dimensional dynamics from large neural ensembles. In: Advances in neural information processing systems, pp 2391-2399.
Picado-Muino D, Borgelt C, Berger D, Gerstein G, Grun S (2013) Finding neural assemblies with frequent item set mining. Front Neuroinform 7:9.
Pienkowski M, Harrison RV (2005) Tone responses in core versus belt auditory cortex in the developing chinchilla. Journal of Comparative Neurology 492:101-109.
Pillow JW, Shlens J, Paninski L, Sher A, Litke AM, Chichilnisky EJ, Simoncelli EP (2008) Spatio-temporal correlations and visual signalling in a complete neuronal population. Nature 454:995-999.
Pitkow X, Liu S, Angelaki DE, DeAngelis GC, Pouget A (2015) How can single sensory neurons predict behavior? Neuron 87:411-423.
Pnevmatikakis EA, Soudry D, Gao Y, Machado TA, Merel J, Pfau D, Reardon T, Mu Y, Lacefield C, Yang W (2016) Simultaneous denoising, deconvolution, and demixing of calcium imaging data. Neuron.
Pola G, Thiele A, Hoffmann KP, Panzeri S (2003) An exact method to quantify the information transmitted by different mechanisms of correlational coding. Network 14:35-60.
Ponce-Alvarez A, Nácher V, Luna R, Riehle A, Romo R (2012) Dynamics of cortical neuronal ensembles transit from decision making to storage for later report. The Journal of neuroscience 32:11956-11969.
Pouzat C, Mazor O, Laurent G (2002) Using noise signature to optimize spike-sorting and to assess neuronal classification quality. Journal of Neuroscience Methods 122:43-57.
Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1996) Numerical recipes in C: Citeseer.
Quian Quiroga R, Panzeri S (2009) Extracting information from neuronal populations: information theory and decoding approaches. Nat Rev Neurosci 10:173-185.
Rabinovich M, Huerta R, Laurent G (2008) Transient dynamics for neural processing. Science 321:48-50.
Radons G, Becker J, Dülfer B, Krüger J (1994) Analysis, classification, and coding of multielectrode spike trains with hidden Markov models. Biological cybernetics 71:359-373.
Rao RP, Sejnowski TJ (2003) Self–organizing neural systems based on predictive learning. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 361:1149-1175.
Riehle A, Grammont F, Diesmann M, Grun S (2000) Dynamical changes and temporal precision of synchronized spiking activity in monkey motor cortex during movement preparation. J Physiol Paris 94:569-582.
Rigotti M, Barak O, Warden MR, Wang X-J, Daw ND, Miller EK, Fusi S (2013) The importance of mixed selectivity in complex cognitive tasks. Nature 497:585-590.
Roudi Y, Nirenberg S, Latham PE (2009) Pairwise maximum entropy models for studying large biological systems: when they can work and when they can't. PLoS Comput Biol 5:e1000380.
Roweis S (1998) EM algorithms for PCA and SPCA. Advances in neural information processing systems:626-632.
Roweis S, Ghahramani Z (1999) A unifying review of linear Gaussian models. Neural computation 11:305-345.
Rucci M, Victor JD (2015) The unsteady eye: an information-processing stage, not a bug. Trends in neurosciences 38:195-206.

Rullen RV, Thorpe SJ (2001) Rate coding versus temporal order coding: what the retinal ganglion cells tell the visual cortex. Neural computation 13:1255-1283.
Rutkowski RG, Miasnikov AA, Weinberger NM (2003) Characterisation of multiple physiological fields within the anatomical core of rat auditory cortex. Hearing research 181:116-130.
Saal HP, Vijayakumar S, Johansson RS (2009) Information about complex fingertip parameters in individual human tactile afferent neurons. The Journal of neuroscience 29:8022-8031.
Sadacca BF, Mukherjee N, Vladusich T, Li JX, Katz DB, Miller P (2016) The Behavioral Relevance of Cortical Neural Ensemble Responses Emerges Suddenly. The Journal of neuroscience 36:655-669.
Sadtler PT, Quick KM, Golub MD, Chase SM, Ryu SI, Tyler-Kabara EC, Byron MY, Batista AP (2014) Neural constraints on learning. Nature 512:423-426.
Sakata S, Harris KD (2009) Laminar structure of spontaneous and sensory-evoked population activity in auditory cortex. Neuron 64:404-418.
Sally SL, Kelly JB (1988) Organization of auditory cortex in the albino rat: sound frequency. Journal of neurophysiology 59:1627-1638.
Santhanam G, Byron MY, Gilja V, Ryu SI, Afshar A, Sahani M, Shenoy KV (2009) Factor-analysis methods for higher-performance neural prostheses. Journal of neurophysiology 102:1315-1330.
Savin C, Joshi P, Triesch J (2010) Independent component analysis in spiking neurons. PLoS Comput Biol 6:e1000757.
Schneidman E, Berry MJ, 2nd, Segev R, Bialek W (2006) Weak pairwise correlations imply strongly correlated network states in a neural population. Nature 440:1007-1012.
Schreiner C, Mendelson J, Raggio M, Brosch M, Krueger K (1997) Temporal processing in cat primary auditory cortex. Acta Oto-Laryngologica 117:54-60.
Schreiner CE, Froemke RC, Atencio CA (2011) Spectral processing in auditory cortex: Springer.
Scott BH, Malone BJ, Semple MN (2011) Transformation of temporal processing across auditory cortex of awake macaques. Journal of neurophysiology 105:712-730.
Scott J, Pillow JW (2012) Fully Bayesian inference for neural models with negative-binomial spiking. In: Advances in neural information processing systems, pp 1898-1906.
Seidemann E, Meilijson I, Abeles M, Bergman H, Vaadia E (1996) Simultaneously recorded single units in the frontal cortex go through sequences of discrete and stable states in monkeys performing a delayed localization task. Journal of Neuroscience 16:752-768.
Shadlen MN, Newsome WT (1998) The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. J Neurosci 18:3870-3896.
Shamir M, Sompolinsky H (2006) Implications of neuronal diversity on population coding. Neural computation 18:1951-1986.
Shlens J, Field GD, Gauthier JL, Grivich MI, Petrusca D, Sher A, Litke AM, Chichilnisky EJ (2006) The structure of multi-neuron firing patterns in primate retina. J Neurosci 26:8254-8266.
Shusterman R, Smear MC, Koulakov AA, Rinberg D (2011) Precise olfactory responses tile the sniff cycle. Nature neuroscience 14:1039-1044.
Smaragdis P, Brown JC (2003) Non-negative matrix factorization for polyphonic music transcription. In: Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on., pp 177-180: IEEE.
Smith A, Brown EN (2003) Estimating a state-space model from point process observations. Neural computation 15:965-991.
Sompolinsky H, Yoon H, Kang K, Shamir M (2001) Population coding in neuronal systems with correlated noise. Physical Review E 64:051904.
Song S, Sjöström PJ, Reigl M, Nelson S, Chklovskii DB (2005) Highly nonrandom features of synaptic connectivity in local cortical circuits. PLoS Biol 3:e68.

Sra S, Dhillon IS (2006) Nonnegative matrix approximation: Algorithms and applications: Computer Science Department, University of Texas at Austin.
Steinbach M, Karypis G, Kumar V (2000) A comparison of document clustering techniques. In: KDD workshop on text mining, pp 525-526: Boston.
Steinmetz PN, Roy A, Fitzgerald P, Hsiao S, Johnson K, Niebur E (2000) Attention modulates synchronized neuronal firing in primate somatosensory cortex. Nature 404:187-190.
Steriade M, Timofeev I, Grenier F (2001) Natural waking and sleep states: a view from inside neocortical neurons. Journal of neurophysiology 85:1969-1985.
Steriade M, Contreras D, Dossi RC, Nunez A (1993) The slow (< 1 Hz) oscillation in reticular thalamic and thalamocortical neurons: scenario of sleep rhythm generation in interacting thalamic and neocortical networks. The Journal of neuroscience 13:3284-3299.
Stevenson IH, Kording KP (2011) How advances in neural recording affect data analysis. Nature neuroscience 14:139-142.
Stiebler I, Neulist R, Fichtel I, Ehret G (1997) The auditory cortex of the house mouse: left-right differences, tonotopic organization and quantitative analysis of frequency representation. Journal of Comparative Physiology A 181:559-571.
Stopfer M, Jayaraman V, Laurent G (2003) Intensity versus Identity Coding in an Olfactory System. Neuron 39:991-1004.
Theunissen F, Roddey JC, Stufflebeam S, Clague H, Miller J (1996) Information theoretic analysis of dynamical encoding by four identified primary sensory interneurons in the cricket cercal system. Journal of neurophysiology 75:1345-1364.
Tipping ME, Bishop CM (1999) Probabilistic principal component analysis. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 61:611-622.
Tkacik G, Marre O, Amodei D, Schneidman E, Bialek W, Berry MJ, 2nd (2014) Searching for collective behavior in a large network of sensory neurons. PLoS Comput Biol 10:e1003408.
Tolhurst DJ, Movshon JA, Dean AF (1983) The statistical reliability of signals in single neurons in cat and monkey visual cortex. Vision research 23:775-785.
Torre E, Picado-Muino D, Denker M, Borgelt C, Grun S (2013) Statistical evaluation of synchronous spike patterns extracted by frequent item set mining. Front Comput Neurosci 7:132.
Truccolo W, Hochberg LR, Donoghue JP (2010) Collective dynamics in human and monkey sensorimotor cortex: predicting single neuron spikes. Nat Neurosci 13:105-111.
Truccolo W, Eden UT, Fellows MR, Donoghue JP, Brown EN (2005) A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. J Neurophysiol 93:1074-1089.
Trussell LO, Fischbach GD (1989) Glutamate receptor desensitization and its role in synaptic transmission. Neuron 3:209-218.
Tuckwell HC (1988) Introduction to theoretical neurobiology: Cambridge University Press.
Uchida N, Poo C, Haddad R (2014) Coding and transformations in the olfactory system. Annual review of neuroscience 37:363-385.
Valtschanoff JG, Weinberg RJ, Kharazia VN, Schmidt HH, Nakane M, Rustioni A (1993) Neurons in rat cerebral cortex that synthesize nitric oxide: NADPH diaphorase histochemistry, NOS immunocytochemistry, and colocalization with GABA. Neuroscience letters 157:157-161.
van Steveninck RRdR, Lewen GD, Strong SP, Koberle R, Bialek W (1997) Reproducibility and variability in neural spike trains. Science 275:1805-1808.
van Vreeswijk C (2010) Stochastic models of spike trains. In: Analysis of Parallel Spike Trains, pp 3-20: Springer.
van Vreeswijk C, Sompolinsky H (1996) Chaos in neuronal networks with balanced excitatory and inhibitory activity. Science 274:1724-1726.

VanRullen R, Thorpe SJ (2002) Surfing a spike wave down the ventral stream. Vision research 42:2593-2615.
VanRullen R, Guyonneau R, Thorpe SJ (2005) Spike times make sense. Trends in neurosciences 28:1-4.
Vinje WE, Gallant JL (2000) Sparse coding and decorrelation in primary visual cortex during natural vision. Science 287:1273-1276.
Wang H, Nie F, Huang H, Makedon F (2011) Fast nonnegative matrix tri-factorization for large-scale data co-clustering. In: IJCAI Proceedings-International Joint Conference on Artificial Intelligence, p 1553.
Wehr M, Zador AM (2003) Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. Nature 426:442-446.
Wei J, Bai W, Liu T, Tian X (2015) Functional connectivity changes during a working memory task in rat via NMF analysis. Frontiers in behavioral neuroscience 9.
Wild S, Curry J, Dougherty A (2004) Improving non-negative matrix factorizations through structured initialization. Pattern Recognition 37:2217-2232.
Wilke SD, Eurich CW (2002) Representational accuracy of stochastic neural populations. Neural computation 14:155-189.
Woolsey TA, Van der Loos H (1970) The structural organization of layer IV in the somatosensory region (SI) of mouse cerebral cortex: the description of a cortical field composed of discrete cytoarchitectonic units. Brain research 17:205-242.
Wu W, Liu S (2015) Neural decoding in motor cortex using state space models with hidden states. In: Advanced State Space Methods for Neural and Clinical Data (Chen Z, ed): Cambridge University Press.
Wu W, Kulkarni JE, Hatsopoulos NG, Paninski L (2009) Neural decoding of hand motion using a linear state-space model with hidden states. IEEE transactions on neural systems and rehabilitation engineering: a publication of the IEEE Engineering in Medicine and Biology Society 17:370.
Wu W, Gao Y, Bienenstock E, Donoghue JP, Black MJ (2006) Bayesian population decoding of motor cortical activity using a Kalman filter. Neural computation 18:80-118.
Wu W, Black MJ, Mumford D, Gao Y, Bienenstock E, Donoghue JP (2004) Modeling and decoding motor cortical activity using a switching Kalman filter. Biomedical Engineering, IEEE Transactions on 51:933-942.
Xu-Friedman MA, Regehr WG (2004) Structural contributions to short-term synaptic plasticity. Physiological Reviews 84:69-85.
Xu W, Liu X, Gong Y (2003) Document clustering based on non-negative matrix factorization. In: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval, pp 267-273: ACM.
Xue Y, Tong CS, Chen Y, Chen W-S (2008) Clustering-based initialization for non-negative matrix factorization. Applied Mathematics and Computation 205:525-536.
Yoshimura Y, Callaway EM (2005) Fine-scale specificity of cortical networks depends on inhibitory cell type and connectivity. Nature neuroscience 8:1552-1559.
Yoshimura Y, Dantzker JL, Callaway EM (2005) Excitatory cortical neurons form fine-scale functional networks. Nature 433:868-873.
Yu BM, Cunningham JP, Santhanam G, Ryu SI, Shenoy KV, Sahani M (2009) Gaussian-Process Factor Analysis for Low-Dimensional Single-Trial Analysis of Neural Population Activity. Journal of neurophysiology 102:614-635.
Yu Y, Crumiller M, Knight B, Kaplan E (2010) Estimating the amount of information carried by a neuronal population. Front Comput Neurosci 4:10.
Zohary E, Shadlen MN, Newsome WT (1994) Correlated neuronal discharge rate and its implications for psychophysical performance.

Zuo Y, Safaai H, Notaro G, Mazzoni A, Panzeri S, Diamond ME (2015) Complementary contributions of spike timing and spike rate to perceptual decisions in rat S1 and S2 cortex. Current Biology 25:357-363.

Zylberberg J, Cafaro J, Turner MH, Shea-Brown E, Rieke F (2016) Direction-selective circuits shape noise to ensure a precise population code. Neuron 89:369-383.