Characterization of information and causality measures for the study of neuronal data

Daniel Chicharro Raventós

Tesi Doctoral UPF / 2011

Supervised by

Dr. Ralph Gregor Andrzejak

Department of Information and Communication Technologies

Barcelona, February 2011

Dipòsit Legal:
ISBN:

Acknowledgments

I am indebted to a lot of people for their help during the PhD. I would thus like to thank

- Ralph G. Andrzejak, for supervising my thesis and everything else.
- The five members of the committee, for accepting to evaluate this thesis.
- Gustavo Deco, for giving me the opportunity to work in his group.
- Thomas Kreuz, for giving me the opportunity to work with him and to stay in San Diego for a while. And for a nice trip the Mount Hood.
- Anders Ledberg, for his comments and discussions, helping me substantially with my work.
- Larissa Albantakis, for her comments and help reading my work.
- Andres Buehlmann, the necessary condition for all my computer calculations.
- Mario Pannunzi for this stuff about matlab.
- Rita Almeida and Ernest Montbrió, for help when teaching Statistics.
- Alex Roxin for all the smart discussions in the journal clubs.
- My officemates Carolina, Larissa, Elena, and Joana, for bearing my sometimes unhinged humor.
- Larissa, Joana, Marina, Mario, Andrea, Andres, Étienne, Johan, Laura, Pedro, Tim, Yota, Gustavo, Ernest, Sara and Albeno for good times at lunch time, and thanks specially to those of you who had the patience to repeatedly invite me to coffee despite my negatives.

- The people of psychology, to Núria Sebastián for a great Christmas dinner.
- Daniela for popping in the office talking so much and being so energetic
- Gisela, Nano and Francesca, for trusting me to be a good subject in their experiments.
- Lydia, for all her help.
- Rupan, for being a tender parody of all the worse qualities of a human being.
- Pau, Taken, Croce, Kiko, Xevo, for the old good times.
- Pau, Jan, Jordi, because sometimes is time for fun.
- Josep, Bernat, for a great trip, and because humor is the most ephemeral and supreme of arts.
- All the taxpayers that make my grants possible.
- The 'Comissionat per a Universitats i Recerca del Departament d'Innovació, Universitats i Empresa de la Generalitat de Catalunya i el Fons Social Europeu', that supported me with an FI grant for the thesis.
- All the people that should be here and are not.
- My parents, for all my live.
- Ariadna Soy, because you are the one.

Abstract

We study two methods of data analysis which are common tools for the analysis of neuronal data. In particular, we examine how causal interactions between brain regions can be investigated using time series reflecting the neural activity in these regions. Furthermore, we analyze a method used to study the neural code that evaluates the discrimination of the responses of single neurons elicited by different stimuli. This discrimination analysis is based on the quantification of the similarity of the spike trains with time scale parametric spike train distances. In each case we describe the methods used for the analysis of the neuronal data and we characterize their specificity using simulated or exemplary experimental data. Taking into account our results, we comment the previous studies in which the methods have been applied. In particular, we focus on the interpretation of the statistical measures in terms of underlying neuronal causal connectivity and properties of the neural code, respectively.

Resum

Estudiem dos mètodes d'anàlisi de dades que són eines habituals per a l'anàlisi de dades neuronals. Concretament, examinem la manera en què les interaccions causals entre regions del cervell poden ser investigades a partir de sèries temporals que reflecteixen l'activitat neuronal d'aquestes regions. A més a més, analitzem un mètode emprat per estudiar el codi neuronal que avalua la discriminació de les respostes de neurones individuals provocades per diferents estímuls. Aquesta anàlisi de la discriminació es basa en la quantificació de la similitud de les seqüències de potencials d'acció amb distàncies amb un paràmetre d'escala temporal. Tenint en compte els nostres resultats, comentem els estudis previs en els quals aquests mètodes han estat aplicats. Concretament, ens centrem en la interpretació de les mesures estadístiques en termes de connectivitat causal neuronal subjacent i propietats del codi neuronal, respectivament.

Contents

Index of figures x

Index of tables xi

1 Introduction 1

2 Studying causality in the brain with time series analysis 3

2.1 Criteria and assumptions for assessing causality 3

2.1.1 Introduction 3

2.1.2 Delay coordinates and the mapping of nearest neighbors 8

2.1.3 Comparison with Granger causality 14

2.2 Sensitivity and specificity of the measures implementing the MNN criterion 18

2.2.1 Measures implementing the MNN criterion . . . 18

2.2.2 Simulated dynamics 21

2.2.3 The influence of the coupling strength and the noise levels 22

2.2.4 The influence of the embedding parameters . . . 28

2.3 Inferring and quantifying causality in the brain 31

3 Studying the neural code with spike train distances 39
3.1 Stimuli discrimination 39
3.1.1 Introduction 39
3.1.2 Discrimination analysis 43
3.1.3 Spike train distances 48
3.2 Discrimination analysis with Poisson spike trains 53
3.2.1 Spike train distances for time-independent Poisson spike trains 53
3.2.2 Information and discriminative precision for time-dependent Poisson spike trains 56
3.3 Information and discriminative precision for experimental responses to transient constant stimuli 68
3.3.1 Dependence on the measure and the classifier . . 70
3.3.2 Temporal accumulation of information 71
3.3.3 Distribution and redundancy of the information along the spike trains 75
3.4 Information and discriminative precision for experimental responses to time-dependent stimuli 82
3.4.1 Dependence of the information and the discriminative precision on the firing rate 84
3.4.2 The time scale characteristic of reliable patterns . 87
3.5 The interpretation of discrimination analysis . . 93
3.5.1 The dependence on the classifier and the measure 94
3.5.2 The dependence on the length of the spike trains for different codes 95
3.5.3 The dependence on the number of trials 96
3.5.4 The application of discrimination analysis to experimental data: transient constant vs. time-dependent stimuli 96
3.5.5 Spike train distances to study population coding . 106
3.6 Conclusions 107

4 Conclusions 109

A The relation between the dimension of an attractor and the number of neighbors 113

B Comparison of the discriminative precision and the spike timing precision 115

C Electrophysiology of the songbird's recordings 117

List of Figures

2.1 Dependence of the measures on the coupling strength for coupled Lorenz dynamics. 23
2.2 Dependence of the difference of the measures for opposite direction on the coupling strength for coupled Lorenz dynamics superimposed with noise 24
2.3 The influence of the embedding parameters 30

3.1 Measures of spike train dissimilarity in dependence on the rates of two time-independent Poisson spike trains. . . . 54
3.2 Dependence of the mutual information on the measure and the classifier. 57
3.3 Maximal mutual information and optimal time scale in dependence on the measure and the classifier. 58
3.4 Dependence of the maximal mutual information and the optimal time scale on the latency difference and the length of the after-transient interval, for spike trains simulating the response to transient constant stimuli. 62
3.5 Dependence of the maximal mutual information and the optimal time scale on the length of the after-transient interval and on the classifier for phasic rate responses. . . . 66
3.6 Dependence of the maximal mutual information and the optimal time scale on the number of trials. 67
3.7 Time-dependent average rates for each stimulus of the four exemplary cells responding to transient constant stimuli. 69

3.8 Dependence of the mutual information on the measure and the classifier for exemplary cells. 72
3.9 Maximal mutual information and optimal time scale in dependence on the measure and the classifier for the exemplary cells. 73
3.10 Temporal accumulation of information. Dependence of the maximal mutual information and the optimal time scale on the length of the recordings used in the discrimination analysis. 74
3.11 Temporal distribution of information for cell g_2. 77
3.12 Temporal distribution of information for cell v_1. 78
3.13 Redundancy of information along the recordings for cell g_2. 80
3.14 Redundancy of information along the recordings for cell v_1. 81
3.15 Responses of two exemplary auditory cells with low rates elicited by conspecific songs. 85
3.16 Responses of two exemplary auditory cells with high rates elicited by conspecific songs. 86
3.17 Mutual information and optimal time scale in dependence on the average number of spikes per song. 88
3.18 The calculation of the characteristic time scale of the reliable patterns. 90
3.19 Correlation of the average mutual information, the optimal time scale of discrimination, and the average number of spikes per song with the characteristic time scale of the reliable patterns. 92

List of Tables

2.1 Expected dependence of the MNN measures on the distances for a coupling from X to Y 21

3.1 Limits of the spike train dissimilarity measures 49

Chapter 1

Introduction

In this thesis we study two different methods of data analysis that are commonly applied to analyze neural activity. First, we consider the investigation of causal interactions between brain regions using time series analysis. Second, we address how the neural code is studied by analyzing the similarity of the responses of single neurons elicited by different stimuli.

Characterizing causal interactions is a general issue relevant in many fields apart from neuroscience. Only quite recently the causal connectivity between brain regions has started to be examined (see Pereda et al., 2005; Bressler and Seth, 2010, and references therein), so that many of the methods used have been taken from other fields (see Granger, 1980; Pearl, 2009). We focus on one approach which was developed in the field of nonlinear time series analysis (Schiff et al., 1996; Arnhold et al., 1999). In this thesis we characterize the specificity of the measures proposed in this approach that have been commonly used to quantify the causal interactions (see Pereda et al., 2005, for a review), and a new measure we introduced (Chicharro and Andrzejak, 2009) which outperforms the specificity of the previous measures. We also compare this approach to Granger causality (Granger, 1963, 1969, 1980), another approach often applied to study neural signals. Furthermore, we discuss the applicability of these methods to neural data and the problems that generally prevent from interpreting the results in terms of underlying neuronal causal inter-

actions. How the measures should be interpreted is important because the measures of causality are often appreciated because they allow to examine the effective connectivity (Friston, 1994), in contrast to other statistical measures which reflect only the correlations in the data.

In contrast to the generality of causal analysis, the analysis of the similarity of the responses of single neurons elicited by different stimuli was specifically introduced to study the neural code in sensory systems (Victor and Purpura, 1996). In particular, it was designed to evaluate the relevance of spike timing in the encoding of the stimuli, in contrast to the simplest rate encoding which was frequently considered. The method consists in a discrimination analysis that compares the capability of a neuron to discriminate between different stimuli based on the similarity between the spike trains. The discrimination performance is evaluated in dependence on the spike timing sensitivity, which is varied using time scale parametric spike train distances to quantify the similarity between the spike trains. As a result, the discrimination analysis provides an estimate of the maximal discrimination performance, of the spike timing precision necessary for this discrimination performance, and of the improvement in discrimination when considering the spike timing with respect to a rate code. We here examine the degree to which the quantities obtained from the discrimination analysis are informative about the neural code. In particular, it is important to consider if the discriminative precision can be related to some characteristic time scale of the code, and what the actual estimated discrimination performance and the improvement when considering spike timing tell us about what and how is being encoded.

Given the clear distinction of our work into two different parts, we will introduce each study separately. In Chapter 2 we present our work related to causal analysis. In Chapter 3, we present the work on the study of the neural code using spike trains similarity. Finally, in Chapter 4, we discuss the elements in common of the two studies and we point to present and future work to continue this research.

Chapter 2

Studying causality in the brain with time series analysis

2.1 Criteria and assumptions for assessing causality

2.1.1 Introduction

Causal interactions have been studied in many different fields in the last century, and different frameworks have been proposed (see Pearl, 2009; Greenland and Brumback, 2002; Pearl, 2010, for an overview and comparison). For example, the framework of counterfactual analysis has been more studied in the statistician community (Rubin, 1974), while Structural Equation Modeling (Wright, 1921) has been commonly applied in econometrics (Haavelmo, 1943) and social sciences (Duncan, 1975). Structural Equation Modeling has also been applied in neuroscience, in particular in neuroimaging (McIntosh and Gonzalez-Lima, 1991, 1994). A similar approach designed specifically to take into account the particularities of neuroimaging data is the framework of Dynamic Causal Modeling (Friston et al., 2003; Daunizeau et al., 2009). A common characteristic of these approaches is that a model is proposed *a priori* relating the variables which causal connections are studied. This model reflects anatomically motivated assumptions about the connectivity. Furthermore, depending on the complexity of the model, the variables included can directly correspond to the recorded signals (for example, the blood oxygen

level-dependent (BOLD) signal), or explicitly account, for example, for hemodynamic responses introducing hidden variables in the model. In any case, these methods are hypothesis-driven in the sense that models are proposed and compared after fitting the parameters with the data.

In contrast to these hypothesis-driven approaches, other methods to study causal interactions are data-driven in the sense that conclusions about the existence and strength of the causal interactions are only based on the analysis of experimental data. In general it is not possible to reliably infer the existence of a causal interaction from observed data without any intervention on the systems (Pearl, 2009). However, under some conditions, causality can be inferred. For example, for the simple case of two systems X and Y from which two time series are recorded, consider that y_{i+1} corresponds to the future value of Y, and $y^{1:i} = \{y_i, y_{i-1}, ...y_1\}$, $x^{1:i} = \{x_i, x_{i-1}, ...x_1\}$ to the past of Y and X, respectively. If temporal precedence is imposed as a physical requirement for the causal interactions, causality from X to Y can be assessed examining the conditional independence of y_{i+1} on $x^{1:i}$, once conditioned first on $y^{1:i}$. The use of conditional independence to identify causal interactions is not restricted to time series and can be generally used to construct causal graphs where each node represents a variable (Pearl, 2009). For time series analysis, the criterion of conditional independence corresponds to the criterion formulated by Granger (1969, 1980), for the study of econometric data. In fact, the same formulation has been proposed repeatedly in different fields, comprising ethology (Marko, 1973), information theory (Massey, 1990), and nonlinear time series analysis (Schreiber, 2000).

The criterion of conditional independence provides a general test for causality because it involves the comparison of probability distributions. However, this generality makes the criterion difficult to test in practice, and further assumptions are usually used to test for causality between time series. For linear Gaussian stationary stochastic processes, Granger causality *in mean* (Granger, 1963) is the most widely applied method. Instead of examining the conditional independence of the probability distribution, it restricts itself to considering the conditional independence of the mean, which can be quantified in terms of the improvement in the

linear predictability of Y considering the past of X (Wiener, 1956; Lütkepohl, 2006). This allows testing for causality by applying linear regression or fitting autoregressive models (Geweke, 1982). This linear Gaussian Granger causality has been applied also in neuroscience to study the causal connectivity. A network of brain regions is defined using, for example, local field potentials (e. g. Brovelli et al., 2004) or BOLD signals (e. g. Roebroeck et al., 2005) to identify each node. However, the linearity and Gaussianity assumptions are often not adequate for neuronal electrophysiological signals. Several extensions have been proposed to test for causality on nonlinear data (e. g. Hiemstra and Jones, 1994; Ancona et al., 2004). In particular, the transfer entropy (Schreiber, 2000), which is the Kullback-Leibler distance (Cover and Thomas, 2006) quantifying directly the conditional independence criterion from the probability distribution (Barnett et al., 2009; Amblard and Michel, 2010), is increasingly applied to characterize causal interactions (e. g. Vicente et al., 2010; Besserve et al., 2010). These extensions form part of the variety of measures that have been proposed in the recent years for a nonlinear analysis of neurophysiological signals (see Pereda et al., 2005, for a review).

In this thesis we focus on an alternative criterion to study causality, which also belongs to the group of data-driven approaches. We will refer to it as the criterion of the mapping of nearest neighbors (MNN criterion). Measures based on this criterion have been developed in the field of dynamical systems (Ott, 2002) and nonlinear time series analysis (Kantz and Schreiber, 2003). Therefore, they are specially appropriate to study directional couplings between nonlinear deterministic self-sustained dynamical systems, in contrast to Granger causality, which is designed for stochastic dynamics.

Since the work of Pecora and Carroll (1990), much attention has been paid to interactions between nonlinear dynamical systems leading to synchronization (see Pikovsky et al., 2001; Boccaletti et al., 2002, Chapter 14 in Kantz and Schreiber, 2003, and Chapter 10 in Ott, 2002). Measures to detect different types of synchronization, comprising generalized synchronization (Rulkov et al., 1995), or phase synchronization (Rosenblum et al., 1996) were proposed, and have been applied to study interdepen-

dencies between electroencephalographic (EEG) and magnetoencephalographic (MEG) signals (e. g. Stam and van Dijk, 2002; Tass et al., 2003). However, these measures are not sensitive to the direction of the coupling. Alternatively, the direction of the coupling can be studied quantifying how nearest neighbors in the delay coordinates space (Takens, 1981) of the driven system are mapped to the delay coordinates space of the driving system, and oppositely. Several measures have been proposed to quantify the mapping of the neighbors in the delay coordinates space (e. g. Schiff et al., 1996; Arnhold et al., 1999; Quian Quiroga et al., 2000). These measures have been applied, for example, to characterize the changes in the interdependence of electroencephalographic (EEG) signals recorded from patients suffering from epilepsy (Arnhold et al., 1999) depending on the epileptic state. The use of a method specific for nonlinear deterministic dynamics is in this case motivated by the traces of nonlinear determinism characteristic of the epileptic activity (Andrzejak et al., 2001), associated with high levels of synchronization during the epileptic seizures (see Stam, 2005, for a review).

Independently of which criterion is used to test for the existence of causal interactions, a measure needs to be derived from this criterion which is sensitive and specific enough to account for the causal interactions. Furthermore, the estimation of the measure has to be characterized to identify possible biases arising from the finite sampling of the data in experimental studies. For example, for information theory measures, like the transfer entropy, there is a rich bibliography of different estimation methods (see Paninski, 2003b; Hlaváčkova-Schindler et al., 2007, for a review), which includes also estimators based on nearest neighbors statistics in the delay coordinates space (Kraskov et al., 2004). For these type of measures there is a clear difference between the definition of the measure and its estimation, because the entropy (Shannon, 1948) or the Kullback-Leibler distance (Cover and Thomas, 2006) are well-founded quantities with a clear interpretation in terms of the amount of information of the stochastic variables. By contrast, such a clear difference is blurred for other measures introduced *ad-hoc* to specifically implement one of the criteria of causality. This is the case of the measures derived

from the criterion of the mapping of nearest neighbors. The proliferation of different variants of the original MNN measures (Arnhold et al., 1999; Quian Quiroga et al., 2002; Pereda et al., 2001; Bhattacharya et al., 2003; Andrzejak et al., 2003; Kantz and Schreiber, 2003) has to be thus understood as the attempt to improve the estimation of directional couplings using the appropriate normalization of the measure. This normalization is necessary because it has been shown that the original measures (Arnhold et al., 1999), can be biased by asymmetries in the statistical properties of the time series, which are not necessarily directly related to the directional couplings (Quian Quiroga et al., 2000; Schmitz, 2000). These asymmetries can be due to differences in the dynamics, but also to different levels of measurement noise. The principal result of this part of the thesis is to identify the various sources of bias and show that most of them can be avoided by using an appropriate normalization. We furthermore diminish the remaining bias by introducing a measure based on ranks of distances, instead of distances.

This part is organized as follows. After this introduction (Section 2.1.1), in Section 2.1.2 the reconstruction of the state-space of the dynamics by delay coordinates is justified and used to motivate the criterion of the mapping of nearest neighbor. Furthermore, this criterion is compared to the Granger causality criterion based on conditional independence (Section 2.1.3). In Section 2.2 we present a characterization of the measures implementing the MNN criterion (Chicharro and Andrzejak, 2009). In particular, in Section 2.2.1 we describe the principal MNN measures proposed to study directional couplings, and a new measure we introduced (Chicharro and Andrzejak, 2009) that substitutes distances statistics by rank statistics. In Section 2.2.2 we describe the simulated models we use to characterize the measures. The results are presented in Sections 2.2.3 and 2.2.4. In Section 2.3 we discuss how these results contribute to the assessment of causal interactions, the caveats of the criterion of the mapping of nearest neighbors and more generally the problems and challenges to study causal effects in the brain.

2.1.2 Delay coordinates and the mapping of nearest neighbors

Delay coordinates reconstruction of the state-space

The scenario contemplated by the MNN criterion consists in the case where two time series $\{x_i\}$, $\{y_i\}$, $i = 1, 2, ..., N$ are recorded from two systems X and Y, respectively. These systems are assumed to be separate deterministic stationary dynamics which both exhibit an independent self-sustained motion. It is further assumed that if there is a coupling it is unidirectional and too weak to induce a synchronized motion. Under these assumptions, directional couplings can be detected by quantifying the probability with which close states of the driven dynamics are mapped to close states of the driving dynamics, and viceversa. Assuming that X is the driving system and Y the driven system, the dynamics can be generally represented as:

$$\dot{\mathbf{x}}(t) = \mathbf{F}(\mathbf{x}(t)) \tag{2.1a}$$

$$\dot{\mathbf{y}}(t) = \mathbf{G}(\mathbf{y}(t), \mathbf{x}(t))., \tag{2.1b}$$

where $\mathbf{x}(t)$, $\mathbf{y}(t)$ are multivariate variables representing all the degrees of freedom of the systems. However, these variables are not observable, and the dynamics have to studied only from the measured time series. Information from the underlying dynamics can be obtained from a measured signal using a state-space reconstruction of the dynamics with delay coordinates. The possibility of the reconstruction, under some assumptions, is assured by Takens's theorem (Takens, 1981; Sauer et al., 1991).

Consider a realization $\mathbf{x}(t)$ of the deterministic dynamical system X evolving in an attractor of dimension D_X. From this dynamics we measure a signal

$$x_i = g(\mathbf{x}(t_i)) \tag{2.2}$$

using the measurement function $g(\cdot)$. These measurements result in the time series $\{x_i\}$. Delay vectors are built, using an embedding dimension m and a time delay τ, by constructing:

$$\mathbf{x}_i = (x_i, ..., x_{i-(m-1)\tau}). \tag{2.3}$$

Hence, a temporal sequence of delay vectors $\{\mathbf{x}_i\}$ is obtained for $i = \eta, ..., N$, where η is the embedding window $\eta = (m-1)\tau$. Taken's theorem states that a bijective function between the original dynamics X and its reconstruction $\{\mathbf{x}_i\}$ exists, provided that the embedding dimension m is higher than $2\mathrm{D}_X$. The assumptions required by Taken's theorem are: First, that the underlying dynamics is stationary. Second, that both the dynamics and the measurement function are generic, in the sense that the measured signal reflects all the degrees of freedom of X. Finally, that the measurement function is invertible. Furthermore, the proof of existence of the bijective function assumes infinitely long noise-free time series, in which case the reconstruction does not crucially depend on the specific value chosen for the time delay τ. For the reconstruction from experimental data, when only a finite noisy sampling is available, this parameter has to be adjusted to obtain a good reconstruction. The selection of m and τ is a matter of study by itself, and different methods have been proposed to find optimal values (e. g. Fraser and Swinney, 1986; Pecora et al., 2007). However, in practice, the reconstruction is often done for a range of m and τ, and one has also to consider the variability of the results depending on the parameters.

We here give an heuristic argument for the existence of a function between the original dynamics in X and the delay vectors reconstructed space (Ott, 2002; Chicharro, 2007), but we do not prove the necessity of $m > 2\mathrm{D}_X$ for this function to be bijective. This argument is enough to later motivate the criterion of the mapping of nearest neighbors. Considering that the dynamics are deterministic, and further assuming they are reversible, considering Equation 2.1a, each past state of the autonomous driving system X can uniquely be determined by $\mathbf{x}(t_i)$, so that

$$\mathbf{x}(t_{i-k\tau}) = \mathbf{L}_k(\mathbf{x}(t_i)) \tag{2.4}$$

for $k = 0, ..., (m-1)$, where $\mathbf{L}_k$ represents the backward application k iterative times of the discrete map related to function $\mathbf{F}$ in Equation 2.1a, when using a time step τ. Furthermore, given the measurement function in Equation 2.2, the delay vector of Equation 2.3 can be determined as $\mathbf{x}_i = (g(\mathbf{L}_0(\mathbf{x}(t_i))), ..., g(\mathbf{L}_{(m-1)}(\mathbf{x}(t_i)))$, which can be expressed in a

compressed way as:

$$\mathbf{x}_i = \mathbf{H}(\mathbf{x}(t_i)), \tag{2.5}$$

where $\mathbf{H}$ is bijective if the required conditions are fulfilled.

The criterion of the mapping of nearest neighbors

The argumentation above leading to Equation 2.5 starts from Equation 2.1a, which reflects the autonomy of the driving system. Oppositely, as noticed in Schiff et al. (1996), for the driven system, an analogous argumentation using Equation 2.1b shows that the delay vector $\mathbf{y}_i$ does not reconstruct only the dynamics Y, but contains also information about the X dynamics. In particular, the delay vector can be expressed as:

$$\mathbf{y}_i = (g(\tilde{\mathbf{L}}_0(\mathbf{y}(t_i), \mathbf{x}(t_i))), ..., g(\tilde{\mathbf{L}}_{m-1}(\mathbf{y}(t_i), \mathbf{x}(t_i)))), \tag{2.6}$$

or in a compressed form

$$\mathbf{y}_i = \tilde{\mathbf{H}}(\mathbf{y}(t_i), \mathbf{x}(t_i)). \tag{2.7}$$

In this case, the degree to which $\mathbf{y}_i$ provides a good reconstruction of the joint dynamics $X \times Y$ depends on the strength of the coupling (Stark et al., 1997). For the case of a weak coupling the condition of genericity is not fulfilled and the degrees of freedom of X are hardly reflected in a short time series $\{y_i\}$. However, a weak coupling is already enough to break the bijective projection between the delay vectors and the underlying dynamics of Y.

The asymmetry between Equations 2.5 and 2.7 is the fundament of the MNN criterion, together with the asymmetry between Equations 2.1a and 2.1b. Below we provide arguments for the validity of this criterion. They do not constitute a rigorous proof, but they are an extended and more detailed explanation of the justification provided by Schiff et al. (1996). Consider the delay vector $\mathbf{y}_i$ and its neighbor $\mathbf{y}_j$, such that $||\mathbf{y}_i - \mathbf{y}_j|| < \varepsilon_y$, where we take $\varepsilon_y \to 0$ and $||\cdot||$ represents a distance in the space of delay vectors. According to Equation 2.6 in order to get $||\mathbf{y}_i - \mathbf{y}_j|| < \varepsilon_y$ it is necessary that:

$$||g(\tilde{\mathbf{L}}_k(\mathbf{y}(t_i), \mathbf{x}(t_i))) - g(\tilde{\mathbf{L}}_k(\mathbf{y}(t_j), \mathbf{x}(t_j)))|| < \varepsilon_{y,k} \tag{2.8}$$

holds for each component of the delay vectors, $k = 0, ..., m-1$, with $\varepsilon_{y,k} \to 0$. Equation 2.8 does not imply that $||\mathbf{y}(t_i) - \mathbf{y}(t_j)|| < \tilde{\varepsilon}_y$, and $||\mathbf{x}(t_i) - \mathbf{x}(t_j)|| < \tilde{\varepsilon}_x$, where $\tilde{\varepsilon}_{x,y} \to 0$. However, the fulfillment of these two conditions is sufficient for Equation 2.8 to hold. By contrast, other possible ways to fulfill Equation 2.8 depend on the concrete form of $\tilde{\mathbf{L}}_k$ and the concrete combination of the terms in it involving $\mathbf{y}(t_i)$, $\mathbf{x}(t_i)$, and $\mathbf{y}(t_j)$, $\mathbf{x}(t_j)$, respectively, so that they may fulfill it for some particular t_i, but not in all the attractor of the dynamics. Given that, it is expected that the fulfillment of Equation 2.8 increases the probability that $||\mathbf{x}(t_i) - \mathbf{x}(t_j)|| < \tilde{\varepsilon}_x$ holds, with respect to alternatively choosing a random t_j. Assuming that $||\mathbf{x}(t_i) - \mathbf{x}(t_j)|| < \tilde{\varepsilon}_x$ holds, given Equation 2.5, for a bijective function $\mathbf{H}$,

$$||\mathbf{x}(t_i) - \mathbf{x}(t_j)|| < \tilde{\varepsilon}_x \Leftrightarrow ||\mathbf{x}_i - \mathbf{x}_j|| < \varepsilon_x, \tag{2.9}$$

where $\varepsilon_x \to 0$. Equation 2.9 relies on the geometrical structure of the attractor of the dynamical systems, which ensures that for either $\tilde{\varepsilon}_x \to 0$ or $\varepsilon_x \to 0$ it is possible to find neighbors close enough to be projected to the corresponding neighbors by either $\mathbf{H}$ or its inverse. In Appendix A we show how the number of close neighbors is related to the dimension of the system's attractor. Accordingly, we obtain that

$$||\mathbf{y}_i - \mathbf{y}_j|| < \varepsilon_y \Rightarrow ||\mathbf{x}_i - \mathbf{x}_j|| < \varepsilon_x. \tag{2.10}$$

Oppositely, this line of arguments cannot be applied in the opposite direction due to the asymmetry between Equations 2.5 and 2.7. In particular $||\mathbf{y}(t_i) - \mathbf{y}(t_j)|| < \tilde{\varepsilon}_y \not\Leftrightarrow ||\mathbf{y}_i - \mathbf{y}_j|| < \varepsilon_y$. Furthermore, $\mathbf{y}(t_i)$ does not appear in Equation 2.1a, and the argument involving the equation analogous to Equation 2.8 in the other direction does not hold neither. In consequence,

$$||\mathbf{x}_i - \mathbf{x}_j|| < \varepsilon_x \not\Rightarrow ||\mathbf{y}_i - \mathbf{y}_j|| < \varepsilon_y. \tag{2.11}$$

Although Equations 2.10 and 2.11 are formulated as implications, in practice it is better to think about the mapping of the nearest neighbors in terms of probabilities. Theoretically, this is because Equation 2.8 does

not imply $||\mathbf{x}(t_i) - \mathbf{x}(t_j)|| < \tilde{\varepsilon}_x$, but only increases the probability that this holds. Furthermore, for experimental finite noisy data, the measurement function may be not invertible, the bijectivity between the underlying dynamics space and the delay vectors space is not assured, and nearest neighbors can be not sufficiently close so that Equation 2.10 holds for all the points. Furthermore, the use of two different indexes for ε_x, ε_y indicates that, for dynamics with different properties, the probabilities of mapping the nearest neighbors should not be compared directly against each other for the two directions, but have to be compared first to a reference specific to each dynamic. Therefore, we can formulate the criterion of the mapping of nearest neighbors as follows: *for a coupling from X to Y, the increase in the probability of the nearest neighbors in the reconstructed space of Y to be mapped to nearest neighbors in the reconstructed space of X is higher than the increase in the probability in the opposite direction,*

$$\Delta P(||\mathbf{x}_i - \mathbf{x}_j|| < \varepsilon_x | \; ||\mathbf{y}_i - \mathbf{y}_j|| < \varepsilon_y) > \Delta P(||\mathbf{y}_i - \mathbf{y}_j|| < \varepsilon_y | \; ||\mathbf{x}_i - \mathbf{x}_j|| < \varepsilon_x). \tag{2.12}$$

The degree to which the increases in the probabilities are different depends on the properties of the dynamics (e. g. their dimensions, their Lyapunov spectrum), and on the strength of the coupling, the impact of which depends itself on the properties of the dynamics. Furthermore, the MNN criterion is expected to hold only for a limited range of coupling strengths. For uncoupled systems, no increase in the mapping probability should be observed, and can only result from a bias caused by the different properties of the dynamics. For a strong coupling, the systems can achieve generalized synchronization (Rulkov et al., 1995), involving the existence of a function

$$\mathbf{y}(t) = \phi(\mathbf{x}(t)) \tag{2.13}$$

between the dynamics of the driving system X and the driven system Y. This functional relationship, together with Equations 2.8 and 2.9, leads to

$$||\mathbf{x}_i - \mathbf{x}_j|| < \varepsilon_x \Rightarrow ||\mathbf{y}_i - \mathbf{y}_j|| < \varepsilon_y. \tag{2.14}$$

Hence, considering Equations 2.10 and 2.14, for a coupling strong enough to produce generalized synchronization, the directionality of the coupling cannot be assessed using the MNN criterion. In fact, a smoother relationship

$$\mathbf{y}_i = \tilde{\phi}(\mathbf{x}_i) \tag{2.15}$$

in the delay vectors space can be attained even when ϕ is not smooth (Rulkov and Afraimovich, 2003; He et al., 2003). For finite experimental data the closeness of the nearest neighbors is limited and therefore the smoothness of the functionality becomes more relevant (So et al., 2002; Barreto et al., 2003). Considering this, the effective impact of generalized synchronization on the nearest neighbors statistics for a particular coupling is specific of each pair of dynamics.

The argumentation above is based on the deterministic nature of the dynamics, which allows the reconstruction of the underlying dynamics using the delay coordinates. Nonetheless, the state-space reconstruction by delay coordinates has been extended to noisy dynamics (Casdagli et al., 1991). In this case the relation between the space of X and the delay vector space does not correspond to a bijective function as in Equation 2.5, but can be represented by a conditional probability distribution. An effect in the probability of mapping the nearest neighbors is also expected for stochastic dynamics. Consider a bivariate autoregressive process of order p:

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} = \sum_{j=1}^{p} \mathbf{A}_j \begin{pmatrix} x_{i-j} \\ y_{i-j} \end{pmatrix} + \begin{pmatrix} u_i^{(x)} \\ u_i^{(y)} \end{pmatrix}, \tag{2.16}$$

where the matrices $\mathbf{A}_j$ are lower triangular matrices such that only Y depends on X, and the innovations $u_i^{(x)}$ and $u_i^{(y)}$ are independent Gaussian white noise terms with zero mean. Each component of the delay vector of X depends on its own past and on the X innovations:

$$x_{i+j\tau} = L_{-j}(x_{(i-1)}, .., x_{i-p}, u_i^{(x)}, ..., u_{i+j\tau}^{(x)}), \tag{2.17}$$

while each component of Y depends on the past of the two processes and

the innovations of both processes:

$$y_{i+j\tau} = \tilde{L}_{-j}(x_{(i-1)}, .., x_{i-p}, y_{(i-1)}, .., y_{i-p}, u_i^{(x)}, \\ ..., u_{i+(j-1)\tau}^{(x)}, u_i^{(y)}, ..., u_{i+j\tau}^{(y)}), \tag{2.18}$$

Since the innovations are independent white noise terms, for a sufficiently high embedding window $(m-1)\tau$, the neighbors in the delay vectors space mainly result from common past trajectories. Therefore, in probabilistic terms, the same arguments used for the deterministic dynamics are valid for stochastic dynamics to justify the MNN criterion.

Overall, the principle of the nearest neighbors statistics can be used to assess the direction of the coupling for a limited range of coupling strength. Furthermore, given the arguments presented above, it is necessary to assume that the driving system is autonomous, that is, that in Equation 2.1a only the X dynamics appear, so that Equation 2.9 holds. For the bivariate case this means that the coupling is unidirectional. A bidirectional coupling is expected to lead to an increase in the probability of the mapping of the nearest neighbors in both directions and for which directions the increase is higher depends on the dynamics and the nature and strength of the couplings. For multivariate systems, another system Z can drive Y together with X, but the effect on the nearest neighbors statistics will then depend on the properties of the three dynamics.

2.1.3 Comparison with Granger causality

Given that the Granger causality criterion of conditional independence is an alternative general test for causality between time series, it is worth comparing with it the criterion of the mapping of nearest neighbors. We start by briefly reviewing Granger causality. Consider a multivariate stationary stochastic process $\mathbf{W}$ which can be divided into the processes X, Y, and Z. Causal interactions are studied between X and Y, and Z represents the rest of the processes. The condition for the existence of causality from X to Y is:

$$P(y_{i+1}|y^{1:i}, z^{1:i}, x^{1:i}) \neq P(y_{i+1}|y^{1:i}, z^{1:i}), \tag{2.19}$$

that is, that y_{i+1} is not independent of the past of X when previously conditioned on the past of all the other processes. The conditional distributions have to be examined for all the concrete values of the conditioning variables. This condition is general under two assumptions. First, that causal interactions follow the arrow of time; second, that $\mathbf{W}$ comprises any process interdependent with X and Y, to assure, for example, that the effect of an unobserved common driver is not mistaken as a direct causal dependence from X to Y. The first assumption is, *per se*, not very restrictive, and is motivated by the study of causality from time series. The second assumptions is stronger for experimental data, because in general it is not possible to record all the influencing processes. Given the necessity of these assumptions, Granger causality corresponds only to a specific subcase in the framework proposed by Pearl to study causality (Pearl, 2009). In Pearl's formulation it is not assumed that all the variables are observed, involving that causality cannot be assessed, in general, without intervention in the system.

We now compare the MNN criterion to Granger causality. Granger causality allows assessing causal interactions for any multivariate system as long as all the processes are observed, even when bidirectional couplings exist. Oppositely, the criterion of the mapping of nearest neighbors can only be applied when the couplings are unidirectional or, more generally, the driving system is autonomous. On the other hand, the MNN criterion does not require that all the degrees of freedom of the underlying dynamics are observed, only that they are reflected in the recorded signal. The fact that the whole past of the processes is used for partialization in the probability distributions of Equation 2.19 has been claimed to play a role equivalent to the delay coordinates reconstruction of the dynamics (Vicente et al., 2010). However, when X and Y are dynamical systems, y_{i+1} may not reflect the degrees of freedom of Y on which the coupling from X operates. If several steps in the future were considered to better reflect the influence of X on Y, also indirect causal influences through Z would be reflected (Lütkepohl, 1993).

Another apparent difference between the two approaches is that, while in Equation 2.12 the delay vectors are contemporaneous, in Equation

2.19, an asymmetry exists between the driving and driven processes, since only the future of the driven system is examined. In fact, the MNN criterion can be formulated also considering the future step. Given that the embedding obtained with the delay coordinates reconstruction is smooth, the counterpart of equation 2.1a in the space of the delay vectors can be obtained using the bijective function of Equation 2.5 and its inverse:

$$\mathbf{x}_{i+1} = \mathbf{H} \circ \mathbf{L}_{-1} \circ \mathbf{H}^{-1} \mathbf{x}_i. \tag{2.20}$$

This means that, considering only the first component of $\mathbf{x}_{i+1}$, Equation 2.10 can be extended to:

$$||\mathbf{y}_i - \mathbf{y}_j|| < \varepsilon_y \Rightarrow ||x_{i+1} - x_{j+1}|| < \varepsilon'_x. \tag{2.21}$$

Oppositely, a map analogous to the one of Equation 2.20 cannot be obtained to express $\mathbf{y}_{i+1}$ in terms only of $\mathbf{y}_i$, which further supports together with Equation 2.11 that:

$$||\mathbf{x}_i - \mathbf{x}_j|| < \varepsilon_x \nRightarrow ||y_{i+1} - y_{j+1}|| < \varepsilon'_y. \tag{2.22}$$

This predictive formulation of the MNN criterion was used in Schiff et al. (1996) and Le Van Quyen et al. (1999). While Schiff et al. (1996) took the inequality of Equation 2.12 in the same direction, Le Van Quyen et al. (1999) used the inverse inequality, following the argumentation valid for the regime of generalized synchronization where Equation 2.15 holds (see Arnhold et al., 1999, for a discussion on the correct direction of the inequality in Equation 2.12). To better compare the MNN criterion with Granger causality, we can assume that conditioning on the past of the processes or on a delay vector which accounts only for a limited past of the time series are equivalent. In practice, also in Equation 2.19 the past considered is limited due to finite sampling. Given that, the criterion of the mapping of nearest neighbors can be related to the following condition on the probability distributions:

$$D(P(x_{i+1}|y^{1:i}), P(x_{i+1})) > D(P(y_{i+1}|x^{1:i}), P(y_{i+1})), \tag{2.23}$$

where $D(\cdot, \cdot)$ is an appropriate measure of the difference of the probability distributions. While the difference between the probabilities in Equation 2.19 can be tested with any arbitrary statistics on the probabilities, here the direction of the inequality, inherited from Equation 2.12, is only valid if D is chosen to reflect the local dispersion of the neighbors. Therefore, considering the distinction made in Section 2.1.1 between criteria, measures, and estimation strategies, we see that the Granger causality criterion is more robust in the sense that it is clearly defined, independently of the statistics used to test it. In fact, the use of nearest neighbors statistics to characterize the probabilities is not exclusive of the MNN criterion and has been used in some extensions to nonlinear dynamics of the formulation of Granger causality for linear Gaussian processes in terms of predictability improvement (Chen et al., 2004; Feldmann and Bhattacharya, 2004). In these extensions the predictability is examined locally using local linear maps or a zero-order predictor.

Equation 2.23 is not a condition of conditional independence like Equation 2.19. Causality from X to Y cannot be assessed by examining the conditional independence

$$P(y_{i+1}|x^{1:i}) \neq P(y_{i+1}), \tag{2.24}$$

that is, without conditioning y_{i+1} on its own past. This is so because, for a causal influence in the opposite direction, $x^{1:i}$ is dependent on $y^{1:i}$, which makes y_{i+1} dependent on $x^{1:i}$ if the process has memory, leading already to the inequality of Equation 2.24. Accordingly, the condition of causality of Equation 2.23 has to compare the degree of conditional dependence in the two directions. The argumentation in Section 2.1.2 justifies this condition for dynamical systems reconstructed via delay coordinates for a given range of coupling strengths. In particular, the existence of generalized synchronization marks an upper bound for the coupling strengths for which the direction of the causal interaction can be assessed. For other type of processes, like the stochastic processes also discussed in Section 2.1.2, we do not have a clear way to know for which range of coupling strengths the criteria is expected to hold. Nonetheless, this limitation is not exclusive of the MNN criterion, and affects also the Granger causality

criterion. For example, the transfer entropy, which directly compares the two probability distributions of Equation 2.19, is nonmonotonic with the coupling strength (e. g Kaiser and Schreiber, 2002). This is due to the necessity to partialize in the own past in Equation 2.19. When X and Y get synchronized, the past of X does not contain further information than the past of Y.

2.2 Sensitivity and specificity of the measures implementing the MNN criterion

We here describe the measures that have been proposed to implement the criterion of the mapping of nearest neighbors and we introduce a new measure that substitutes distance statistics by rank statistics to attenuate the biases arising from differences in the statistical properties of the time series. We describe the simulated dynamics we use to characterize the measures and discuss the sources of bias. We then analyze the specificity of the measures in dependence on the coupling strength, the levels of noise and the embedding parameters used in the delay coordinates reconstruction.

2.2.1 Measures implementing the MNN criterion

We consider the case of a bivariate system such that the dynamics X and Y are assumed to be separate deterministic stationary dynamics which both exhibit an independent self-sustained motion. It is further assumed that if there is a coupling it is unidirectional and too weak to induce a synchronized motion. From the dynamics X and Y scalar time series x_i and y_i ($i = 1, \dots N$) are simultaneously measured. The dynamics are reconstructed using delay coordinates $\mathbf{x}_i = (x_i, \dots x_{i-(m-1)\tau})$, $\mathbf{y}_i = (y_i, \dots y_{i-(m-1)\tau})$ with embedding dimension m and delay τ ($i = 1, \dots N^* = N-(m-1)\tau$) (Kantz and Schreiber, 2003). When using nearest neighbors statistics to examine the direction of the coupling there are two different possibilities, the fixed mass (Arnhold et al., 1999) or the fixed distance approach (Cenys et al., 1991). We here follow the fixed mass approach,

which means that we select a fixed number of nearest neighbors k instead of all the neighbors closer than a fixed distance. By $v_{i,j}$ and $w_{i,j}$ ($j = 1, \dots k$) we denote the time indices of the k nearest neighbors of $\mathbf{x}_i$ and $\mathbf{y}_i$, respectively. These k neighbors are chosen excluding temporal neighbors within $|v_{i,j} - i| \leq W$ and $|w_{i,j} - i| \leq W$, where W is the Theiler window (Theiler, 1986). This window is necessary because, due to the autocorrelation in each system, the nearest neighbors on the same trajectory have a higher probability to be mapped to nearest neighbors in the other system even if they are uncoupled. For each $\mathbf{x}_i$, the k-mean squared Euclidean distance to its k nearest neighbors is $R_i^k(X) = \frac{1}{k} \sum_{j=1}^{k} |\mathbf{x}_i - \mathbf{x}_{v_{i,j}}|^2$, and the conditional k-mean distance is $R_i^k(X|Y) = \frac{1}{k} \sum_{j=1}^{k} |\mathbf{x}_i - \mathbf{x}_{w_{i,j}}|^2$. The mean distance to all other points is $R_i(X) = \frac{1}{N^*-1} \sum_{j=1, j\neq i}^{N^*} |\mathbf{x}_i - \mathbf{x}_j|^2$. [1] Based on these distances one can define:

$$S(X|Y) = \frac{1}{N^*} \sum_{i=1}^{N^*} \frac{R_i^k(X)}{R_i^k(X|Y)} \tag{2.25}$$

$$H(X|Y) = \frac{1}{N^*} \sum_{i=1}^{N^*} \log \frac{R_i(X)}{R_i^k(X|Y)} \tag{2.26}$$

$$N(X|Y) = \frac{1}{N^*} \sum_{i=1}^{N^*} \frac{R_i(X) - R_i^k(X|Y)}{R_i(X)} \tag{2.27}$$

$$M(X|Y) = \frac{1}{N^*} \sum_{i=1}^{N^*} \frac{R_i(X) - R_i^k(X|Y)}{R_i(X) - R_i^k(X)} \tag{2.28}$$

Equations 2.25, 2.26 were defined in Arnhold et al. (1999). Equation 2.27 was proposed as a normalized measure in Quian Quiroga et al. (2002), however it attains values of one only for synchronized periodic dynamics. Therefore, Equation 2.28 was derived from Equation 2.27 in Andrzejak et al. (2003). Moreover, Equation 2.28 was introduced independently

[1] Due to the exclusion of the vectors in the Theiler window W, N^* has to be adjusted, and cases $i < W$, $N^* - i < W$ need to be further distinguished. To simplify the notation we write all formulas for $W = 0$.

from Andrzejak et al. (2003) in Kantz and Schreiber (2003). Both approaches lead to very similar results, and we therefore do not consider Equation 2.27 but use Equation 2.28 instead. These measures have in common the quantification of conditional dispersion by $R_i^k(X|Y)$, related to the criterion of the mapping of the nearest neighbors. The other terms provide a reference specific to each system to see the impact of the conditioning. The measure S compares $R_i^k(X|Y)$ to the k-mean distance to the true nearest neighbors $R_i^k(X)$, which is the minimum value attainable, so that S values close to one are expected for a strong coupling. For uncoupled dynamics, $\mathsf{E}[R_i^k(X|Y)] = \mathsf{E}[R_i(X)]$, where $\mathsf{E}[\cdot]$ denotes the expected value across independent realizations of the dynamics. This determines the value of S, small but higher than zero for uncoupled dynamics. By contrast, the mean distance $R_i(X)$ is used in H as a reference, which is the expected value of the conditional k-mean distance for uncoupled dynamics. Accordingly, given the logarithm, H is supposed to be zero for uncoupled dynamics and increase to an upper bound determined by $R_i(X)/R_i^k(X)$. Both references are used in M, so that a value of zero is expected for uncoupled dynamics and a value of one for strong couplings. The degree to which these expected values hold will be examine below.

In Chicharro and Andrzejak (2009) we proposed the following rank-based statistics: For each $\mathbf{x}_i$, let $g_{i,j}$ denote the rank that the distance between $\mathbf{x}_i$ and $\mathbf{x}_j$ takes in a sorted ascending list of distances between $\mathbf{x}_i$ and all $\mathbf{x}_{j\neq i}$. The conditional k-mean rank is then $G_i^k(X|Y) = \frac{1}{k}\sum_{j=1}^{k} g_{i,w_{i,j}}$, and we define

$$L(X|Y) = \frac{1}{N^*}\sum_{i=1}^{N^*}\frac{G_i(X) - G_i^k(X|Y)}{G_i(X) - G_i^k(X)}. \tag{2.29}$$

where $G_i(X) = \frac{N^*}{2}$ and $G_i^k(X) = \frac{k+1}{2}$ denote the mean and minimal k-mean rank, respectively. This measure L has the same normalization as M but the distance-based statistics is substituted by rank-based statistics. For the opposite direction $S(Y|X),\ H(Y|X),\ M(Y|X)$, and $L(Y|X)$ are defined by exchanging the role of X and Y in the above definitions. Furthermore, we use the notation A to refer to the group of

Table 2.1: Expected dependence of the MNN measures on the distances for a coupling from X to Y

Measure	Uncoupled dynamics	Synchronized dynamics
$S(X\|Y)$	$R_i^k(X)/R_i(X)$	1
$H(X\|Y)$	0	$R_i(X)/R_i^k(X)$
$M(X\|Y)$	0	1
$L(X\|Y)$	0	1

all measures S, H, M, L [2]. Since Equation 2.12 involves the comparison of the mapping in both directions, $A(X|Y)$ by itself is not specific and $\Delta A = A(X|Y) - A(Y|X)$ is needed. The aim of the analysis is to examine to what extent $\Delta A > 0$ is a sensitive and specific condition for assessing that a unidirectional coupling from X to Y exists. Notice that the applicability of the criterion of the mapping of the nearest neighbors is a necessary but not a sufficient condition for $\Delta A > 0$ being reliable to assess causality.

2.2.2 Simulated dynamics

We restrict ourselves to the study of simulated systems, so that we can control the strength of the coupling and generate a sufficient number of realizations of the dynamics to obtain average results. In particular, we analyze uncoupled as well as unidirectionally coupled non-identical Lorenz dynamics (Lorenz, 1963) superimposed with different types of noise. The Lorenz dynamics correspond to the following differential equations for system X:

$$\begin{aligned} \dot{v}_1 &= 10(v_2 - v_1) \\ \dot{v}_2 &= 39v_1 - v_2 - v_1 v_3 \\ \dot{v}_3 &= v_1 v_2 - \tfrac{8}{3} v_3 \end{aligned} \tag{2.30}$$

[2] A Matlab code to calculate the measures implementing the MNN criterion is available at http://pre.aps.org/supplemental/PRE/v80/i2/e026217

and for system Y:

$$\begin{aligned} \dot{w}_1 &= 10(w_2 - w_1) + \varepsilon(v_1 - w_1) \\ \dot{w}_2 &= 35w_1 - w_2 - w_1 w_3 \\ \dot{w}_3 &= w_1 w_2 - \tfrac{8}{3} w_3. \end{aligned} \tag{2.31}$$

A unidirectional diffusive coupling from X to Y exists, with a strength controlled by ε. The dynamics were integrated using a 4-th order Runge-Kutta algorithm with fixed step size of 0.005 and a sampling interval of 0.03 time units. We used random initial conditions and applied 10^6 pre-iterations to diminish transients. As deterministic time series we use $\tilde{x}_i = v_1(t_i)$ and $\tilde{y}_i = w_1(t_i)$ and autoregressive processes as noise time series: $\xi_{i+1}^{X,Y} = a^{X,Y}\xi_i^{X,Y} + \zeta_{i+1}^{X,Y}$. Here $\zeta_i^{X,Y}$ denotes uncorrelated Gaussian noise with zero mean and unit variance. All examples studied here can then be written in the general form: $x_i = d^X\tilde{x}_i + n^X\xi_i^X, y_i = d^Y\tilde{y}_i + n^Y\xi_i^Y$ for $i = 1, \ldots N = 2048$. Throughout all simulations we use fixed values of $k = 5$, and $W = 50$ and set m and τ as specified below.

2.2.3 The influence of the coupling strength and the noise levels

As a first example we use Lorenz dynamics superimposed with Gaussian uncorrelated noise ($d^{X,Y} = 1$, $a^{X,Y} = 0$). Apart from uncoupled dynamics ($\varepsilon = 0$), we study coupled dynamics with ε separated by factors of 1.05 between 0.05 and 18. Within this set of coupling strengths, $\varepsilon_{\mathrm{GS}} = 9.28$ is the lowest value for which generalized synchronization is attained. This can be determined from the comparison of two replicas of the driven Y dynamics started at different initial conditions, but driven by the same realization of X (Kocarev and Parlitz, 1996). According to Equation 2.13, the two replicas are identically synchronized when they both depend only on X. Apart from noise-free dynamics we use the noise amplitudes $I^{X,Y} = [0.125 \times 1.5^n]\sigma^{X,Y}$ for $n = 0, \ldots 12$, where $\sigma^{X,Y}$ denote the standard deviation of $\tilde{x}_i$ and $\tilde{y}_i$, respectively. The noise is superimposed either only to the driver: $n^X \in I^X$, $n^Y = 0$; to the driver and to the response: $n^X \in I^X$, $n^Y \in I^Y$; or only to the response: $n^X = 0$, $n^Y \in I^Y$.

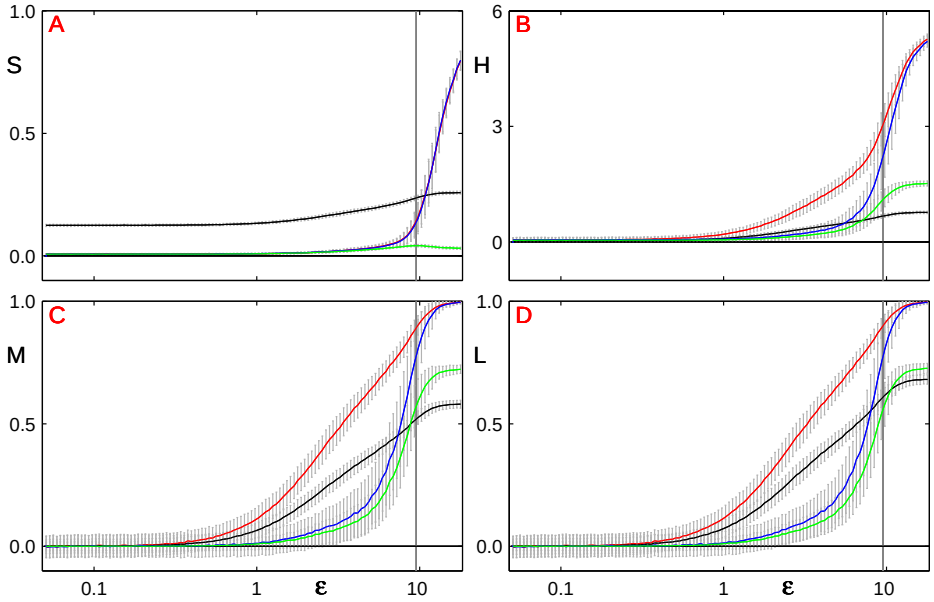

Figure 2.1: Dependence of A on ε for coupled Lorenz dynamics ($m = 8, \tau = 4$). For the noise-free case with $A(X|Y)$ in red and $A(Y|X)$ in blue. For $n^X = 0.95\sigma^X, n^Y = 0$ with $A(X|Y)$ in black and $A(Y|X)$ in green. In panel A the blue curve almost covers the red one. Error bars depict $\pm$ one standard deviation. Vertical lines mark ε_{GS}, the coupling strength for which Generalized synchronization is attained. Reproduced from Chicharro and Andrzejak (2009).

For the noise-free dynamics the coupling direction is correctly detected by $\Delta A > 0$ to a different degree for H, M, and L (Figure 2.1). For asymmetric noise levels some biases occur resulting in $\Delta A < 0$, and thus in the detection of the wrong coupling direction, for some range of ε. Comparing H, M, and L, the rank-based measure L is least affected by asymmetric noise. For the measure S, a more complicated picture is obtained which we will explain in terms of the various sources of biases in the following.

In Figure 2.1 we show the two directions of the measures so that it

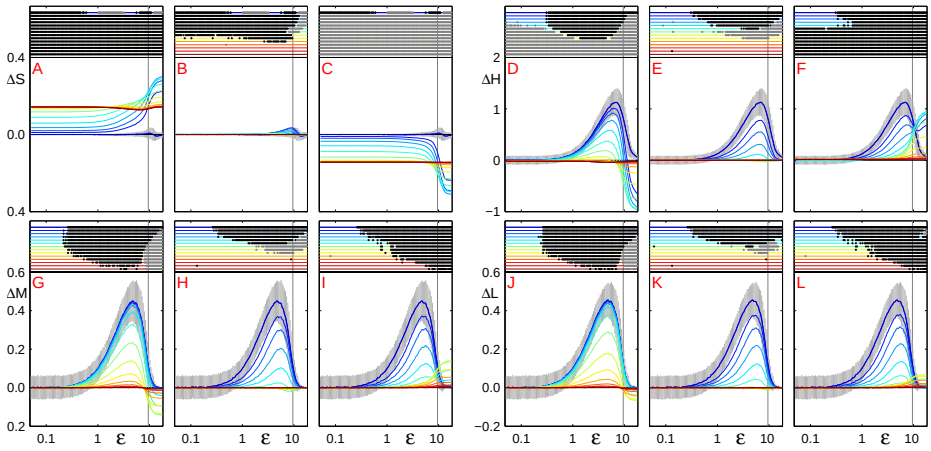

Figure 2.2: Dependence of ΔA on ε for $m = 8$, $\tau = 4$ for coupled Lorenz dynamics superimposed with noise. We used 1000 independent realizations of $\tilde{x}_i$ and $\tilde{y}_i$ for each ε and added independent realizations of ξ_i^X and ξ_i^Y for each noise amplitude specified in the text (A, D, G, J: noise on X; B, E, H, K: noise on X and Y; C, F, I, L: noise on Y). From blue to red colors indicate increasing noise levels. Error bars depict $\pm$ one standard deviation and are shown for $n^{X,Y} = 0$ only. Black and gray dots indicate significantly positive and negative ΔA values, respectively (Wilcoxon signed rank test at p=0.001.). Vertical lines mark $\varepsilon_{\mathrm{GS}}$. Reproduced from Chicharro and Andrzejak (2009).

is visible that both increase for an increasing coupling strength. In Figure 2.2 we more systematically study the dependence of $\mathsf{E}[\Delta A]$ on the coupling strength and the noise levels. We simplify the notation $\mathsf{E}[\Delta A]$ to ΔA and apply the Wilcoxon test to determine whether nonzero values of ΔA are significant (see black and gray dots in Figure 2.2). From a first inspection it can be seen that ΔS is the measure that is most biased. To understand some of the sources of bias we have to take into account that, as indicated by Arnhold et al. (1999), the effective dimension D_X is reflected in the proportion between the distances appearing in Table 2.1

as:

$$\mathsf{E}[R_i^k(X)/R_i(X)] \propto (k/N)^{2/D_X}, \tag{2.32}$$

and analogously for Y. The effective dimension is related to the degrees of freedom of the dynamics reflected in the reconstructed delay coordinates space but also to the level of measurement noise. For stochastic or very noisy dynamics, the effective dimension corresponds to the dimension m of the embedding, and the reconstructed space is filled. However, for chaotic dynamics the dimension can be fractal, because the trajectories of the dynamics do not fill all the space (see Chapter 3 Ott, 2002). Assuming ergodicity, and according to Table 2.1, we have that for zero or very small couplings, or for high levels of noise, $S(X|Y) \propto (k/N)^{2/D_X}$. Hence, $\mathsf{E}[\Delta S] > 0$ for $D_X > D_Y$. Therefore, generally ΔS is nonzero even for uncoupled noise-free deterministic dynamics with slightly different effective dimensions.

For the Lorenz dynamics used here we have $D_X > D_Y$ at $\varepsilon = 0$ and accordingly get $\Delta S > 0$ (Figures 2.1 A, 2.2 A-C). Upon increasing of ε, for $n^{X,Y} = 0$, D_Y at first increases due to the incorporation of the driver's degrees of freedom. However, when the coupling strength increases towards and beyond ε_{GS}, D_Y decreases due to the collapse of the joint dynamics to the synchronization manifold (Quian Quiroga et al., 2000). For the given dynamics this local maximum of D_Y is reflected in $\Delta S < 0$ found for an intermediate range of ε. For the noisy dynamics, increasing the noise level, the effective dimension converges to the embedding dimension m. Therefore, for asymmetric noise levels of X and Y the difference in D_X and D_Y and thereby ΔS is dominated by this asymmetry. Even for symmetric noise levels $n^X = n^Y$, measured relative to the respective standard deviation of X and Y, the impact of this noise can be asymmetric depending on the fine structure of the dynamics.

In contrast to S, the measure H seems better at detecting the existence of a directional coupling, at least for low levels of noise. For uncoupled noise-free Lorenz dynamics we get $\Delta H = 0$ (Figures 2.1 B, 2.2 D-F). With increasing coupling strength the stronger mapping of close states in Y to close states in X leads to $\Delta H > 0$. However, for asymmetric noise levels, at zero and low ε we see that $\Delta H > 0$ for $n^X > n^Y$, and $\Delta H < 0$

for $n^X < n^Y$. Here the bias results from the influence of the different effective dimensions on a bias inherent to the nonlinear dependence on the conditional distance $R_i^k(X|Y)$ (Andrzejak et al., 2006a). Since for uncoupled or weakly coupled dynamics $\mathsf{E}[R_i^k(X|Y)] = \mathsf{E}[R_i(X)]$, each term of the N^* summands in Equation 2.28 corresponds to $\gamma(\delta) = \log \frac{\mu}{\mu+\delta}$, where $\mu = R_i(X)$ and δ accounts for the deviation from the mean. Due to the concavity of the logarithm, $\gamma(\delta) + \gamma(-\delta) > 0$, and therefore a positive bias is expected for H. When the dynamics are comparable, like for the case of symmetric noise, this bias is comparable for $H(X|Y)$ and $H(Y|X)$, and it turns out not to be significant for ΔH. However, since the variance of $R_i^k(X|Y)$ around its mean $R_i(X)$ depends on the properties of the dynamics, the biases in opposite directions are not compensated for asymmetric levels of noise. In particular, for $n^X > 0$, $n^Y = 0$, given that the noise is uncorrelated, this means that the reconstructed space of X becomes more circularly symmetric, which results in a smaller variance of $R_i^k(X|Y)$. Therefore, the bias due to the nonlinearity is smaller for $H(X|Y)$ and $\Delta H < 0$ (Figure 2.2 D). For strong couplings and a moderate level of noise, there is another bias coherent to this one. Since $\mathsf{E}[R_i^k(X|Y)] = \mathsf{E}[R_i^k(X)]$ for ε close to ε_{GS}, a term proportional to $(k/N)^{D_X/2}$ contributes to $H(X|Y)$. In consequence, for $D_Y > D_X$ the value of ΔH increases, and viceversa.

Characterizing the measures S and H we already have discussed all the sources of bias that prevent from a correct assessment of the direction of the coupling. We briefly review them before examining M and L. First, the relation between the distances and the effective dimension shown in Equation 2.32 indicates that, according to Table 2.1, the references chosen in measures S and H are not adequate for weak and strong couplings respectively. The effective dimension does not correspond to the underlying dimension of the system, but to the dimension of the reconstructed state-space, which apart from the dimension of each of the systems, depends on the strength of the coupling, as well as on the noise level and color, and on the parameters m and τ used for the delay coordinates. Second, the nonlinear dependence on $R_i^k(X|Y)$ also introduces a bias. We commented this bias for H, because it is the predominant source of bias

for this measure. However, for S this bias is also present given the nonlinearity $\frac{1}{(\cdot)}$, which, given its convexity, has an influence opposite to the one in H. The impact of the bias is determined by the variance of $R_i^k(X|Y)$. This variance also depends in a non-trivial way on intrinsic properties of X and Y, the noise level and color, and the parameters m, τ, and k. Therefore, the accuracy of the measures depends on a lot of factors and although taking into account these sources of bias helps to qualitatively describe Figure 2.2, the actual noise levels or coupling strength for which we observe a change in the significance of ΔA result from the details of the dynamics and the particular combination of parameters chosen.

It is important to distinguish these biases, which arise from the estimation of the nearest neighbors statistics and the selection of a particular value of reference in each measure, from the fundamental limitation in the criterion caused by generalized synchronization (Section 2.1.2). As discussed in Section 2.1.3, this problem does not only affect these measures but is shared by any measure implementing the Granger causality criterion. In Figure 2.1 we see that for M, and L, $A(X|Y)$ and $A(Y|X)$ converge to 1, when there is no noise. For H, $A(X|Y)$ and $A(Y|X)$ converge to each other. For this simulated system, $A(X|Y)$ remains higher until they converge, but this is not guaranteed and depends on the smoothness of the function in Equation 2.15. The difference ΔA is nonmonotonic, starting to decrease when the coupling strength approaches ε_{GS} (Figures 2.2 D-L). In this region and for higher couplings a reliable assessment of the direction of the coupling is compromised. Therefore, even if $\Delta A > 0$, one should also keep track of the values in the two separate directions to know how reliable this nonzero ΔA value is, to infer a coupling from X to Y. However, for noisy dynamics, $A(X|Y)$ and $A(Y|X)$ decrease, hiding the unreliability of the assessment of the coupling direction.

We now continue the analysis of the bias for M and L. For independent dynamics the expected values of $M(X|Y)$, $M(Y|X)$ and thereby the one of ΔM are all zero (Figures 2.1 C, 2.2 G-I). The measure M depends linearly on $R_i^k(X|Y)$, and thus the bias caused by the nonlinearities described above for S and H does not affect M. The bias related to Equation 2.32 does not affect the mean of M. However, factorizing the

term $R_i(X)$ in $M(X|Y)$, we see that the slope of the linear dependence is determined by the denominator $(R_i(X)(a(k/N)^{2/D_X} - 1))$, where a is some constant. Therefore, for zero coupling the across realization variance of ΔM depends on D_X and D_Y, in contrast to its mean. For any noise level and any type of symmetry, ΔM is zero for small couplings. For weak coupling strengths, ΔM is positive for low noise levels and, if the noise is symmetric, becomes nonsignificant when the noise increases. However, for symmetric noise levels, even if the range of the parameters in which a wrong direction is indicated is smaller, for intermediate coupling strengths and intermediate levels of noise, $\Delta M < 0$ is still obtained, as it was for H in this regime. The fact that for asymmetric noise levels ΔM is still positive for higher levels of noise cannot be explained by the sources of bias here discussed, and result from the particularities of the dynamics.

The advantages gained by virtue of the appropriate normalization of M are inherited by L. The remaining dependence on the statistical properties of the time series for ε close to ε_{GS} is diminished since in contrast to distributions of distances, distributions of ranks of distances are always uniform. Notice that although this bias is still present, its absolute value is substantially lower for L than for M (Figure 2.1D). Also the region where $\Delta L < 0$ for symmetric noise is smaller than for M. Therefore, although the use of the appropriate normalization is the main advantage shared by L and M, the use of ranks helps to further attenuate the effect of the different statistical properties of the time series.

2.2.4 The influence of the embedding parameters

To further study the specificity of $\Delta A > 0$ for the assessment of couplings from X to Y, we examine the influence of the embedding parameters m and τ. We now focus on independent dynamics. At first we consider two examples of uncoupled Lorenz dynamics superimposed with asymmetric noise levels. While both examples share $\varepsilon = 0$, $d^{X,Y} = 1$, $a^Y = 0$, and $n^X = [0, 0.25, 0.5, 1, \ldots, 128]$, we have $a^X = 0$, $n^Y = 4$ in the first and $a^X = 0.97$, $n^Y = 0$ in the second example. Here $n^{X,Y}$ are given

as unitless amplitudes not related to the standard deviation of $\tilde{x}_i$ and $\tilde{y}_i$. Again nonzero values for ΔH and ΔS are obtained for these independent dynamics. Due to the dominance of the bias related to asymmetric effective dimensions the sign of ΔS is independent of the embedding parameters (Figure 2.3 A, E), although its magnitude increases with the embedding window, since $D_X = m$ for the noisy dynamics. Oppositely, for H the only bias for uncoupled dynamics is the nonlinear dependence on $R_i^k(X|Y)$. The variance of $R_i^k(X|Y)$ and thereby the sign of ΔH depends on the embedding parameters and the asymmetry of the noise amplitudes (Figures 2.3 B, F). In contrast to the false detection found using S or H for almost all the noise levels, the mean values of ΔM and ΔL are never significantly different from zero. Only their standard deviation are influenced by the relative noise levels and the embedding parameters, as discussed in Section 2.2.3. In consequence, not a single false positive is obtained by M and L (Figures 2.3 C, D, G, H).

As a last example we use purely stochastic time series. As discussed in Section 2.1.2, the criterion of the mapping of nearest neighbors, although thought for deterministic dynamics, is still applicable for stochastic dynamics when formulated in terms of probabilities. Here we examine the influence of the autocorrelation of each of the stochastic dynamics in dependence on the embedding parameters. In particular the stochastic dynamics are autoregressive processes with asymmetric degrees of autocorrelation $d^{X,Y} = 0$, $a^Y = 0.5$, $n^{X,Y} = 1$, and $a^X = [0.99,\ 0.98,\ 0.95,\ 0.9,\ 0.8,\ 0.65,\ 0.5,\ 0.35,\ 0.2,\ 0]$. If the embedding window $(m-1)\tau$ is not too big for a given autocorrelation decay, a stronger autocorrelation is reflected in a lower effective dimension. Also the variance of $R_i^k(X|Y)$ for the uncoupled systems depends on the autocorrelation, that determines the shape of the cloud of delay vectors in the reconstructed space. Therefore both the bias related to Equation 2.32 and to the nonlinearity in the measures are affected by the autocorrelation, so that simple linear properties of the dynamics are enough to distort the detection of the coupling. Given the exponential dependence of the autocorrelation on $a^{X,Y}$, these effects are mainly observed for high values of $a^{X,Y}$.

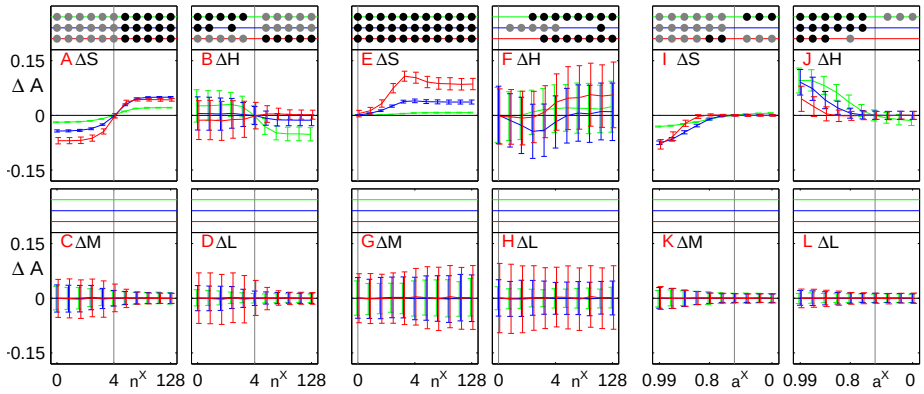

Figure 2.3: Analogous to Fig. 2.2 but for independent dynamics. Green: $m = 4, \tau = 1$; blue: $m = 6, \tau = 4$; red: $m = 8, \tau = 12$. (A-D) Uncoupled Lorenz dynamics superimposed with asymmetric levels of white noise in dependence on n^X. Here vertical lines mark n^Y. Slight offsets on the abscissa are used to distinguish different error bars. (E-H) Same as (A-D) but here for uncoupled Lorenz dynamics with asymmetric levels of Gaussian autocorrelated noise. (I-L) Same as (A-D) but for purely stochastic time series with asymmetric autocorrelation strengths in dependence on a^X. Here vertical lines mark a^Y. Reproduced from Chicharro and Andrzejak (2009).

For the measure S, the bias related to Equation 2.32 leads to $\Delta S < 0$ for $a^X > a^Y$ (Figure 2.3 I). Since $a^Y = 0.5$ we can consider that $D_Y \cong m$. Therefore, for a^X close to one, the bias is higher for higher values of m, because the autocorrelation in X reduces D_X relatively more with respect to $D_Y \cong m$. For lower values of a^X, if $(m-1)\tau$ is too big, in particular if τ is high, the components of the delay vector are less correlated, so that the reduction of the effective dimension is lower. Therefore, in this range the bias is higher for smaller $(m-1)\tau$. For the measure H, $\Delta H > 0$ for $a^X > a^Y$ (Figure 2.3 J). For H, the bias is caused by the nonlinearity, and thus by the variance of $R_i^k(X|Y)$ around $R_i(X)$. For high embedding windows and in particular for high τ the impact of the autocorrelation is reduced and the state space is filled more homoge-

neously. In consequence, we see that the bias decreases monotonically with $(m-1)\tau$, in contrast to the more complicated dependence observed for S. Finally, like for the two previous examples, ΔM and ΔL exhibit not a single false positive coupling detection (Figure 2.3 K,L).

2.3 Inferring and quantifying causality in the brain

In this part of the thesis we focused on one specific approach to assess causal interactions. In Section 2.1.2 we described the criterion of the mapping of nearest neighbors (MNN), and compared it to the criterion of Granger causality. We reexpressed the MNN criterion of Equation 2.12 in terms of conditional probability distributions (Equation 2.23) to be more comparable to the Granger causality criterion of conditional independence (Equation 2.19). We concluded that, given the assumptions and the argumentation justifying the MNN criterion, its applicability is more restrictive than the one of Granger causality. However, regarding the comparison of the criteria, it remains to be discussed to which degree the different assumptions are fulfilled in practice. In Section 2.2 we described different measures implementing the MNN criterion and analyzed how they are affected by different sources of bias related to the properties of the individual time series, including the properties of the underlying dynamics and of the measurement noise. We now summarize these results and turn back to the comparison with Granger causality, now addressing which are the problems to assess causal dependencies from neuronal data.

Regarding the measures implementing the criterion of the mapping of nearest neighbors (Equation 2.12), both ΔS and ΔH are of weak specificity. Moreover, ΔS is of poor sensitivity, while the sensitivity of ΔH is stronger but spoiled by its limited specificity. Oppositely, by virtue of their common appropriate normalization, the measures ΔM and ΔL are more sensitive and more specific for directional couplings. Furthermore, the use of rank statistics endows ΔL with a higher noise robustness as compared to ΔM. However, perhaps because S and H were the measures first introduced in Arnhold et al. (1999), they have been used almost uniquely when comparing the performance of these type of mea-

sures with other approaches using simulated data (e. g. Smirnov and Andrzejak, 2005; Ansari-Asl et al., 2006; Lungarella et al., 2007; Palus and Vejmelka, 2007; Kreuz et al., 2007a).

The influence of the individual properties of the dynamics of each system in the assessment of causal interactions and the effect of asymmetric levels of noise is not a problem specific of the MNN measures considered here. This problem has been pointed out for measures relying on the Granger causality criterion, like the linear Gaussian Granger causality (Nalatore et al., 2007) and a measure based on phase responses (Smirnov and Bezruchko, 2003). In Smirnov and Bezruchko (2003) a modification of a measure proposed by Rosenblum and Pikovsky (2001) for assessing causal relations between phases was introduced to correct a bias related to the difference in the characteristic frequencies of the dynamics. Nalatore et al. (2007) used a Kalman filter to explicitly model the asymmetric measurement noise that bias also Granger causality. Apart from the modification of the measures, also the use of surrogate time series (Schreiber and Schmitz, 2000) can help to avoid false detections. Ideally, the surrogate data should maintain the properties of the individual dynamics, but destroy any interdependence produced by the causal interactions. This corresponds to testing the null hypothesis that the dynamics are uncoupled. Although different types of surrogates have been proposed (Quian Quiroga et al., 2002; Andrzejak et al., 2003; Thiel et al., 2006), for nonlinear dynamics it is difficult to generate independent surrogates while maintaining the same nonlinear structure of the individual dynamics. When causality is studied in an event-related framework, it is easier to construct surrogates by shuffling randomly the index of the trial for one of the dynamics (Andrzejak et al., 2006a).

The characterization of the sensitivity and specificity of the measures implementing a given criterion is necessary to assess causality from a system X to a system Y. However, it seems a more fundamental issue to ask wether the assumptions needed to apply the criterion are fulfilled for a set of experimental data. In neuronal applications, the data generally consist in multivariate time series reflecting the activity in different areas of the brain, and multiple areas are expected to interact with each other.

Therefore, Granger causality seems more useful since it does not assume the existence of an autonomous driving system. However, the assumption of a well defined set of processes, and that this set comprises all the processes influencing X and Y, is in fact also a strong assumption. For EEG recordings, a time series is recorded at each electrode implanted in the brain, while for BOLD signals each voxel (volumetric pixel) corresponds to a time series. Therefore, the first problem for the application of any of the criteria is the identification of the systems between which the causal interactions are studied. The brain is an extended system, and the electrodes reflect the activity of extended regions. The signals recorded at close electrodes are correlated, as well as they are close voxels in neuroimaging recordings. This problem is not specific for the study of causal interactions, but it is common to any approach studying functional or effective connectivity in brain networks (see Bullmore and Sporns, 2009; Lehnertz et al., 2009; Rubinov and Sporns, 2010, for a review). This is a problem not only for the identification of the nodes in the network and the interpretation of the dependencies found between them, but also for the quantification of these dependencies, since the superposition of different sources affects the assessment of causality (e. g. Schmitz, 2000; Nolte et al., 2006). The identification of the systems is further hindered by other factors associated with the nature and setup of the recordings. For EEG data, the use of a common reference for the voltage at each electrode, can induce spurious dependencies (Guevara et al., 2005; Schiff, 2005), reflecting the activity in the electrode of reference. Alternatively, a bipolar montage can be used analyzing the time series resulting from the difference in voltage between adjacent electrodes. However, it is then more difficult to interpret the causal interactions between the time series. For the BOLD signals, the activity of the neurons is only indirectly reflected and depends on the hemodynamic response (Deshpande et al., 2010), affecting also the reliable assessment and the interpretation of causal dependencies.

The identification problem is by itself enough to indicate that, although being a more general criterion, Granger causality cannot be applied to neuronal data in a straightforward manner. However, if one obviates the problems in the interpretation of causal dependence arising from

the definition of the nodes, Granger causality seems better than the MNN criterion. The Granger causality criterion (Equation 2.19) is formulated in a multivariate framework, while the MNN criterion (Equation 2.12) is restricted to the bivariate case, or to specific cases of a multivariate framework in which X is autonomous. Furthermore, the fact that Granger causality can also be formulated in the spectral domain (Geweke, 1982; Ding et al., 2006) makes it more appealing to study causal interactions in the brain, since specific brain rhythms are related to brain functions (e. g. Brovelli et al., 2004). In consequence, Granger causality has gained more popularity in the recent years, specially for the analysis of LFP's and neuroimaging (see Gourevitch et al., 2006; Bressler and Seth, 2010, for a review). This led to the proposal of measures of Granger causality accounting to some extent for the effect of exogenous variables (Guo et al., 2008), or to consider the brain as an extended system (Barrett et al., 2010).

However, apart from the identification problem, there is a further practical limitation which prevents from testing reliably the condition in Equation 2.19. In this equation the Granger causality criterion is general for multivariate systems because the partialization includes Z, which accounts for the rest of the set apart from X and Y. This means that it is necessary to sample a high dimensional space that comprises the past of all the recorded signals. In practice a good sampling of the probability distributions cannot be achieved. Therefore, in the majority of applications of Granger causality to neuronal data, the partialization on Z is removed from Equation 2.19 (e. g. Brovelli et al., 2004; Roebroeck et al., 2005; Bressler et al., 2007, 2008; Zhou et al., 2009; Kayser and Logothetis, 2009; Besserve et al., 2010; Vicente et al., 2010). The remaining condition is still different from the MNN criterion because it considers the conditioning on the past of the own process, but strictly can only be applied to a bivariate system. This approach is often referred to as pairwise Granger causality (PGC) (Geweke, 1982), in contrast to the more general conditional Granger causality (CGC) (Geweke, 1984). When applied to multivariate systems PGC can lead to the false positive detection of causal interactions due to the unconsidered dependencies (Cadotte et al., 2008).

Even for the pairwise Granger causality, accounting for the whole past of the systems is too difficult in general. When nonlinear measures of Granger causality are used, the past is reduced to one or a few samples at particular time lags (e. g. Besserve et al., 2010; Vicente et al., 2010). In general, it is more usual to apply linear Gaussian Granger causality, and model the data with bivariate autoregressive models (e. g. Brovelli et al., 2004; Roebroeck et al., 2005; Kayser and Logothetis, 2009).

There is also another aspect of Granger causality which deserves some discussion. In Section 2.1.3 we concluded that the Granger causality criterion should be preferred because the condition it involves is better defined and more general, since the driving system is not required to be autonomous, which means that bidirectional causality is testable. For brain networks, considering the existence of structural feedback connections between different regions, bidirectional causal interactions are expected. The applicability in presence of bidirectional couplings is possible because, when testing for causality from X to Y, equation 2.19 compares two probabilities of y_{i+1}, and when testing causality in the opposite directions, two probabilities of x_{i+1} are compared. Oppositely, independently of which direction of causality is tested, the MNN criterion compares probabilities of X and Y (Equation 2.23), and the assumption that only in one direction the causal interaction is nonzero is implicit. This is why in Sections 2.2.3, 2.2.4 we used ΔA and not the measures $A(X|Y)$ and $A(Y|X)$ separately to assess the existence of a causal dependence. But for the Granger causality criterion the analogous measures $A_{X\to Y}$ and $A_{Y\to X}$ are associated separately to Equation 2.19 and its analogous for the opposite direction, testing for causality from X to Y and from Y to X, respectively. This has induced a frequent mistake that stems already from the terminology used by Granger, who associated this measures to the *strength* of the causal effect (Granger, 1963, 1969). However, as mentioned in Section 2.1.3, the partialization on the own past of the system implies that each measure $A_{X\to Y}$ and $A_{Y\to X}$ separately is expected to behave nonmonotonically for strong causal interactions, as it is ΔA in Figure 2.1. The Granger causality criterion does not consider which is the change in the probability of y_{i+1} caused by the past of X, but which is the

change that cannot already be explained by the own past of Y. Accordingly, it is designed to test for the existence of a nonzero causal interaction, but not to quantify the strength of the causal effect. While the MNN measure have been used to quantify the degree of interdependence using the sum $A(X|Y) + A(Y|X)$ (Quian Quiroga et al., 2002; Kreuz et al., 2007b; Andrzejak et al., 2011), an analogous measure $A_{X\to Y} + A_{Y\to X}$ would not be adequate for this purpose.

This difference between assessing the existence of a causal interaction and quantifying its strength is frequently disregarded. The first problem is that, when estimating any measure implementing the Granger causality criterion based on a finite data sampling, a nonzero value will be obtained which significance has to be evaluated. For the linear Gaussian Granger causality a significance level can be assigned analytically (Geweke, 1982; Lütkepohl, 2006). However, in general, some bootstrapping is necessary to obtain some confidence intervals. Alternatively, for brain networks a common practice is to select a threshold and to consider only interactions above this value (Bullmore and Sporns, 2009). Furthermore, the quantification of the causal effect is implicit in the comparison of causal dependencies across frequencies (e. g. Brovelli et al., 2004; Ding et al., 2006), and the comparison of causal effects between the opposite directions or between different combinations of nodes (e. g. Roebroeck et al., 2005; Bressler et al., 2008). It is also usual to compare not just which are the significant causal interactions but also their strength for different tasks or conditions (e. g. Roebroeck et al., 2005).

Overall, we see that, despite its formal advantages, the measures of Granger causality are not much more useful than the ones implementing the MNN criterion to study causality from neuronal data. At this point it is necessary to ask what information is expected to be obtained from any measure assessing causality in a data-driven approach. Accepting, according to Pearl (2009), that causality cannot be detected in general without intervention, Granger causality measures or the MNN measures characterized here should be considered as statistical measures of dependence which are sensitive to some particular properties of the dynamics analyzed. These measure are different from other statistical measures like

correlation, coherence or mutual information principally because they are designed to be asymmetric. Adopting this point of view, the main question would be which is the advantage of using these measures which are more elaborated and usually more difficult to estimate instead of symmetric measures of dependence. This question is similar to asking why should nonlinear measure be used instead of the simpler linear ones. The support to the use of the measures of causality must come from showing that they allow characterizing relevant aspects of the dependence not reflected in the other measures.

For example, in the field where the MNN measures were developed, the study of EEG data from epileptic patients, a key objective has been to find changes in the activity which can be used to predict a seizure (see Mormann et al., 2007, for a review). If the causal measures, which consider more explicitly the temporal dynamics, would result in a better predictor, this would justify their use. The use of nonlinear instead of linear measures was supported in this context by showing, for example, that they can more reliably localize the focal hemisphere where the epileptic seizure originates (e. g. Andrzejak et al., 2006b, 2011). By contrast, to our knowledge, the advantage of using the MNN measures has always been justified *a priori* by the causal information they provide, but has not been supported by the comparison with other simpler measures. In some studies the MNN measures have been shown to be sensitive to changes between different states of the brain like, for example, awake and deep sleep (Pereda et al., 2001), or different conditions, like listening to different types of music (Bhattacharya et al., 2001). In these cases, without a comparison with the changes in more elementary properties like the autocorrelation of the dynamics or the correlation between the dynamics, it is difficult to judge the value of the measures.

The same argument is valid for Granger causality measures. The distinction between functional and effective connectivity (Friston, 1994), implies that while functional connectivity only reflects statistical dependencies, effective connectivity should be informative about real causal interactions. Recognizing some of the limitations of Granger causality, Roebroeck et al. (2009) proposed to consider the data-driven Granger causal-

ity analysis as an exploratory approach, which can provide useful information, complementary to the information from structural connectivity, to construct models. This models should then be tested in a hypothesis-driven confirmatory analysis. Similarly, in other studies, in particular when applying the pairwise Granger causality, it is adverted that, despite the terminology, *causal* interactions should not be understood as effective connectivity (Brovelli et al., 2004; Cadotte et al., 2008). However, it is then not clear what extra information is obtained with respect to the analysis of, for example, the coherence.

The study of causality in this thesis started with the concrete goal of improving the specificity of the MNN measures (Andrzejak et al., 2008; Chicharro et al., 2008; Chicharro and Andrzejak, 2009). This work is here presented in Section 2.2. This methodological study has continued by considering more generally the interpretation of causality measures. The necessity to distinguish between the assessment of the existence of a causal interaction and the quantification of its strength has led to the ongoing research (Chicharro and Andrzejak, 2010), here only briefly discussed. Although focusing on the criterion of the mapping of nearest neighbors and the measures implementing it, we provided a wider context justifying the criterion and describing its assumptions, as well as comparing it to the Granger causality criterion (Section 2.1.2). Furthermore, we discussed the problems of assessing causality from neuronal data. We showed that principal limitations are related to the assumptions and formulation of the criteria, and further limitations arise from the availability of finite data sets.

Chapter 3

Studying the neural code with spike train distances

3.1 Stimuli discrimination

3.1.1 Introduction

This second part of the thesis focuses on the study of the responses of sensory neurons to external stimuli. In contrast to the case addressed in the previous part, in this context it is known *a priori* that the causal interactions are unidirectional, from the stimulus to the neuron. The question is what properties of the stimulus are exactly encoded and how. In analogy to a communication channel (Shannon, 1948; Cover and Thomas, 2006), the mutual information has been used to examine the capacity of the neurons to convey information about the stimulus (see Rieke et al., 1997, and references therein). Alternatively, one can use some decoding strategy to see how well the stimuli can be discriminated from the elicited spike trains (see Quian Quiroga and Panzeri, 2009, for a comparison of information theory and decoding approaches). We here study a discrimination analysis that combines a measure of dissimilarity of the spike trains with a classifier to assign each spike train to the stimulus which is most likely to have elicited it. This allows one to calculate a measure of discrimination performance, like the mutual information between the stimuli and the predicted stimuli.

A wide variety of methods have been applied to quantify the similarity between spike trains (e. g. Mainen and Sejnowski, 1995; Victor and Purpura, 1996; Hunter et al., 1998; van Rossum, 2001; Schreiber et al., 2003; Kreuz et al., 2007a). A particular class of measures designed to study the neural code are time scale parametric spike train distances, mainly represented by the distances proposed by Victor and Purpura (1996) and van Rossum (2001). These spike train distances have in common that they depend on some parameter τ that determines the temporal scales in the spike trains to which the distance is sensitive. For the limit $\tau \to \infty$ these distances are only sensitive to the number of spikes in each spike train, while in the limit $\tau \to 0$ they are spikes coincidence detectors. These limits reflect the assumption of a rate code and a coincidence code, respectively. When used for the discrimination analysis, the time scale parametric spike train distances provide a quantification of the discriminative precision, associated with the optimal time scale τ^* for which the maximal mutual information between the stimuli and the predicted stimuli is found. This precision is conceptually different from the single spike precision, that indicates the jitter of each reliable spike from trial to trial. It is also different from other types of precision proposed to study the neural code (see Borst and Theunissen, 1999, for a review), like the spike timing precision obtained with the direct method used to calculate the mutual information (Strong et al., 1998). In this part of the thesis we will focus on the calculation of the mutual information and the discriminative precision from the discrimination analysis, and only in Appendix B we will compare the interpretation of these types of precisions.

Discrimination analysis has been applied to study experimental data using different dissimilarity measures. For example, the Victor distance (see Victor, 2005, for a review) has been applied to study the precision in the visual system (e. g. Victor and Purpura, 1998; Mechler et al., 1998), the auditory system (e. g. Machens et al., 2001) or the olfactory system (e. g. Macleod et al., 1998). The van Rossum distance has recently been used in a series of studies on the time scale of discrimination of natural sounds in songbirds (e. g. Narayan et al., 2006; Wang et al., 2007). The correlation-based reliability measure of Schreiber et al. (2003) has also

been used to analyze auditory responses (Wang et al., 2007). A common conclusion of many of these studies is that the optimal time scale found indicates that the temporal structure of the responses and not only the total spike counts are relevant for the neural encoding of the stimuli.

Apart from applications to experimental data, the time scale parametric spike train distances have been previously characterized (Victor and Purpura, 1997; Kreuz et al., 2007a; Dauwels et al., 2009; Kreuz et al., 2009; Paiva et al., 2010; Kreuz et al., 2011). Victor and Purpura (1997) characterized the Victor distance and showed examples of simulated data for which the discrimination analysis resulted in an optimal time scale consistent with *a priori* expectations given the construction of the spike trains. The main focus of other studies was to compare the sensitivity of the time scale parametric distances and other measures of spike train similarity to discriminate between different stimuli, without addressing the meaning of the optimal time scale at which the sensitivity was optimized (e. g. Kreuz et al., 2007a; Paiva et al., 2010). We here follow Victor and Purpura (1997) considering the time scale parametric spike train distances integrated in the discrimination analysis. We are interested in examining how informative they are about the neural code, the mutual information, and discriminative precision obtained with this analysis. In a first step, we use simulated Poisson spike trains to characterize the influencing factors on the estimation of the mutual information and the discriminative precision. In particular, we consider how the mutual information and discriminative precision depend on the measure, on the classifier used in the discrimination analysis, and on the number of trials available for each stimuli. We also study the influence of the length of the recordings used to calculate the spike train distances. Furthermore, we consider how the mutual information between the stimuli and the predicted stimuli is related to the mutual information between stimuli and responses. We also question what can be concluded from the discriminative precision about the time scale in which information is contained in the responses.

Apart from the simulated data, we also use exemplary experimental recordings. This allows us to check if the dependencies observed for simple Poisson spike trains are consistent with those found for real responses.

Moreover, we examine the degree to which the complex temporal structure of the spike trains is well represented in a single quantity, namely the optimal time scale τ^* associated with the discriminative precision. Data used in this study were delivered via neurodatabase.org, a neuroinformatics resource funded by the Human Brain Project (http://neurodatabase.org) (Gardner, 2004), and provided by the Laboratory of Auditory Neuroethology and Communication of Tim Gentner, at the University of California, San Diego.

The remainder of the Chapter is organized as follows: We first describe the discrimination analysis used to calculate the mutual information and discriminative precision (Section 3.1.2), as well as the measures of spike train dissimilarity studied (Section 3.1.3). In Section 3.2 we apply the measures to simulated time-independent Poisson processes (Section 3.2.1), and the discrimination analysis to time-dependent Poisson processes (Section 3.2.2). Examples of time-dependent Poisson processes are presented to study the dependence of the mutual information and the optimal time scale on different factors. We examine the dependence on the measure and classifier used for the discrimination analysis. We simulate different examples representing possible types of responses to the transient presentation of a constant stimuli and study the dependence of the mutual information and the optimal time scale on the length of the after-transient interval included in the discrimination analysis. Furthermore, we analyze the influence of the number of trials available for each stimulus. In Section 3.3 we apply the discrimination analysis to spike trains elicited by transient constant stimuli. We examine the dependence on the measure and classifier in Section 3.3.1, and the temporal accumulation of information obtained from increasing the length of the recordings used to calculate the spike train dissimilarity measures (Section 3.3.2). We also analyze the temporal distribution of information and the redundancy between the information contained in different intervals of the recordings (Section 3.3.3), and examine how this temporal distribution is reflected in the optimal time scale related to the discriminative precision. In section 3.4 we apply the discrimination analysis to spike trains elicited by time-dependent stimuli. Finally, in Section 3.5 we discuss our results and how

to evaluate the quantities obtained from the discrimination analysis. For the sake of completeness, in Appendix B, we compare the discriminative precision with the spike timing precision obtained when the direct method (Strong et al., 1998) is used to calculate the mutual information.

3.1.2 Discrimination analysis

The mutual information $I(\mathbf{R}, S)$ (Cover and Thomas, 2006) is used to characterize the encoding of the stimuli S by the responses $\mathbf{R}$ elicited by them. For the spike trains, the space of $\mathbf{R}$ is determined by the times of the spikes. If the spike trains are binned in N_b bins and r_i is the number of spikes in each bin, $\mathbf{R} = \{r_1, r_2, ...r_{N_b}\}$ will have a high dimensionality for small bins. The mutual information is calculated as:

$$I(\mathbf{R}, S) = \sum_{\mathbf{R}} \sum_{S} p(\mathbf{R}, S) \log \frac{p(\mathbf{R}, S)}{p(\mathbf{R})p(S)}, \tag{3.1}$$

where $p(\mathbf{R}, S)$ and $p(\mathbf{R})$, $p(S)$ are the joint and marginal distributions, respectively. The reliable estimation of this quantity from experimental data is hindered by the curse of dimensionality caused by sampling limitations (see for example Borst and Theunissen, 1999; Paninski, 2003b; Victor, 2006; Panzeri et al., 2007, for a discussion of different estimation techniques and bias corrections).

One of the alternatives to estimate $I(\mathbf{R}, S)$ consists in using a classifier to predict to which stimulus a response pertains. Given N distinct stimuli $\{S\} = S_1, ..., S_i, ..., S_N$ and M trials for each stimulus, a decoding algorithm assigns each spike train to one of the stimuli $\{S\}$ resulting in a confusion matrix $\mathcal{C}(S_k^P, S_l)$ that indicates the number of times a spike train corresponding to a trial of the stimulus S_l is assigned to the stimulus S_k^P. The confusion matrix is normalized to a probability matrix $P(S_k^P, S_l)$ dividing by the total number of spike trains NM. The decoding performance can be quantified by the mutual information $I(S^P, S)$ between the stimuli and the predicted stimuli (Quian Quiroga and Panzeri, 2009), analogously to Equation 3.1. Since the classification is based only on $\mathbf{R}$, S^P is conditionally independent from S given $\mathbf{R}$, that is $S \rightarrow \mathbf{R} \rightarrow S^P$

forms a Markov Chain. In consequence, the data processing inequality (Cover and Thomas, 2006) assures that $I(S^P, S) \leq I(\mathbf{R}, S)$. Therefore, the mutual information obtained from the classifier is a lower bound for the total amount of information between the stimuli and the responses.

Different types of classifiers have been proposed depending on the nature and assumptions made about the stimuli and responses (e. g. Bialek et al., 1991; Rolls et al., 1997). The discrimination analysis proposed by Victor and Purpura (1996, 1997) is based on a classifier in which the spike train distances are used to compute the distance from a spike train $s_{l,j'}$ to a predicted stimulus S_k^P as

$$d(S_k^P, s_{l,j'}; \tau) = [\langle (D(s_{k,j}, s_{l,j'})[\tau])^z \rangle_j]^{1/z}, \tag{3.2}$$

so that the average of the spike train distances $D(s_{k,j}, s_{l,j'})[\tau]$ between the spike train $s_{l,j'}$, corresponding to the j'-th trial of the l-th stimulus, and all the trials $j = 1, ...M$ of the k-th stimulus is calculated (note that trial $j = j'$ is excluded if $k = l$). The classification is carried out by assigning each spike train to the stimulus that minimizes this average distance. The classification depends on the parameter τ, leading to a whole range of, in general, different confusion matrices. The parameter z determines which geometry is considered for the average. For positive exponents the preeminent contribution comes from the larger distances, while for negative exponents the closest spike trains have a higher weight. The differences in the confusion matrices obtained for different exponents can help to infer the structure of the clusters of trials of the same stimulus and consequently the variability of the spike trains (Victor and Purpura, 1997). The mutual information between the actual and the assigned stimuli, given the probability matrix $P(S_k^P, S_l; \tau)$, is obtained for each time scale:

$$I(S^P, S; \tau) = \sum_{k=1}^{N} \sum_{l=1}^{N} P(S_k^P, S_l; \tau) \log \frac{P(S_k^P, S_l; \tau)}{P(S_k^P; \tau) P(S_l; \tau)}. \tag{3.3}$$

The time scale τ^* for which the maximal mutual information $I_{max}(S^P, S)$ is found gives the best discrimination of the stimuli. The percentage re-

duction of uncertainty is given by:

$$I^*(S^P, S; \tau) = \frac{I(S^P, S; \tau)}{H(S)}, \tag{3.4}$$

where $H(S)$ is the entropy of the stimuli distribution (Cover and Thomas, 2006)

$$H(S) = -\sum_{l=1}^{N} P(S_l) \log P(S_l). \tag{3.5}$$

The measure $I^*(S^P, S; \tau)$ is normalized in the range $[0\ 1]$, which facilitates the comparison among results obtained for different confusion matrices, and furthermore it better quantifies the goodness of the decoding (Borst and Theunissen, 1999). Since the stimuli distribution $P(S)$ is well known, $H(S)$ can be calculated exactly, so that $I^*(S^P, S; \tau)$ is also a lower bound of $I^*(\mathbf{R}, S)$, defined analogously to Equation 3.4. How close $I^*_{max}(S^P, S)$ is to $I^*(\mathbf{R}, S)$ depends on the particular representation of the stimuli in the $\mathbf{R}$ space. This is because, even with an optimal classifier, $I^*_{max}(S^P, S) \leq I^*(\mathbf{R}, S)$ (Quian Quiroga and Panzeri, 2009). Furthermore, for a suboptimal classification, the difference between $I^*_{max}(S^P, S)$ and $I^*(\mathbf{R}, S)$ will depend on the measure used to calculate the dissimilarity and on the exponent z in Equation 3.2. The classifier also depends on the number of trials M, which affects the distribution of distances between one particular spike train and the spike trains elicited for all the trials of a particular stimulus. In the limit of $M \to \infty$, this distribution of distances will be determined by the distribution of $p(\mathbf{R}|S)$ and will converge. However, this does not mean that $I_{max}(S^P, S)$ has to increase monotonically with M. Furthermore, although the estimation of the mutual information with the classifier avoids the bias caused by the estimation of probability distributions in high-dimensional spaces, one has to correct for the upward bias associated with the baseline level of correct classifications that would already result from a random assignment of the trials to a predicted stimulus. This bias is estimated by a random resampling of the spike trains. The trials are shuffled across stimuli and the average mutual information is calculated across multiple repetitions of

the random shuffling. This correction is subtracted from the mutual information calculated for the original set, and the mutual information is considered different from zero only for positive values.

While $I^*_{max}(S^P, S)$ is simply the maximum of $I^*(S^P, S; \tau)$, getting the optimal time scale τ^* is more complicated. This is because $I^*(S^P, S; \tau)$ can have piecewise constant intervals (e. g. Victor and Purpura, 1996; Di Lorenzo and Victor, 2003; Roussin et al., 2008; Huetz et al., 2006, 2009). Often experimental studies that apply the discrimination analysis lack an explicit explanation of how to calculate τ^*. In Huetz et al. (2006) and Wohlgemuth and Ronacher (2007), the average of τ is taken when a plateau is obtained. In Roussin et al. (2008) they comment that the influence of temporal coding may be underestimated because they consider as rate coding any case in which $I^*_{max}(S^P, S)$ is obtained for $\tau = \infty$, even if lower time scales provide the same information. Instead of using the position of the peak to indicate the discriminative precision, Reich et al. (2001b) use τ_{cut}, the minimal time scale at which half of the maximal information is retrieved.

We here follow two different criteria, each emphasizing different aspects of the shape of $I(S^P, S; \tau)$ and its estimation. For simulated data, we can generate K independent realizations of the set of trials for each stimuli, and consider the average and the variability of the results of the discrimination analysis across realizations. For each realization we retrieve a value of τ^* as the mean of all τ values for which $I^*_{max}(S^P, S)$ is obtained. Accordingly, we report the average of $\langle \tau^* \rangle$ across realizations and we quantify its variability by the range of τ^*. Alternatively, we calculate the average of $I^*(S^P, S; \tau)$ for the different realizations, and then we calculate $\tau^*_{\langle I \rangle}$ only once from the average shape. In this case we additionally report the interval containing values of $I^*(S^P, S; \tau)$ higher than a given percentage of the maximum (Wohlgemuth and Ronacher, 2007). Since the maximum and mean operations are not commutative, $\langle \tau^* \rangle$ and $\tau^*_{\langle I \rangle}$ will generally not coincide. Furthermore, while for the first calculation the range of τ^* quantifies the variability of the estimator, in the second case the interval of high values is related to the flatness of $I^*(S^P, S; \tau)$. The same procedure can be applied for experimental data

carrying out the calculations across K bootstrapping realizations instead of independently generated data. For each bootstrapping realization the discrimination analysis is repeated randomly selecting a subset M' of the trials for each stimulus.

Apart from τ^*, to analyze the contribution of temporal coding it is necessary to compare the maximum mutual information $I^*_{max}(S^P, S)$ with the information that can be retrieved based only on a rate code, $I^*_{count}(S^P, S)$. One can calculate an index of temporal coding like the percentage of increase (e. g. Mechler et al., 1998; Di Lorenzo and Victor, 2003; Roussin et al., 2008)

$$\Theta = \frac{I^*_{max}(S^P, S) - I^*_{count}(S^P, S)}{I^*_{count}(S^P, S)}. \tag{3.6}$$

Alternatively, Reich et al. (2001b) normalized by $I^*_{max}(S^P, S)$ instead of $I^*_{count}(S^P, S)$. Another possibility is to use the difference with no normalization (e. g. Samonds and Bonds, 2004; Di Lorenzo et al., 2009).

To complete the description of the discrimination analysis we also briefly mention an alternative classifier that has been used to construct the confusion matrix (e. g. Machens et al., 2003; Narayan et al., 2006). In this case, for each stimulus S_i the spike train of one of the trials, s_{i,j_T}, is chosen as a template and each spike train $s_{i',j'}$ is assigned to the stimulus with the nearest template. This procedure is repeated for K_T randomly chosen templates. As a measure of discrimination performance, the percentage of correctly assigned spike trains is used. This procedure does not introduce any extra parameter and thus avoids the ambiguity arising when results depend on the exponent z of Equation 3.2. The maximal mutual information and the percentage of correct classifications are not biunivocally related (Thomson and Kristan, 2005) and quantify different aspects of the representation of the stimuli in the responses. For the purpose of our study, we will restrict ourselves to applying the classifier associated with Equation 3.2.

3.1.3 Spike train distances

We here describe two spike train distances, proposed by Victor and Purpura (1996), van Rossum (2001), and a dissimilarity measure based on the correlation-based spike train similarity measure proposed by Schreiber et al. (2003). We furthermore describe the calculation of a spike train distance based on binning (Schnupp et al., 2006). We will refer to these measures as D_{V}, D_{R}, D_{S}, and D_{B}, respectively. These measures quantify the degree of dissimilarity of two spike trains, with spikes occurring at times $\{t_i^{x,y}\} = t_1^{x,y}, t_2^{x,y}, ..., t_{n_{x,y}}^{x,y}$, where $n_{x,y}$ denote the total number of spikes in each train.

Victor distance D_{V}

Victor and Purpura (1996) introduced different families of distances each motivated by different neurobiological mechanisms. These distances are all based on the point process nature of spike trains. We use only the distance denoted $D^{spike}[q]$, which we denote as $D_{\mathrm{V}}[\tau_{\mathrm{V}}]$. Among the distances proposed by Victor and Purpura (1996), this one is the most similar to the other measures considered here. The motivation for D_{V} was to capture the behavior of neurons as coincidence detectors. It is defined as the minimal total cost of a sequence of elementary steps that transform one spike train into another. The allowed operations are insertion, deletion, and temporal shift of a spike. While the cost of deletion and insertion is fixed to 1, the cost of a shift by Δt is $q|\Delta t|$, where q is the cost per time unit. For a shift of length $|\Delta t| = \frac{2}{q}$ the cost of shifting is equal to the one of deleting the pair of spikes. Therefore, $\frac{2}{q}$ is the maximal separation of the spikes for which the shift operation is preferred. Given that, we here use a parameter in units of time defined as $\tau_{\mathrm{V}} = \frac{2}{q}$.

Each value of τ_{V} identifies one member of the family of distances. At one extreme, $D_{\mathrm{V}}[\tau_{\mathrm{V}} = \infty] = |n_x - n_y|$ is only sensitive to the difference in the total number of spikes in each train. Accordingly, it quantifies the similarity of the spike trains assuming a rate code. The distance D_{V} monotonically increases for τ_{V} decreasing. In the other extreme, assuming that the sampling resolution is high enough so that there are no coin-

Table 3.1: Limits of the spike train dissimilarity measures

Measure	$\tau = \infty$	$\tau = 0$
D_{V}	$\lvert n_x - n_y \rvert$	$n_x + n_y$
D^*_{V}	$\frac{\lvert n_x - n_y \rvert}{n_x + n_y}$	1
D_{R}	$(n_x - n_y)^2$	$n_x + n_y$
D_{S}	0	1
D_{B}	$(n_x - n_y)^2$	$n_x + n_y$

cident spikes, $D_{\mathrm{V}}[\tau_{\mathrm{V}} = 0] = n_x + n_y$. This magnitude of the distances for $\tau_{\mathrm{V}} \to 0$ is not normalized, but depends on the total number of spikes $n = n_x + n_y$. A normalized version of the Victor distance, $D^*_{\mathrm{V}} = \frac{D_V}{n_x + n_y}$, was introduced by Kreiman et al. (2000), resulting in the limits indicated in Table 3.1.

van Rossum distance $D_{\mathbf{R}}$

The second spike train distance studied here was introduced by van Rossum (2001). This distance operates on continuous signals obtained by convolution of the spike trains with a truncated exponential function:

$$f_{x,y}(t) = \sum_i^{n_{x,y}} u(t - t_i^{x,y}) e^{-(t - t_i^{x,y})/\tau_{\mathrm{R}}}, \tag{3.7}$$

where u is the Heaviside step function ($u(t - t_i) = 0$ if $t < t_i$ and $u(t - t_i) = 1$ if $t \geq t_i$). Similarly to τ_{V}, the time constant τ_{R} determines the precision to which the distance is sensitive to. The truncated exponential convolution kernel is motivated by its causality and its correspondence to the shape of postsynaptic currents. Given two continuous signals the distance is calculated as the L^2 norm

$$D_R[\tau_{\mathrm{R}}] = \frac{1}{\tau_{\mathrm{R}}} \int_0^\infty [f_x - f_y]^2 dt. \tag{3.8}$$

As the Victor distance, the van Rossum distance avoids a coarse binning procedure. The time constant of the exponential function τ_{R} determines

the interval for which relevant contributions to the integral last. Note that this interval generally extends also beyond the end of the spike trains. This distance might be seen as a rate profile comparison in which a convolution kernel is used to smooth the rate profiles (Dayan and Abbot, 2001). Any other convolution kernel or another L^n norm could be used. We will use D_{R^*} to refer to the van Rossum spike train distance calculated with a rectangular convolution kernel instead of the usual truncated exponential kernel. For the rectangular kernel we will take the parameter τ_{R} to be the standard deviation of the kernel, like for the truncated exponential kernel. The dependence of D_{R} for the two extremes of τ_{R} is independent of the kernel function used up to a constant factor. At one extreme, $D_{\mathrm{R}}[\tau_{\mathrm{R}} \to 0] \sim n_x + n_y$ is a coincidence detector. In the opposite limit $D_{\mathrm{R}}[\tau_{\mathrm{R}} \to \infty] \sim (n_x - n_y)^2$. In consequence, in contrast to D_{V}, D_{R} is not monotonic with τ_{R}, and the difference between the magnitude of the values at the two extremes (Table 3.1) is in general higher.

Schreiber dissimilarity D_{S}

The third measure we study is a dissimilarity measure derived from the correlation-based reliability measure of Schreiber et al. (2003). At first the spike trains are convolved with a Gaussian convolution kernel to obtain two continuous signals

$$s_{x,y}(t) = \sum_{i}^{n_{x,y}} \frac{1}{\sqrt{2\pi\tau_{\mathrm{S}}^2}} e^{-(t-t_i^{x,y})^2/2\tau_{\mathrm{S}}^2}, \tag{3.9}$$

which become vectors $\vec{s}_{x,y}$ when the convolution is applied for a discrete number of sampling times. Analogous to the continuous signals constructed for D_R, the length of $\vec{s}_{x,y}$ is determined not only by the length of the spike trains but also by the width τ_{S} of the convolution kernel. In contrast to the truncated exponential convolution kernel used for D_{R}, the symmetric Gaussian convolution kernel is motivated by the idea that the variability in the spike times is a jitter which can be either positive or

negative. The similarity measure is defined as the cross correlation

$$S_{corr} = \frac{\vec{s}_x \cdot \vec{s}_y}{|\vec{s}_x||\vec{s}_y|}. \tag{3.10}$$

A dissimilarity measure based on S_{corr} is implemented as $D_{\mathrm{S}} = 1 - S_{corr}$. Here the standard deviation τ_{S} of the Gaussian convolution kernel is the parameter controlling the precision to which D_{S} is sensitive to. The measure S_{corr} corresponds to an angle in the space of the spike trains (Paiva et al., 2009). Therefore, the dependence of S_{corr} on the width of the convolution kernel τ_{S} differs significantly from the one of D_{R} on τ_{R} (Table 3.1). Again, this dependence is independent of the particular convolution kernel. The limit $D_{\mathrm{S}}[\tau_{\mathrm{S}} \to 0]$ is consistent with the ones obtained for $D_{\mathrm{V}}[\tau_{\mathrm{V}} \to 0]$ and $D_{\mathrm{R}}[\tau_{\mathrm{R}} \to 0]$, corresponding to the maximal sensitivity to spike timing. Assuming no exact coincidence of spikes, in this limit $\vec{s}_x \cdot \vec{s}_y = 0$ and thus $D_{\mathrm{S}} = 1$. By contrast, for $\tau_{\mathrm{S}} \to \infty$ the convoluted signals $\vec{s}_{x,y}$ will be both constant with a different value determined by the respective number of spikes in each train, $n_{x,y}$. However, due to the normalized dot product, $\frac{\vec{s}_x}{|\vec{s}_x|} \cdot \frac{\vec{s}_y}{|\vec{s}_y|}$ will always have the maximum value of 1, so that $D_{\mathrm{S}}[\tau_{\mathrm{S}} \to \infty] = 0$. In consequence, $D_{\mathrm{S}}[\tau_{\mathrm{S}} \to \infty]$ cannot be associated with a rate code, since it is not sensitive to the difference in the number of spikes. Clearly, the dependence of D_{S} on τ_{S} does not reflect different degrees of sensitivity to the temporal structure, going from a rate code to a coincidence code. Accordingly, in contrast to D_{V}, D^*_{V}, and D_{R}, D_{S} does not fulfill the requirements to study the discriminative precision (Victor and Purpura, 1997).

Binning distance $D_{\mathbf{B}}$

The calculation of the mutual information with spike train distances in combination with a classifier has been motivated to avoid the use of binning in the estimation of the probabilities $p(\mathbf{R}, S)$ (Section 3.1.2). To check to which degree binning is inconvenient by itself or becomes problematic only due to the high dimensionality of the $\mathbf{R}$ space, we also study a spike train distance based on binning (Schnupp et al., 2006). For each

spike train, the spikes are binned in N_B bins of width τ_{B}. The distance D_{B} is calculated straightforwardly from the differences between the number of spikes in each bin for two spike trains. Like for D_{R}, the sum of the squared differences is here used unless stated otherwise. We will use D_{B^*} to refer to the bins distance calculated by taking the absolute value of the difference instead of the squared difference. Accordingly, this distance is proportional to D_{R} in the two extremes of τ_{B} (Table 3.1). For $\tau_{\mathrm{B}} \to 0$ the bins will contain at maximum one spike, and the full bins will generally not coincide for each spike train. In consequence, $D_{\mathrm{B}}[\tau_{\mathrm{B}} = 0] = n_x + n_y$. Oppositely, if only one bin is used for the whole length L of the spike trains, we get $D_{\mathrm{B}}[\tau_{\mathrm{B}} = L] = (n_x - n_y)^2$. The main difference with D_{V} and D_{R} is that the rate code is not only represented by $\tau = \infty$, but already by $\tau = L$, the length of the spike trains.

General remarks

We finish the description of the measures with some general remarks. In first place, as already pointed out by Victor and Purpura (1997), for certain values of the parameters the spike train distances do not accomplish the conditions required to be a distance, in the strict sense. In particular, for $D_{\mathrm{V,R,B}}[\tau_{\mathrm{V,R,B}} = \infty]$, zero values are not only obtained for identical spike trains, but also for spike trains with equal number of spikes. Similarly, for the dissimilarity measure $D_{\mathrm{S}}[\tau_{\mathrm{S}} \to \infty]$, zero values are obtained for any pair of spike trains, and it is clearly not a distance. Being aware of these violations we still generally refer to $D_{\mathrm{V,R,B}}$ as spike train distances and consider that identity is redefined in these cases to the classes of spike trains leading to zero distance. In second place, there is no unique way to relate the parameter of the distances to a time scale of the spike trains. For D_{V} we have chosen $\tau_{\mathrm{V}} = \frac{2}{q}$. This selection has been used in several previous studies (Victor and Purpura, 1996; Reich et al., 2001b; Huetz et al., 2009), but it is also common to use $\tau_{\mathrm{V}} = \frac{1}{q}$ (e. g. Victor and Purpura, 1997; Samonds and Bonds, 2004; Huetz et al., 2006). Notice that, although for an individual pair of spikes $2/q$ corresponds to the maximum time shift allowed, for complete spike trains $\tau_{\mathrm{V}} = 2/q$ does not indicate

the time scale of the time shifts actually used in the minimal cost transformation. For D_{R} and D_{S} we take as the time scale parameter the standard deviation of the convolution kernel. An alternative criteria could be the decay time of the convolution kernel (Wang et al., 2007). For D_{B}, the time scale parameter is the width of the bins. This arbitrariness introduces some ambiguity in the alignment of the parameter τ for different measures and the comparison of the obtained τ^*. Furthermore, except for D_{B}, the time scale parameter can take values higher than the length of the spike trains L without being associated with a rate code.

3.2 Discrimination analysis with Poisson spike trains

3.2.1 Spike train distances for time-independent Poisson spike trains

We first examine the spike train distances for pairs of Poisson spike trains with distinct time-independent rates λ_x, λ_y. We consider as an *a priori* condition of consistency that the lowest distance to spike trains generated with rate λ_x should be obtained for $\lambda_y = \lambda_x$. In Figure 3.1 we show the results obtained for D_{V}, D^*_{V}, D_{R}, D_{S}, and D_{B}. We see that D_{V}, D_{R}, and D_{B} have a similar behavior. From high to low τ the consistency is gradually lost. This loss can be understood from the lack of normalization of these spike train distances. The maximum possible value of the distances depends on λ_x and λ_y, that determine the total number of spikes $n = n_x + n_y$. In particular, for a fixed λ_x, a lower λ_y results in a lower maximum possible value. For example, for D_{V}, when τ approaches the time scale indicated by $\frac{1}{\lambda_{x,y}}$, the effect of the different maximum possible values corresponding to different n becomes progressively more relevant, since insertions and deletions dominate over shifts. In consequence, for small τ the lowest distances to spike trains generated with a rate λ_x are not obtained for $\lambda_y = \lambda_x$ anymore, but for $\lambda_y \ll \lambda_x$.

The loss of consistency of D_{V}, D_{R}, and D_{B} also results in a worse discrimination between the stimuli whose responses are Poisson spike trains with different rates. For example, for $\tau = 10$ ms, a classifier using any of these three metrics assigns all the spike trains to the stimulus eliciting the

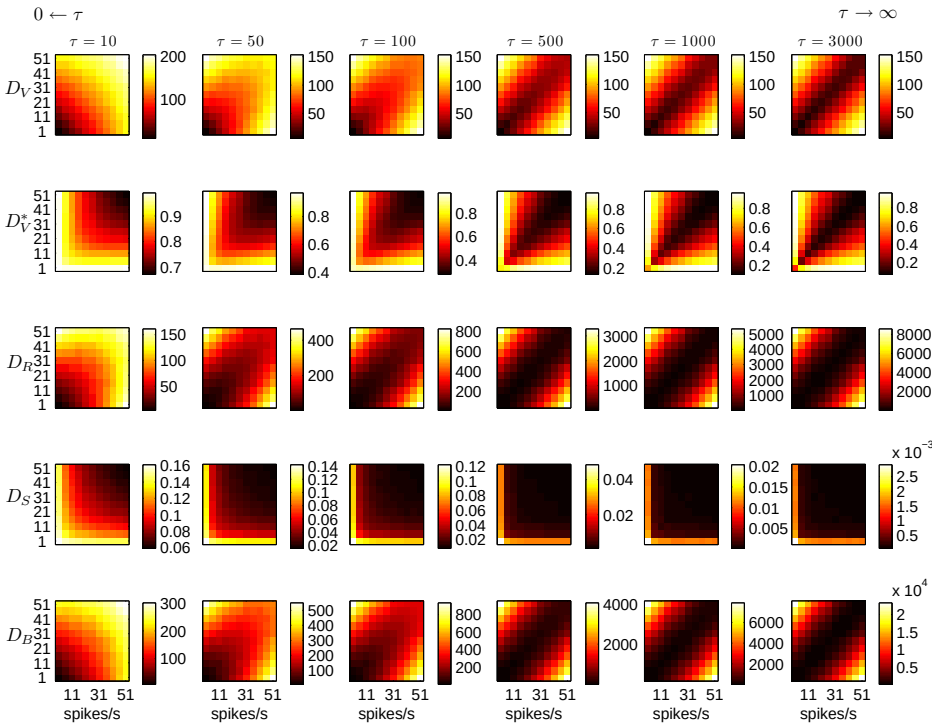

Figure 3.1: Measures of spike train dissimilarity in dependence on the rates of two time-independent Poisson spike trains. Each column corresponds to a fixed value of the time scale parameter τ (ms). $M = 50$ realizations of length 3 s were generated for each rate. Values shown are averages over all the possible pairs of spike trains for two given rates. Notice that the scale is set independently for each subplot.

lowest rate. For non-Poissonian spike trains the time scale τ at which the consistency is lost will depend on their structure and the single spike reliability and precision. The main difference between these three distances is the dependence of their magnitudes on τ. As discussed in Section 3.1.3, this is because of the specific functionality on the sum or the difference of total spikes in the limit of small or high τ, respectively (Table 3.1).

The remaining two distances behave differently. In contrast to D_{V}, in the limit of small τ_{V}, low values of D_{V}^{*} occur for high rates. This indicates that although the absolute transformation cost increases with the rates, the cost normalized by the maximum possible cost is indeed lower. For $\tau = 10$ ms we see that the range of values decreases, approaching 1. Since $D_{\mathrm{V}}^{*}[\tau_{\mathrm{V}} \to 0] \to 1$, towards this limit the distance progressively will loose all its specificity. However, how this affects the classification will depend on wether the distances resulting from the same or different stimuli are still separable. For $\tau_{\mathrm{V}} \to \infty$ the consistency condition is fulfilled, but the distance by construction does not only depend on the difference between the rates, due to the normalization (Table 3.1).

Regarding D_{S}, values close to zero are obtained regardless of the rate difference in the limit $\tau_{\mathrm{S}} \to \infty$. In this limit, since the convolution kernel is broad, the number of spikes contributing to the convolution at each point in time is high enough to give an estimation of the time-independent rate with low variability. In consequence, for large τ_{S}, $\frac{s_i}{|\vec{s}|} \simeq \frac{1}{L}$ independently of the value of the rate. Only for distances involving the lowest rates slightly higher D_{S} values are obtained. This is because for these lowest rates even for large kernel widths a good estimation of the rate is generally not obtained at each point in time. In the limit $\tau_{\mathrm{S}} \to 0$, low values occur for high rates as an effect of the normalization analogous to D_{V}^{*}.

These results confirm that all the measures except D_{S} cover the range of time scales sensitivity going from a coincidence detector to a rate code distance. However, each measure has particular magnitudes and dependencies on the rates, and therefore may lead to a different discrimination of the stimuli associated to the rates. By contrast, $D_{\mathrm{S}} = 0$ independently of the rates for $\tau_{\mathrm{S}} = \infty$ and thus D_{S} cannot be used as a measure to study

the degree of discrimination obtained with a rate code. Therefore, for the rest of this Chapter we will restrict ourselves to the other measures.

3.2.2 Information and discriminative precision for time-dependent Poisson spike trains

Dependence on the measure and the classifier

We here examine the dependence of $I^*(S, S^P; \tau)$ and τ^* on the measure and classifier. We simulate time-dependent Poisson spike trains in which the only source of temporal structure is the modulation of the rates. We compose the spike trains by concatenation of several intervals of length L_i, where $i = 1, ..., N_L$. For each interval time-independent Poisson processes are generated. Furthermore, only two stimuli S_x and S_y are considered, so that the spike trains are fully described by the rates $(\lambda_{x,i}, \lambda_{y,i})$.

In Figure 3.2 we show the dependence of $I^*(S^P, S; \tau)$ on the time scale τ for four examples with different configurations of the time-dependent rates. For all examples $N_L = 4$. In example A the rates on each interval for the two stimuli are such that the stimuli can be well discriminated by the total rates. In example B the total rates still differ but are closer to each other. In examples C and D the total rates are equal, and only the local rates in each interval L_i differ. These local differences in the rate are higher for example C. Further details on the generation of the spike trains are described in the caption of Figure 3.2. We show the results for four different classifiers corresponding to different values of the exponent z (Equation 3.2).

We see that $I^*_{max}(S^P, S)$ does not critically depend on the measure. In particular, the values of $I^*_{max}(S^P, S)$ obtained by D_{B} are not lower than those obtained by the other measures. This indicates that, for these examples, this distance based on binning is as valid as the other more elaborated measures. While $I^*_{max}(S^P, S)$ is almost the same for different measures, the shape of $I^*(S^P, S; \tau)$ varies considerably across them. In the first place, for τ_{B} equal the length L of the spike trains, D_{B} already corresponds to a rate code distance. By contrast, the other measures are still sensitive to the temporal structure as long as $\tau < \infty$. Therefore, for

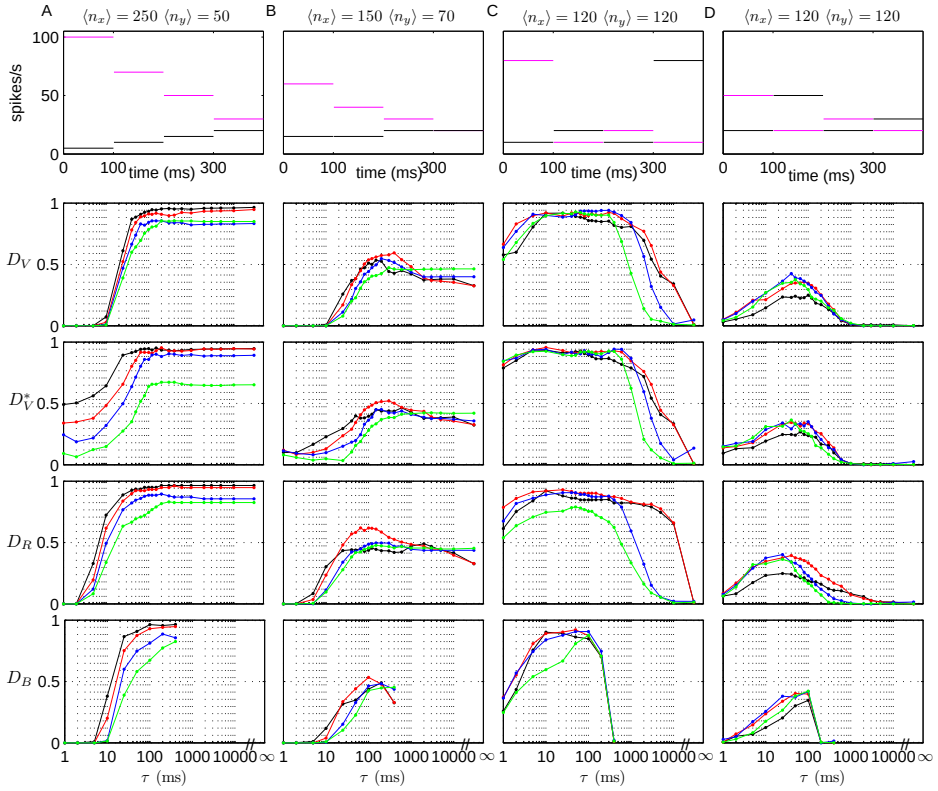

Figure 3.2: Dependence of $I^*(S^P, S; \tau)$ on the measure and the classifier. The different columns correspond to examples A-D, with different combinations of modulated rates, as shown in the first row. In all the cases the responses to two stimuli S_x, S_y are simulated as Poisson processes with piecewise constant modulated rates. For all the examples $N_L = 4$ intervals of length 100 ms are simulated with constant rates $[\lambda_{x,1}, \lambda_{y,1}; \lambda_{x,2}, \lambda_{y,2}; \lambda_{x,3}, \lambda_{y,3}; \lambda_{x,4}, \lambda_{y,4}]$ (spikes/s). A: $100, 5; 70, 10; 50, 15; 30, 20$ spikes/s. B: $60, 15; 40, 15; 30, 20; 20, 20$ spikes/s. C: $80, 10; 10, 20; 20, 10; 10, 80$ spikes/s. D: $50, 20; 20, 50; 30, 20; 20, 30$ spikes/s. We indicate the total average spike counts $\langle n_x \rangle$, $\langle n_y \rangle$. Four classifiers with different exponent (Equation 3.2) are used: $z = -8$ (black), $z = -2$ (red), $z = 2$ (blue), and $z = 8$ (green). Averages across $K = 20$ independent realizations with $M = 20$ trials per stimulus are shown.

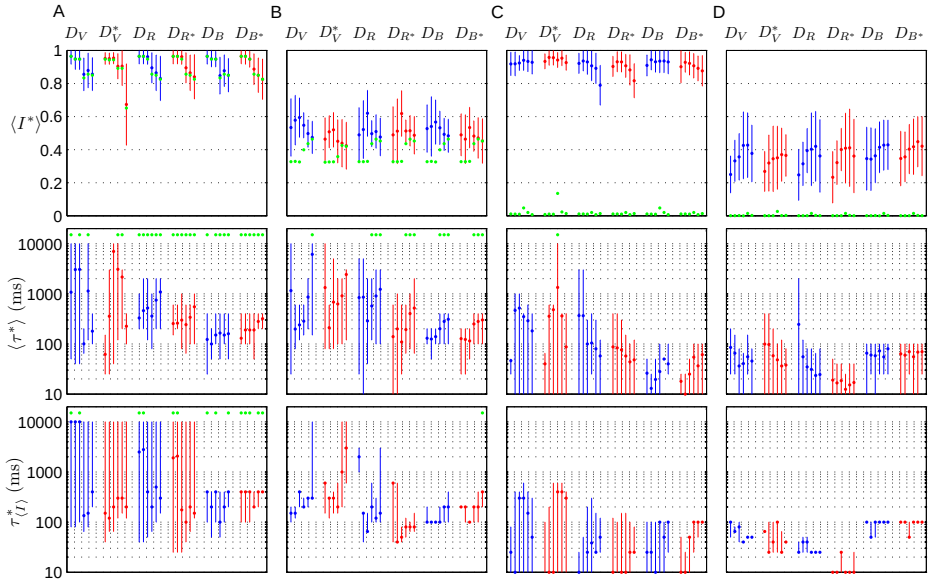

Figure 3.3: Maximal mutual information $I^*_{max}(S^P, S)$ and optimal time scale τ^* in dependence on the measure and the classifier for the examples shown in Figure 3.2. Each column shows the results for the corresponding simulations A-D in Figure 3.2. For each measure the results are shown for exponents $z = [-8, -4, -2, 2, 4, 8]$, from left to right, and alternate blue and red dots are used to distinguish the results of adjacent measures. In the first row, we also show with green dots the mutual information obtained with a rate code, $I^*_{count}(S^P, S)$. In the second and third row we indicate with green dots when $\tau = \infty$ was optimal for some realization. Details on the calculation of the averages and the error bars are provided in Section 3.1.2.

the examples C and D, in which the total rate is the same for the responses to S_x and S_y and a rate code cannot discriminate between them, the other measures still result in high values of $I^*(S^P, S; \tau)$ for time scales higher than L. On the other limit, for $\tau \to 0$, the expected decrease of the mutual information depends on the loss of consistency as discussed in Section 3.2.1. In general, $I^*(S^P, S; \tau)$ decreases more slowly for D^*_{V} than for the rest of the measures. This is due to the normalization and the effect of the maximum possible value of the distance on the loss of consistency, as described in Section 3.2.1.

Since the distances depend on τ smoothly, also $I^*(S^P, S; \tau)$ is expected to be smooth if τ is sufficiently sampled. The flatness of the peak depends on several factors. If high values of $I^*(S^P, S; \tau)$ are obtained like in examples A and C, the peak is flatter due to the saturation of the sensitivity of $I^*(S^P, S; \tau)$ when its maximum is attained. For these examples, since the stimuli are very different, they are more likely to be discriminated better at any time scale. Furthermore, the flatness depends on the sensitivity of each measure at each time scale. The most obvious example is that the sensitivity of D_{V}, D^*_{V}, and D_{R} to temporal structure for $\tau > L$, will result in a flatter peak towards high time scales τ. In the opposite limit $\tau \to 0$, the initial decrease from $I^*_{max}(S^P, S)$ is slower for D_{R}. This occurs because for D_{V} and D_{B} the resolution cannot be further increased when τ is sufficiently small. This saturation of the resolution for small τ is caused by the binning in D_{B} and by the maximal cost of 2 in D_{V}.

The exponent z of the classifier also results in significant changes in the shape of $I^*(S^P, S; \tau)$. Furthermore, z influences $I^*_{max}(S^P, S)$ more than the measure. To see more in detail the dependence of $I^*_{max}(S^P, S)$ and τ^* on the measures and exponents z, we examine in Figure 3.3 the values extracted from the shapes of $I^*(S^P, S; \tau)$ shown in Figure 3.2. We confirm that $I^*_{max}(S^P, S)$ depends more on z than on the measure. For example A, negative exponents are preferred, and for example D the positive ones. For examples B and C small $|z|$ lead to higher $I^*_{max}(S^P, S)$. The particular dependence on z changes across measures, but is qualitatively similar indicating that z is associated with the geometry resulting

from the distances. The variability of $I^*_{max}(S^P, S)$, quantified by the standard deviation across trials, does not depend crucially on the measure or exponent.

The optimal time scale $\langle \tau^* \rangle$ is much more variable across measures. Furthermore, although some trends are observed in dependence on the exponent, the clarity and shape of this dependence change from measure to measure much more than for the dependence of $I^*_{max}(S^P, S)$. The broad error bars indicate the high variability of τ^* from realization to realization of the analysis (Section 3.1.2). The same qualitative conclusions are obtained for $\tau^*_{\langle I \rangle}$ regarding the high variability. Here the error bars quantify the flatness of the peak, showing that in general D_{B} leads to a sharper peak. Comparing across examples, we see that higher values of τ^* are obtained for examples A and B, in which the responses to the two stimuli have different total rates. For example A, values of $\tau^* = \infty$ occur. For examples C and D, τ^* is lower and barely related to the time scale of the intervals ($L_i = 100$ ms). In all these examples this time scale is kept fixed, and the changes in τ^* result from the local balance of the different rates in response to the two stimuli. Furthermore, comparing D_{R} and D_{R^*}, notice that τ^* does not only depend on the distance used but also on the kernel. The rectangular convolution kernel, which is symmetric with respect to the spike time, generally gives the lowest values of τ^*. This is because apart from the higher resolution of D_{R} discussed above, it adds the fact that it is effectively wider than the truncated exponential convolution kernel.

Apart from determining τ^*, to assess the relevance of temporal coding in the discrimination of the stimuli, we need to compare $I^*_{max}(S^P, S)$ and $I^*_{count}(S^P, S)$. For examples C and D, in which the total rates are equal, $I^*_{count}(S^P, S) \simeq 0$ independently of the measure and exponent. For examples A and B, $I^*_{count}(S^P, S)$ depends more on z than on the measure. Accordingly, Θ varies substantially when $I^*_{max}(S^P, S)$ and $I^*_{count}(S^P, S)$ calculated with the same z are used to calculate it. In particular, for example B, while for positive exponents the role of temporal coding is almost irrelevant, for negative exponents $\Theta > 0.5$. Considering the variability observed for τ^* and for the difference between $I^*_{max}(S^P, S)$ and

$I^*_{count}(S^P, S)$, we see that the discriminative precision and the role of temporal coding are more difficult to evaluate than the mutual information between stimuli and predicted stimuli.

Dependence on the length of the spike trains for different simulated codes

In the examples above we have not made any assumption about the cause of the modulation. In particular the modulation can represent the effect of short-term adaptation in the transient responses to the onset of a constant stimulus, or the effect of correlations with a dynamic stimulus. In the case of the transient responses, the selection of a given length of the spike trains for the discrimination analysis involves some assumption about the time during which the constant stimuli are encoded. To further study the relation of the discriminative precision with the properties of the code, we examine how it depends on the length of the interval of the spike train for which the distances are calculated.

For transient responses, latency has been shown to contain a high amount of information about the stimuli (e. g. Reich et al., 2001b; Panzeri et al., 2001). To model this effect we consider two constant stimuli S_x and S_y that elicit, during a short fixed interval, the same elevation of the rate with respect to a baseline level, but with different delays with respect to the onset time. Furthermore, after this transient response, a constant rate is associated to each stimulus. In a first example (E) these after-transient rates are equal, while for a second example (F) the rate is higher for the stimulus with the larger latency. This difference in the after-transient rates, depending on the length L of the spike trains used for the discrimination analysis, suffices to produce a rate code discrimination. Further details of the simulations are given in the caption of Figure 3.4. Here we only show the results for D_{V}, but the conclusions about the estimation and interpretation of the mutual information and discriminative precision are qualitatively similar for the other measures.

We examine how $I^*_{max}(S^P, S)$ and τ^* depend on the latency difference L_2 and on the after-transient interval length L_4. In example E the

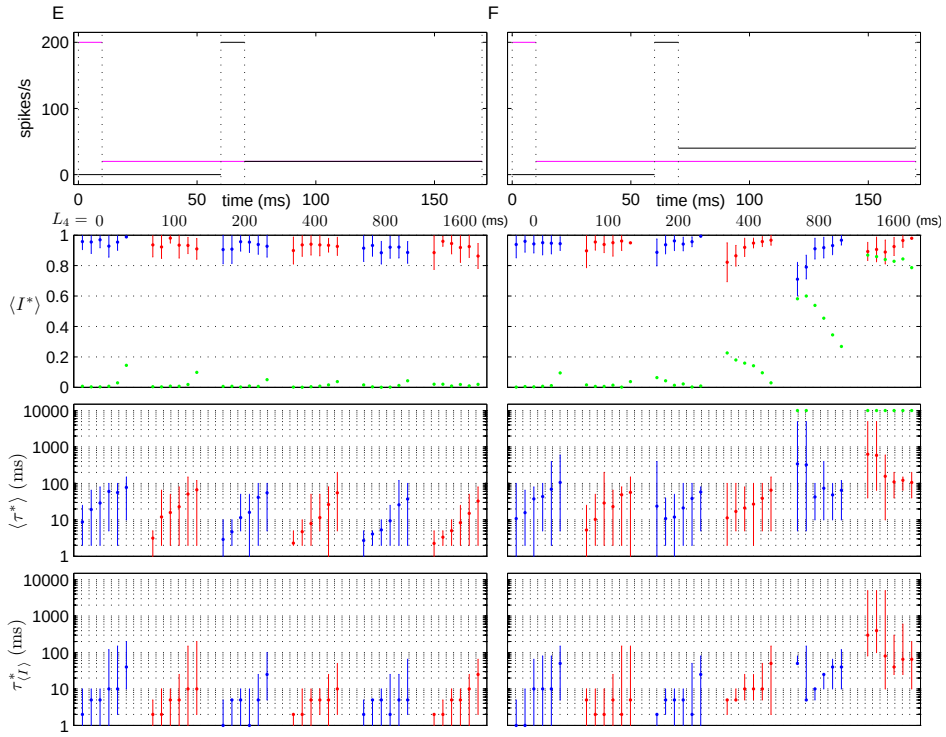

Figure 3.4: Dependence of $I^*_{max}(S^P, S)$ and τ^* on the latency difference L_2 and the length of the after-transient interval L_4, for spike trains simulating the response to transient constant stimuli. Four intervals with constant rates for each stimulus are generated as shown in the first row. $L_1 = 10$ ms models the transient response of stimulus S_x. $L_2 = [0, 10, 25, 50, 100, 200]$ ms models the latency difference. $L_3 = 10$ ms corresponds to the time of the transient response of S_y. $L_4 = [0, 100, 200, 400, 800, 1600]$ ms is the after-transient interval. Following the notation in Figure 3.2, the rates are: E: $[200, 0; 20, 0; 20, 200; 20, 20]$ spikes/s, F: $[200, 0; 20, 0; 20, 200; 20, 40]$ spikes/s. In the first row the modulated rates are shown using $L_2 = 50$ ms and $L_4 = 100$ ms for exemplary purpose. In example E the rates are the same during the after-transient interval. In the other rows, for each length L_4 the results are shown for all the lengths L_2 with the same color, analogously to Figure 3.3. Results are shown for D_{V} with $z = -2$, $K = 20$, and $M = 20$.

after-transient rates are equal, leading to $I^*_{count}(S^P, S) \simeq 0$ independently of L_2 and L_4. Since there is no overlap of the periods of elevated rate and the increase of the rates is high with respect to the baseline rate, high values of $I^*_{max}(S^P, S)$ are obtained independently of the latency difference. In this example, the after-transient interval length L_4 does not affect the amount of information. Therefore, Θ is also similar for different latency differences and after-transient interval length L_4, indicating a constant contribution of temporal coding to discrimination. The optimal time scales $\langle\tau^*\rangle$ and $\tau^*_{\langle I\rangle}$, while almost independent of L_4, increase consistently in dependence on the latency difference L_2. Accordingly, the discriminative precision depends only on L_2.

For example F, for which the after-transient rates differ, we get different results. For small L_4 the dependence on L_2 resembles the one found in example E as expected, since the total number of spikes is still similar for responses to both stimuli. When L_4 increases $I^*_{count}(S^P, S) > 0$ are obtained indicating that a rate code is enough to discriminate. The length L_4 necessary to obtain $I^*_{count}(S^P, S) > 0$ depends on the latency difference L_2. This is because during the latency interval the stimulus with the earlier response has a higher rate that has to be balanced in the after-transient interval by the higher rate of the other stimulus. On the other hand, $I^*_{max}(S^P, S)$ increases with the latency difference L_2 because information can be obtained from both the latency and rate difference if temporal coding is used. Since both $I^*_{count}(S^P, S)$ and $I^*_{max}(S^P, S)$ change with L_2 and L_4, so does Θ. In this example there are two different time scales at which the stimuli can be discriminated, and one or the other is more relevant depending on the balance between the latency difference and L_4. For intermediate values of L_4, $I^*_{max}(S^P, S)$ decreases for small latency differences L_2, showing that there is no optimal time scale at which the two sources of information can be combined. The optimal time scale τ^* does not mainly depend on the latency difference as in example E. For sufficiently large L_4, both $\langle\tau^*\rangle$ and $\tau^*_{\langle I\rangle}$ increase with L_4 indicating that a rate code is enough for discrimination. For small L_4 the dependence on the latency difference L_2 is still visible, in particular for $\langle\tau^*\rangle$. Altogether, we see that when different sources of information

exist that are not on the same time scale, it is difficult to interpret τ^* as something more than the parameter for which the best discrimination is obtained.

We also consider another possible encoding of the transient presentation of constant stimuli. In this case a phasic response is elicited with a different elevation of the rate for each stimulus S_x, S_y. Furthermore, we assume that after this increase the rate returns to the same baseline level independently of the stimulus. A detailed description of this example G is given in the caption of Figure 3.5. For a fixed interval of the transient response L_1 and fixed values of the rates during the transient response to each stimulus, we vary the after-transient interval and thereby the length of the spike trains used for the discrimination analysis L.

We examine the dependence of $I^*_{max}(S^P, S)$ and τ^* on L and on the exponent of the classifier. Like for example E, the after-transient rates do not add any information about the stimuli. Therefore, we expect that also here $I^*_{max}(S^P, S)$ is independent of L. However, if we restrict ourselves to a negative exponent $z = -2$ as in the Figure 3.4, $I^*_{max}(S^P, S)$ increases with L. Since there is no extra information in L_2, this increase can only be caused by a change in the ability of the classifier to extract the existent information in each case. As discussed in Section 3.1.2, $I^*_{max}(S^P, S)$ is a lower bound of the total information $I(\mathbf{R}, S)$. To have a reference of the total information in the responses, we show an estimate of $I(\mathbf{R}, S)$ according to Equation 3.1. We use only the number of spikes in the first interval to represent $\mathbf{R}$, since there is no information about the identity of the stimulus in L_2. This estimate is calculated independently using a larger number of trials per stimulus and checking for its convergence to avoid any upward bias. If we compare across L_2 the maximal $I^*_{max}(S^P, S)$ across z, we see that this maximum remains almost constant for different L_2. Oppositely, $I^*_{count}(S^P, S)$ decreases with L_2 even if the optimal z is chosen independently from the optimal z for $I^*_{max}(S^P, S)$. This is because for Poisson processes the variance is equal to the mean, and thus it is harder to use the different rates in L_1 to discriminate when L_2 increases and the number of spikes in this interval differs more. Therefore, Θ is more sensitive to the selection of L_2 than $I^*_{max}(S^P, S)$. This variability

also affects the discriminative precision. The optimal time scale estimated as the average across realizations $\langle \tau^* \rangle$ results in higher values than $\tau^*_{\langle I \rangle}$, calculated from the average shape of $I^*(S^P, S; \tau)$. The error bars of $\langle \tau^* \rangle$ indicate a high variability across single realizations. From $\tau^*_{\langle I \rangle}$ we see that the peak is sharper for high L_2. Furthermore, for the exponents for which the maximal $I^*_{max}(S^P, S)$ is obtained, τ^* is almost constant independently of L_2 but it does not correspond to the length of L_1.

Dependence on the number of trials

It depends not only on the measure and classifier but also on the number of available trials per stimulus how tight the lower bound $I^*(S^P, S; \tau)$ of $I(\mathbf{R}, S)$ is. In previous examples the number of trials was fixed. We now go back to the examples A-D shown in Figures 3.2 and 3.3 to study how $I^*_{max}(S^P, S)$ and τ^* depend on the number of trials M. We use up to $M = 50$ trials, which is already high compared to the usual number of trials available in experimental studies in which the discrimination analysis has been applied. In the limit of $M \to \infty$ we would expect $I^*_{max}(S^P, S)$ to converge to a constant value, since the trials would reflect the distribution $p(\mathbf{R}|S)$. Furthermore, for a sufficiently good classifier, if $p(\mathbf{R}|S)$ is well reflected in the trials, this will improve the classification. For classifiers that are worse, improving the estimation of $p(\mathbf{R}|S)$ can oppositely further evidence that the classifier is not optimal. To have a reference of the total information, we calculate an estimate of $I(\mathbf{R}, S)$ (Equation 3.1), using the number of spikes in each of the four intervals to represent $\mathbf{R}$. We used a sufficient number of trials so that the values of $I(\mathbf{R}, S)$ converge, avoiding the upward bias. This estimate is also a lower bound of $I(\mathbf{R}, S)$ when calculated with a higher resolution for $\mathbf{R}$.

We see in Figure 3.6 that $I^*_{max}(S^P, S)$ generally increases with M. However, for example B and $z = 8$, we see that it decreases. For $M = 50$, $I^*_{max}(S^P, S)$ attains values lower than $I(\mathbf{R}, S)$ in examples B and D, suggesting that the classifier cannot extract all the information. Oppositely, in examples A and C, for high M we see that $I^*_{max}(S^P, S)$ is a tight lower bound of $I(\mathbf{R}, S)$. Although a clear convergence to a constant value is

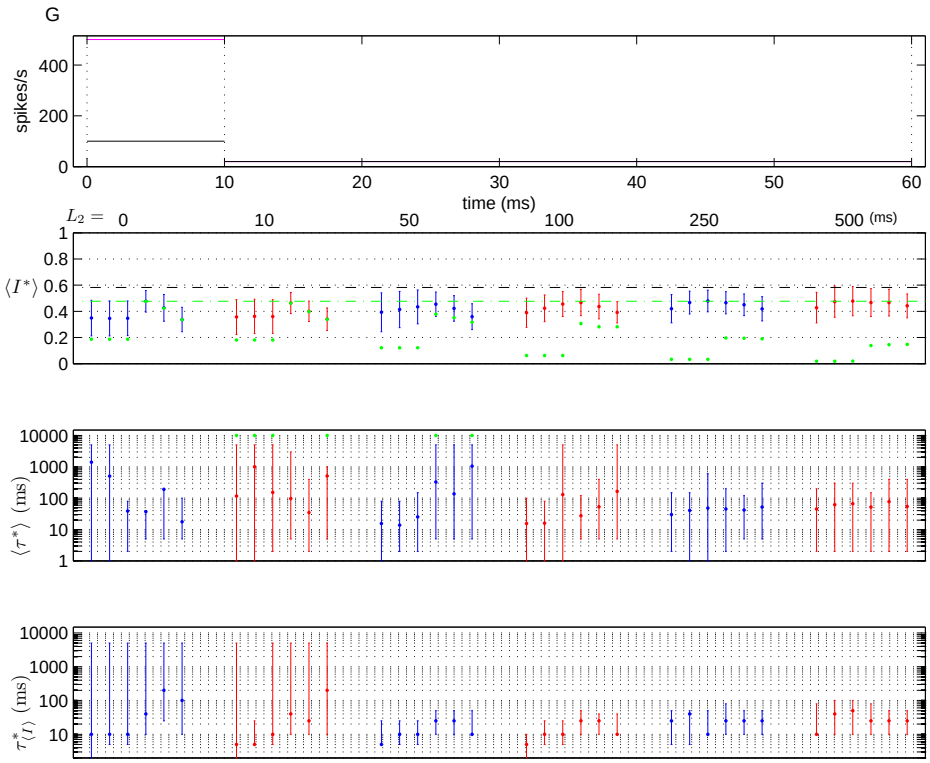

Figure 3.5: Dependence of $I^*_{max}(S^P, S)$ and τ^* on the length of the after-transient interval and on the classifier for phasic rate responses. Two intervals with constant rates for each stimulus are generated, as shown in the first row. The first interval $L_1 = 10$ ms models the phasic responses, while the second, $L_2 = [0, 10, 50, 100, 250, 500]$ ms, is the after-transient interval. The rates, following the notation of Figure 3.2, are $[500, 100; 20, 20]$ spikes/s. The structure of the figure is analogous to Figure 3.3. For each length L_2, results for exponents $z = [-8, -4, -2, 2, 4, 8]$ are shown with the same color. For the mutual information, we also show the maximal $I^*_{max}(S^P, S)$ across all the exponents for $L_2 = 0$ (dotted green line), and the estimate (Equation 3.1) of $I^*(R, S)$ for $L_2 = 0$ (dotted black line). Results are shown for D_{V}, $K = 20$, and $M = 20$.

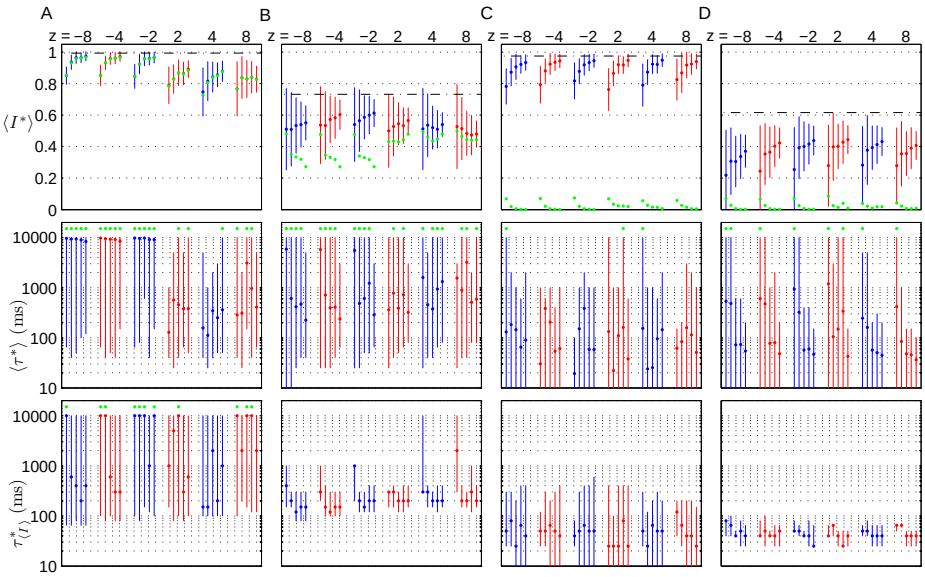

Figure 3.6: Dependence of $I^*_{max}(S^P, S)$ and τ^* on the number of trials M. Each column corresponds to the same examples as in Figure 3.2. The structure of the figure is analogous to Figure 3.3. For each exponent z, results for different numbers of trials $M = [5, 10, 20, 30, 50]$ are shown with the same color. We also indicate an estimate of $I^*(\mathbf{R}, S)$ (dotted black line), taking into account separately the rates in each of the four subintervals of constant rates. Results are shown for D_{V}. $K = 20$ independent realizations were used.

not achieved, the error bars become smaller for increasing M. In contrast to what is observed for $I^*_{max}(S^P, S)$, we see that $I^*_{count}(S^P, S)$ decreases with M except when $I^*_{max}(S^P, S) \simeq I^*_{count}(S^P, S)$. Therefore, the measure of the relevance of the temporal coding Θ is more sensitive than $I^*_{max}(S^P, S)$ to the number of trials. Regarding τ^*, increasing M does not reduce the variability across realizations or the flatness of the peak, and we do not see a clear convergence for increasing M. In contrast to $I^*_{max}(S^P, S)$, the error bars of $\langle \tau^* \rangle$ and $\tau^*_{\langle I \rangle}$ do not decrease with M.

3.3 Information and discriminative precision for experimental responses to transient constant stimuli

We now study the responses of exemplary rat gustatory cells and one monkey visual cell available online at neurodatabase.org (Gardner, 2004). Di Lorenzo and Victor (2003) analyzed the reliability of mean rate response across trials in the nucleus of the solitary tract (NTS) of the rat for four taste quality stimuli: NaCl, sucrose, quinine-HCl, and HCl. The authors applied discrimination analysis with the Victor distance to examine the potential contribution of temporal coding to the discrimination of the stimuli. The analysis was carried out for the total length of the recordings, 10 s, as well as for only the interval related to the phasic response, the first 2 s. We here analyzed three single neurons labeled 11, 9, and 4, to which we will refer as gustatory cells g_1, g_2, and g_3, respectively. For these cells 16, 23, and 19 trials for each stimulus were recorded respectively. In Figure 3.7 A-C we show the time-dependent rates averaged across trials for each stimuli. Notice the initial phasic response which lasts for the first two seconds approximately. Details of the moving window averaging are available in the caption of Figure 3.7.

Aronov et al. (2003) studied the discriminability of transient presentations of stationary gratings of spatial phases in single neurons from the primary visual cortex of anesthetized macaque monkeys. The authors applied the discrimination analysis with the Victor distance to analyze the encoding of spatial phases. Furthermore, Aronov et al. (2003) proposed an extension of the distance to compare the activity of a population of simultaneously recorded neurons depending on the stimuli. It was found that keeping track of which individual neuron fires and considering the temporal structure of the responses contributed to the discriminability of the responses. We here use one exemplary cell labeled $410106t$, to which we will refer as v_1. Data comprise the responses to 64 trials for four spatial phases (0, 90, 180, and 270 degrees). Each spike train corresponds to 236 ms of stimulus presentation plus 710 ms of a uniform field at the mean luminance. Accordingly, the length of the spike trains is ~ 1 s, which is one order of magnitude smaller than for the gustatory cells, as

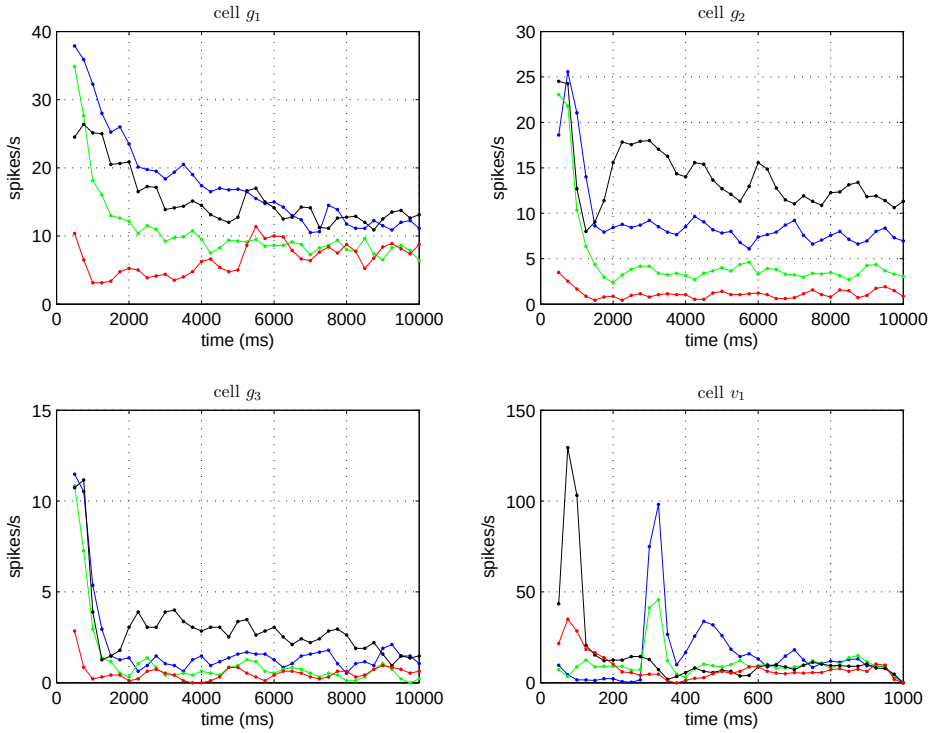

Figure 3.7: Time-dependent average rates for each stimulus of the four exemplary cells responding to transient constant stimuli. For the gustatory cells g_1-g_3 the rates were calculated for a moving window of 500 ms with an step of 250 ms. For the visual cell v_1 the moving window had 50 ms and a step of 25 ms. Different colors correspond to different stimuli.

can be appreciated in Figure 3.7 D. The local peaks in the rate reflect the onset and offset of the stimuli presentation.

3.3.1 Dependence on the measure and the classifier

We proceed in analogy to Section 3.2.2 examining $I^*(S^P, S; \tau)$ for L equal to the whole length of the spike trains, and for different exponents of the classifier and different measures. In Figure 3.8 we see that, for the gustatory cells g_{1-3}, the shape of $I^*(S^P, S; \tau)$ varies less for different exponents than for the visual cell v_1. For the different measures we again see that the normalization for D^*_{V} can lead to nonzero values for small time scales τ. The flatness and location of the peak depends on the measure.

In Figure 3.9 we show in more detail the dependence of $I^*_{max}(S^P, S)$ and τ^*. Like for the simulations using Poissonian spike trains (Section 3.2.2), $I^*_{max}(S^P, S)$ depends more on the classifier than on the measure. In particular, D_{B} can discriminate equally well as the measures without binning. The highest values of $I^*_{max}(S^P, S)$ are generally obtained for the classifiers using low $|z|$ and negative exponents are preferred. Oppositely, for $I^*_{count}(S^P, S)$, positive exponents lead to higher values. In consequence, Θ is more sensitive to the classifier. To minimize this effect $I^*_{max}(S^P, S)$ and $I^*_{count}(S^P, S)$ should be optimized separately across z before calculating Θ. In contrast to the simulated examples, we here cannot generate additional trials to get an unbiased estimation of $I^*(\mathbf{R}, S)$ that takes the temporal structure into account. Therefore, we reduce $\mathbf{R}$ to the total rates to give a reference of the information $I^*(R, S)$ that can be obtained considering only a rate code ($\tau \to \infty$). We use Equation 3.1 to calculate $I^*(R, S)$ and apply the analytical first order correction to the bias (Treves and Panzeri, 1995). Notice that $I^*_{max}(S^P, S)$ is not a lower bound for $I^*(R, S)$, as can be seen for those cells in which temporal coding is relevant. Nonetheless, $I^*(R, S)$ is an upper bound for $I^*_{count}(S^P, S)$ and thus it can be useful to quantify Θ when, like for cell g_1, $I^*_{count}(S^P, S)$ does not attain $I^*(R, S)$. Regarding τ^*, we see that the values of $\langle \tau^* \rangle$ and $\tau^*_{\langle I \rangle}$ for the exponent leading to the optimal $I^*_{max}(S^P, S)$ can differ in an order of magnitude across measures. In particular lower values are gen-

erally obtained for D_{R}, specially when the symmetric rectangular convolution kernel is used. Overall these results are consistent with the ones found for examples A-D of simulated data, indicating that the time scale related to the discriminative precision together with the assessment of the importance of temporal coding are more sensitive to the measure and the classifier than the mutual information.

3.3.2 Temporal accumulation of information

The dependence of the results on the length of the spike trains L used for the discrimination analysis was studied for the gustatory cells in Di Lorenzo and Victor (2003). In particular, the results obtained applying the discrimination analysis for the whole 10 s of the recordings were compared with those obtained for the first 2 s, associated with the phasic response. This comparison revealed that the temporal coding was more relevant during the phasic response. Similarly, Victor and Purpura (1998) found that the discriminability of spatial phases depends on L for V1 visual cells. Since both the gustatory cells and the visual cell respond to a transient presentation of constant stimuli, we expect that L plays an important role in the analysis as found for the simulations in Section 3.2.2.

In Figure 3.10 we show the dependence of $I^*_{max}(S^P, S)$ and τ^* on L. For all the cells the mutual information increases with L until it saturates or slightly decreases again. We compare $I^*_{max}(S^P, S)$ obtained for $z = -2, 2$ with $I^*_{count}(S^P, S)$. We observe a different dependence for each cell. For cell g_2 there is almost no difference among the classifiers, but the difference between $I^*_{max}(S^P, S)$ and $I^*_{count}(S^P, S)$ varies with L. For $L = 500$ ms, the information obtained by rate coding is almost maximal. When L increases, Θ becomes substantial but it decreases for $L > 2$ s, leading again to a rate code for $L > 8$ s. The first 2 s are almost as informative as the total. For $L = 2$ s, $I^*_{max}(S^P, S) > 0.8$ is already close to the value obtained for $L = 10$ s. However, $\langle \tau^* \rangle$ and $\tau^*_{\langle I \rangle}$ increases with L. This resembles example F of Figure 3.4 for which information is contained on two different time scales, related to the latency code and the rate code. In such cases a single quantity τ^* is not enough to reflect the

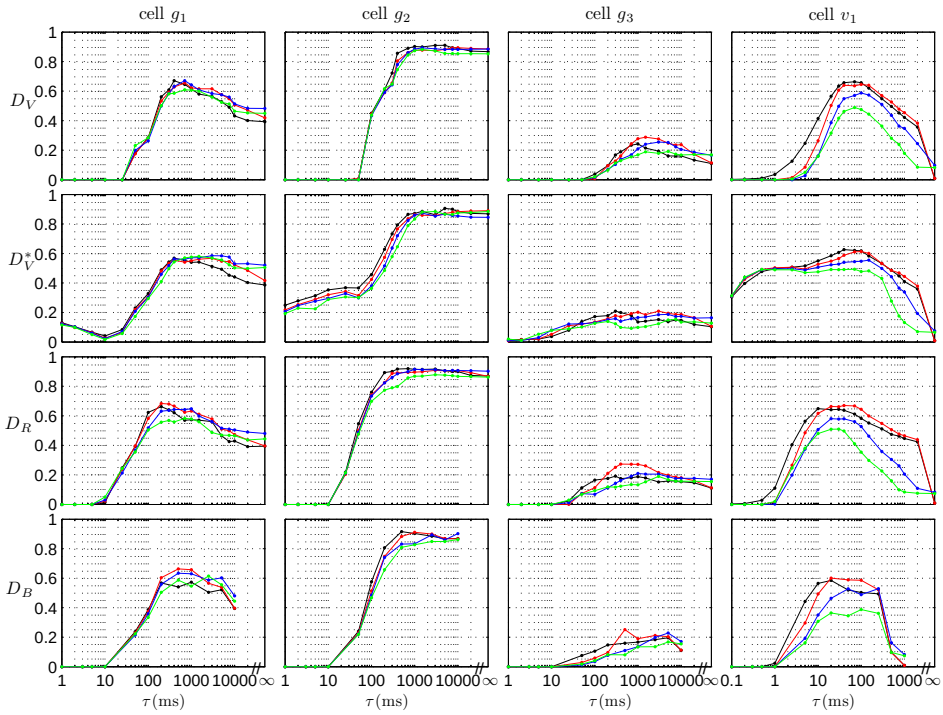

Figure 3.8: Dependence of $I^*(S^P, S; \tau)$ on the measure and the classifier for the four exemplary cells. The figure is analogous to Figure 3.2. Four classifiers with different exponent (Equation 3.2) are used: $z = -8$ (black), $z = -2$ (red), $z = 2$ (blue), and $z = 8$ (green). Results shown are averages across $K = 20$ bootstrapping realizations in which $M' = \frac{3}{4}M$ randomly chosen trials are used.

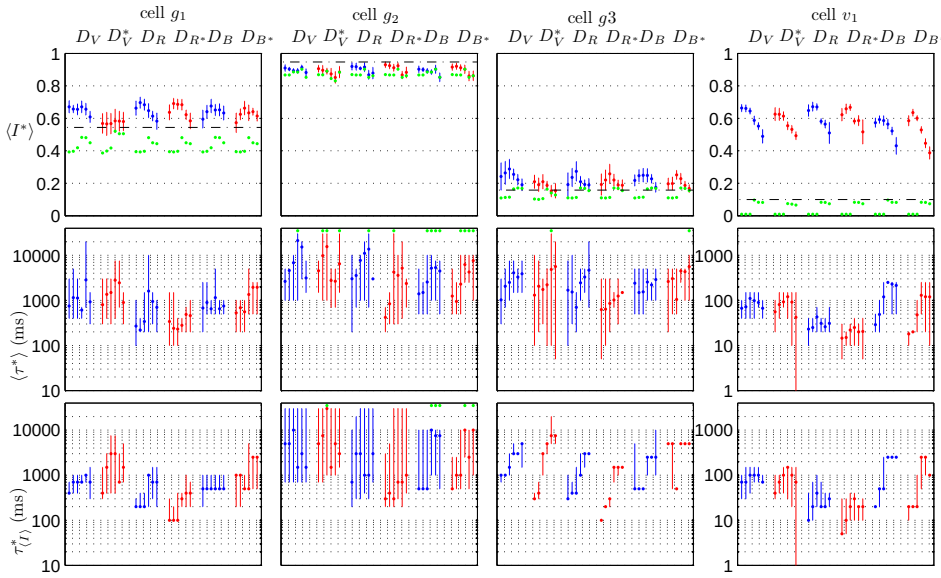

Figure 3.9: Maximal mutual information $I^*_{max}(S^P, S)$ and optimal time scale τ^* in dependence on the measure and the classifier for the exemplary cells. The figure is analogous to Figure 3.3. For each measure the results are shown for exponents $z = [-8, -4, -2, 2, 4, 8]$, from left to right, with the same color. We indicate an estimate of the mutual information obtained without considering temporal coding, $I^*(R, S)$ (dotted black line). Details on the calculation on the averages and the error bars are provided in Section 3.1.2.

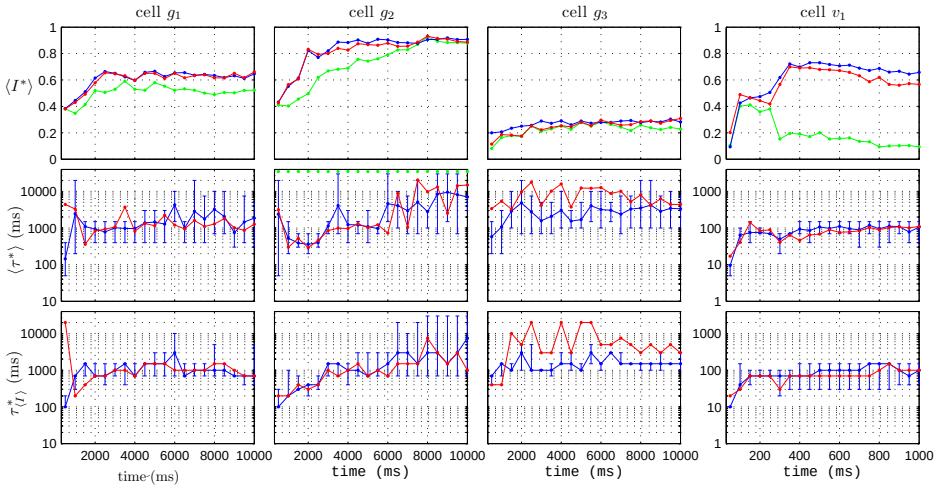

Figure 3.10: Temporal accumulation of information. Dependence of $I^*_{max}(S^P, S)$ and τ^* on the length L of the recordings used in the discrimination analysis. The maximal mutual information $I^*_{max}(S^P, S)$ is shown for $z = -2$ (blue), and for $z = 2$ (red). We also show the mutual information obtained by a rate code $I^*_{count}(S^P, S)$ (green). For the optimal time scales $\langle \tau^* \rangle$ and $\tau^*_{\langle I \rangle}$, results are shown with error bars for $z = -2$ and without for $z = 2$. The length L is increased in steps of 500 ms for the gustatory cells g_1-g_3 and 50 ms for the visual cell v_1. Results are shown for D_{V}. $K = 20$ bootstrapping realizations with $M' = \frac{3}{4}M$ randomly chosen trials were used.

time scale at which information is contained, and intermediate values of τ^* are obtained when none of the sources of information predominates. For example, for $L = 10$ s, the information in the first two seconds can still be obtained with a lower τ, however, using the lower τ would result in contradictory information for the posterior part of the spike trains. This indicates that τ^* depends in a nontrivial way on the balance between how the sources of information are distributed across the total length of the spike train and at which time scale.

We also see in Figure 3.10 that, for cells g_1 and g_3, τ^* is less dependent on L. This suggest that there is no interplay of different time scales. Furthermore, since the values of $I^*_{count}(S^P, S)$ are close to $I^*_{max}(S^P, S)$, the results are similar to the ones shown in Figure 3.5. For the visual cell v_1, the pronounced ON and OFF responses are clearly reflected in $I^*_{max}(S^P, S)$. The information increases abruptly with the ON response and then again with the OFF response. By contrast, $I^*_{count}(S^P, S)$ increases with the ON response but goes down with the OFF response, indicating that the total rates are not enough to distinguish the stimuli without resolving the time of the rate increases. After the OFF response $I^*_{max}(S^P, S)$ decays slightly, in particular for the positive exponents. This indicates that only contradictory information is added. Given the dependence of $I^*_{max}(S^P, S)$ and $I^*_{count}(S^P, S)$ on L, Θ is very sensitive to L. Before the OFF response a rate code accounts for almost all the information, while after it Θ is high. However, the optimal time scale τ^* increases for $L < 150$ ms but then remains constant at a time scale which is neither related to the presentation period nor to the shorter time scale corresponding to the ON and OFF responses.

3.3.3 Distribution and redundancy of the information along the spike trains

Given the above results, it is clear that information is not distributed uniformly along the duration of the recordings. Furthermore, that the accumulation of information saturates faster or at a higher level does not only depend on the temporal distribution of the information but also on

how redundant it is across different intervals. To complement the study of the dependence of $I^*_{max}(S^P, S)$ and τ^* on L, we here examine these quantities for segments of different length and location. Due to the extensiveness of this analysis we restrict ourselves to cells g_2 and v_1, but the conclusions are common to all cells. We consider all nonoverlapping segments with length percentages 100, 80, 40, 20, 10, and 5 of the total length of the recordings. In Figures 3.11 and 3.12 we show the shape of $I^*(S^P, S; \tau)$ for classifiers with $z = -2, 2$. We also provide an estimate of $I^*(R, S)$, considering the total rate for each segment length, which serves as a reference of the total amount of information obtained without temporal coding.

For cell g_2, we see that all the 1 s segments contain similar amounts of information. However, the shape of $I^*(S^P, S; \tau)$ depends on the segment and the classifier. For some segments the same $I^*_{max}(S^P, S)$ is obtained with both classifiers but they produce different shapes of $I^*(S^P, S; \tau)$, and different τ^* and Θ. For the first second, both classifiers indicate some degree of temporal coding, while for the rest the negative exponent gives higher Θ. In particular, for various of the 1 s segments, after the phasic response elicited in the first 2 s, the positive exponent lead to higher $I^*_{max}(S^P, S)$ that nearly match the estimated $I^*(R, S)$. This dependence on the exponent is even higher for $I^*_{count}(S^P, S)$. We found that positive exponents are preferable for the posterior part of the response for all the cells. When examining the segments of length 2 s, the first segment is clearly different from the rest, with temporal coding both indicated by the peak at low τ and the values of $I^*_{max}(S^P, S)$ higher than $I^*(R, S)$. Notice that for the rest of the segments the shapes for the two exponents are almost overlapping, although they differed substantially for the corresponding segments of 1 s. Furthermore in these segments $I^*_{max}(S^P, S)$ matches $I^*(R, S)$. The transition from temporal to rate code is already visible for the first segment of 4 s where the peak of $I^*_{max}(S^P, S)$ is closer to $I^*(R, S)$, and this trend is consolidated for the 8 s and 10 s segments.

For cell v_1 the temporal distribution of information is completely different. Regarding the segments of length 100 ms, the last five segments do no contain information at all. These are the segments posterior to the

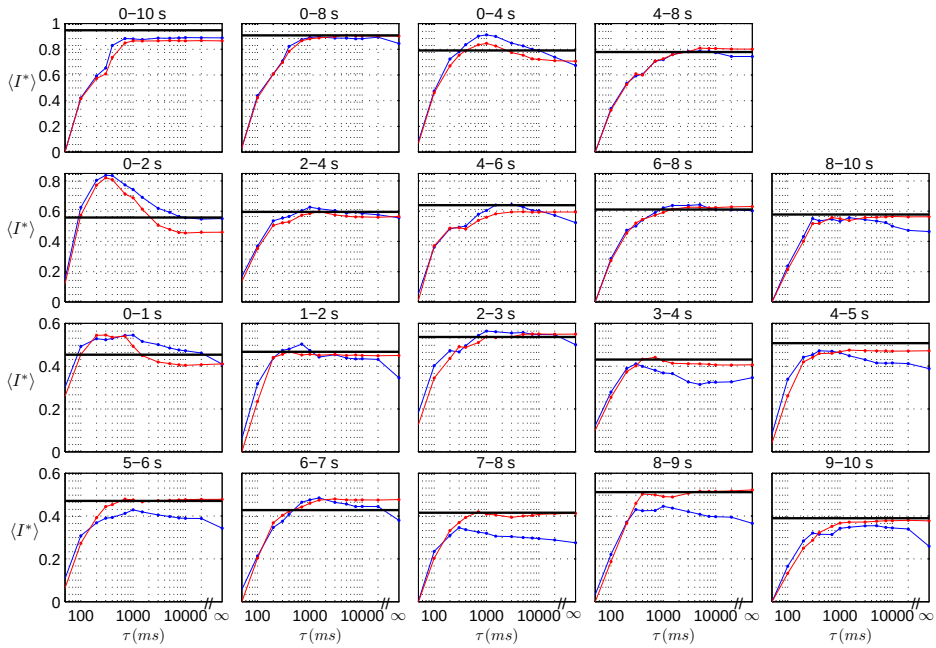

Figure 3.11: Temporal distribution of information along the recordings for cell g_2. We show $I^*(S^P, S; \tau)$ for nonoverlapping segments with different location and different length L. Row 1: from left to right, segments of 10 s, 0.8 s, and 0.4 s. Row 2: segments of 0.2 s. Row 3 and 4: segments of 0.1 s. We indicate the initial and final time for each segment. We show the results for $z = -2$ (blue), and $z = 2$ (red). We also provide an estimate $I^*(R, S)$ (black). Results are shown for D_{V}. $K = 20$ bootstrapping realizations with $M' = \frac{3}{4}M$ randomly chosen trials were used.

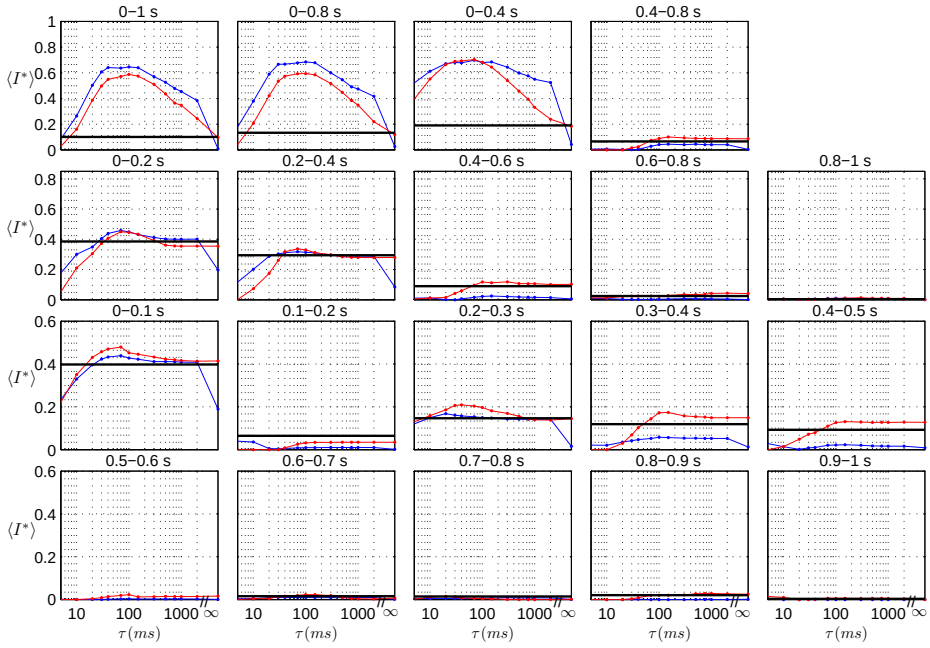

Figure 3.12: Analogous to Figure 3.11 but for cell v_1.

OFF response. For the first five segments, the first segment contains the highest amount of information, the second the lowest, and the other three a similar amount. Notice that the segmentation has not been done taking into account where exactly the ON and OFF responses occur, which means that taking segments located at these specific times could show that information is higher and more concentrated. However, for our purpose, it suffices to show how information arises nonuniformly. For the first, third, and, to a lower degree, the fourth segment, temporal coding retrieves more information. Analogously to cell g_2, for segments four and five the positive exponent substantially increases the information. The concentration of information during the presentation period is confirmed when examining the segments of 200 ms. For this length $I^*_{max}(S^P, S)$ is only slightly higher than $I^*(R, S)$ for the first two segments. By contrast, when these two segments are merged in the first segment of 400 ms, the rate code discriminates clearly worse than if temporal coding is used. The first segment of 400 ms contains more information than the total of 1 s, in agreement with the decay of the accumulated information observed in Figure 3.10. For 800 ms and 1 s the negative exponent retrieves more information, oppositely to what occurs for the shortest segments.

To better appreciate how the local information accumulates, we compare the shape of $I^*(S^P, S; \tau)$ for a given segment to the shape of the two subsegments of half length. In Figures 3.13 and 3.14 we furthermore indicate for each case the redundancy index (Reich et al., 2001a):

$$RI = \frac{I^*_{max,1} + I^*_{max,2} - I^*_{max,joint}}{I^*_{max,1} + I^*_{max,2} - \max(I^*_{max,1}, I^*_{max,2})}, \tag{3.11}$$

which is 0 if each half segment contributes independent information and 1 if they are completely redundant. RI is higher than 1 for contradictory information and negative for synergistic information. However, since the denominator $I^*_{max,1} + I^*_{max,2} - \max(I^*_{max,1}, I^*_{max,2}) = \min(I^*_{max,1}, I^*_{max,2})$, this index is specially sensitive when low information is found in one of the half segments. Here we only use it to qualitatively indicate the accumulation of information.

In Figure 3.13, for cell g_2, RI is about 0.6 or higher for the segments

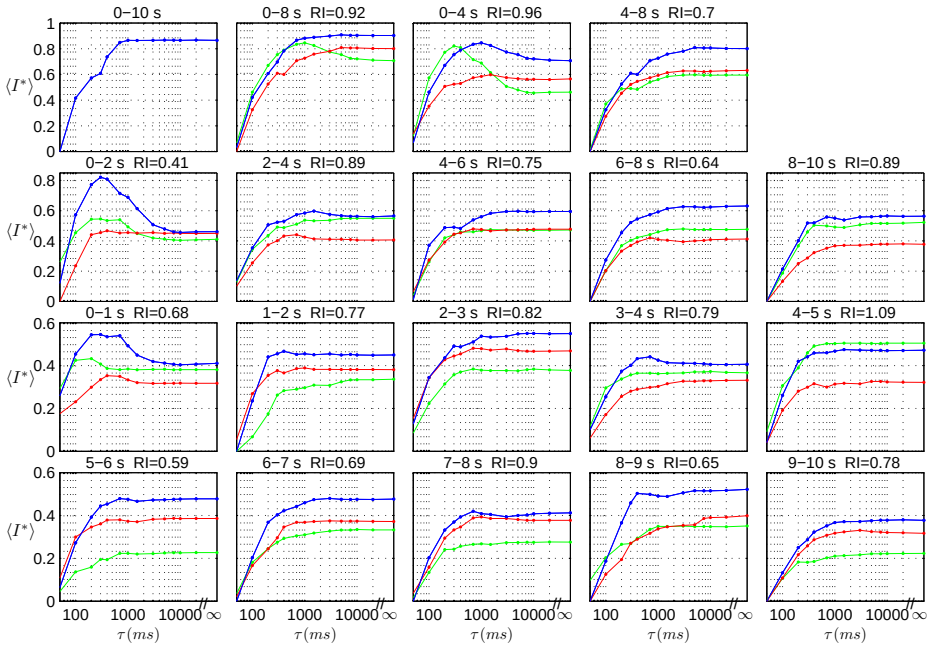

Figure 3.13: Redundancy of information along the recordings for cell g_2. $I^*(S^P, S; \tau)$ for nonoverlapping segments with different location and different length L. The structure of the figure is analogous to Figure 3.11. However, apart from the mutual information $I^*(S^P, S; \tau)$ for each segment (blue), also the mutual information for the two nonoverlapping subsegments of half length are shown, the first in green, the second in red. We provide for each segment the redundancy index (Equation 3.11) quantifying the redundancy between the two subsegments. Results are shown for D_{V} and $z = 2$. $K = 20$ bootstrapping realizations with $M' = \frac{3}{4}M$ randomly chosen trials were used.

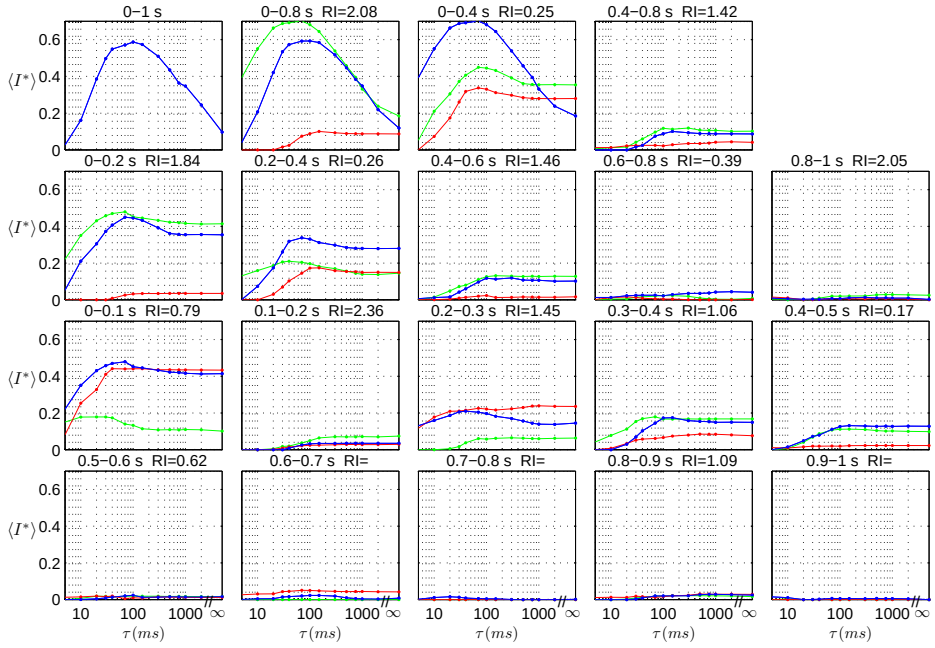

Figure 3.14: Analogous to Figure 3.13 but for cell v_1.

of length 1 s, indicating that there is a substantial redundancy between the information. Indeed, for the subsegments of length 500 ms, that we did not show in Figure 3.11, also a significant degree of information is nearly uniformly distributed across the whole length of the spike trains. For the first second, which shows a clear peak indicating temporal coding, the two subsegments are more consistent with a rate code. For the first 2 s, the lowest value of RI over all time scales is obtained. By contrast, for the first 4 s information is almost completely redundant. The same occurs for the segment of 8 s. This is consistent with Figure 3.10 concerning the saturation of the accumulation of information. The phasic response contains almost all the nonredundant information in the first 2 s. Nonetheless, this information can also be retrieved from the posterior response where it is distributed more uniformly and in a less redundant way.

For cell v_1, we see in Figure 3.14 that RI is much more variable and we find negative values and values above 1. For three segments of length 100 ms RI is not well defined because the information is zero in one of the subsegments. RI shows a high variability when low information values are obtained. We see that almost all the information is contained in the segments 50-100 ms, 250-300 ms and 300-350 ms. The first interval covers the ON response, and the other two cover the OFF response. This is reflected in low RI values for the second segment of 200 ms and the first segment of 400 ms. The part of the spike trains after the OFF response provides only contradictory information as evidenced by the RI of the 800 ms segment.

3.4 Information and discriminative precision for experimental responses to time-dependent stimuli

In the previous analysis, comprising simulations (Section 3.2) and experimental data (Section 3.3), we restricted ourselves to transient constant stimuli. We now turn to time-dependent stimuli to consider the temporal structure in the spike trains caused by the correlation with the time-dependent stimuli. For time-dependent stimuli, one has to distinguish

between the elements S_j, $j = 1, ..., N$, of the discrete set of stimuli $\{S\}$, considered in Equation 3.1, and what is the actual stimuli determining the response of each particular cell at each concrete time. Each S_j consists of a sequence $\mathbf{S}_j(t)$ of stimuli $\mathbf{S}_j(t_i)$. A particular stimulus $\mathbf{S}(t_i)$ can be repeated at different times within or across the stimuli $\{S\}$, following some particular distribution. For neurons with different receptive fields, the same time-dependent stimulus $\mathbf{S}_j(t)$ constitutes a different time-dependent profile $\mathbf{S}_{jk}(t)$ to which the neuron is sensitive, where k refers to the particular neuron responding to S_j. This is because each neuron is sensitive to a particular subspace of the stimulus. The time-dependent profile is a vector which can have components corresponding not only to the instantaneous time t_i but to the past up to $t_i - \Delta t$, depending on the temporal extent of the receptive field. Accordingly, the modulation of the firing rate reflects the interplay of a sequence $\mathbf{S}(t)$ with a particular receptive field and a particular tuning curve. In this context, the discrete set of stimuli $\{S\}$, even if meaningful, for example, considering the perceptual relevance of the sequences chosen, provides simply the categories needed to calculate the mutual information with the classifier (Equation 3.3). Accordingly, the mutual information between the stimuli and predicted stimuli is not a lower bound of the information contained in the responses about the stimuli, but simply a measure of discrimination performance.

We analyze single-unit recordings from the songbird auditory system (Brainard and Doupe, 2000; Theunissen and Shaevitz, 2006). Details of the recordings are described in Appendix B and have been facilitated by Emily Caporello, from the Laboratory of Auditory Neuroethology and Communication of Tim Gentner, at the University of California, San Diego. The data comprise recordings from neurons from Field L and Caudal Mesopallium (CM) of European Starlings. Field L is the avian analog to primary auditory cortex in mammals (Fortune and Margoliash, 1992), and contains neurons that respond to complex temporal, spectral and spectrotemporal acoustical features. Some neurons encode low-level acoustics like power in a given band, while others are sensitive to specific vocalizations. CM can be compared to the superficial layers

of auditory cortex or to secondary mammalian auditory regions, such as the lateral belt in primates or Wernicke's area in humans (Jarvis, 2004). Some neurons in this area are sensitive to higher order acoustic features of vocalizations. Recordings are from 4 European Starlings summing a total of 22 well-isolated single neurons. From these cells, 13 are in CM and 9 in Field L. In the results we do not distinguish between different areas because we found no systematic differences. The stimuli consist of $N = 9$ unfamiliar conspecific songs of 14 s of duration. $M = 6$ trials are available for each song.

In Figures 3.15, 3.16 we present the response from some exemplary cells to show the variability of the responses across cells. The average firing rate is highly variable, however, events corresponding to a reliable response for repeated presentations of the same song are appreciable in all the cells to some degree. This behavior is consistent with the diversity of receptive fields. While cells encoding low-level acoustic features are sensitive to continuous temporal variations of these features, cells encoding higher-order features, or with a more selective tuning curve, fire sparsely.

3.4.1 Dependence of the information and the discriminative precision on the firing rate

We examine the relation between $I^*_{max}(S^P, S)$ and τ^*, and their dependence on the average firing rate. As in Section 3.3, when analyzing the responses to transient constant stimuli, we use bootstrapping to obtain average values. In Figure 3.17 we show $\langle I^* \rangle$ and $\langle \tau^* \rangle$ obtained by repeating the discrimination analysis for all the combinations of $N' = 5$ songs. We use the Spearman rank correlation coefficient to quantify the dependencies. The mutual information and the optimal time scale are negatively correlated. The coefficient is always significant and has an average value of $\langle \rho \rangle = 0.61$, when averaged for groups of $N' = 3 - 7$ songs. Furthermore, $\langle I^* \rangle$ is highly positively correlated with $\bar{n}$, the average number of spikes per song. The coefficient has an average value of $\langle \rho \rangle = 0.71$. Oppositely, $\langle \tau^* \rangle$ is highly negatively correlated to $\bar{n}$ with an average co-

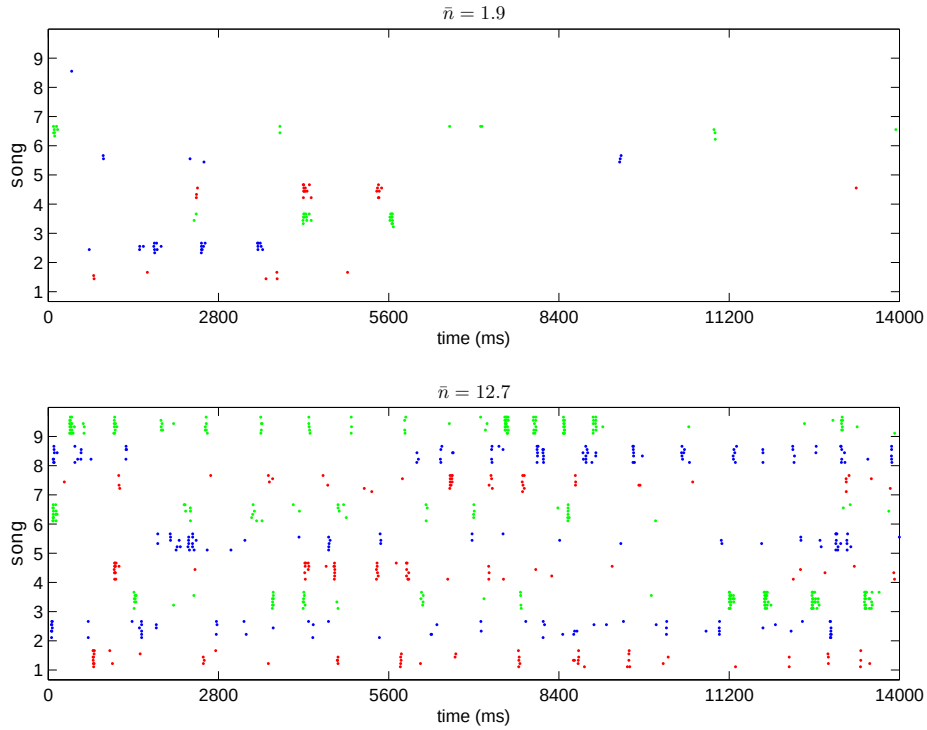

Figure 3.15: Responses of two exemplary auditory cells elicited by $N = 9$ conspecific songs. The spike trains corresponding to $M = 6$ trials elicited by the same song are shown together, and colors are alternated to better differentiate the responses to different songs. The average number of spikes per song is indicated by $\bar{n}$.

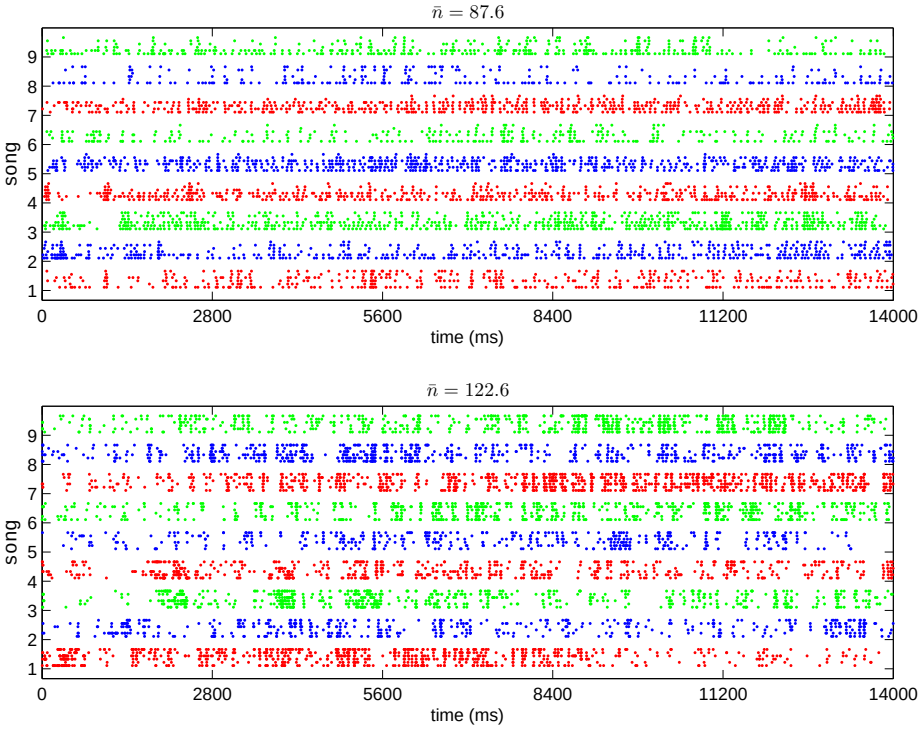

Figure 3.16: Analogous to Figure 3.15 but for two cells with higher average number of spikes per song $\bar{n}$.

efficient $\langle\rho\rangle = -0.78$. A similar dependence was found for $\tau^*_{\langle I\rangle}$.

These results can be explained assuming the correlation between the rate and the profiles $\mathbf{S}_{jk}(t)$ determined by the sensitivity of the k-th cell. The cells with higher rates reflect a lower-level acoustic feature, which, given that the songs are independent, evolves differently at each point in time for each song, therefore increasing the evidence about the identity of the song. For more selective cells, the evidence about the identity of the song is only available occasionally. Therefore, in general, higher rates improve the discrimination, despite it depends on the reliability of the responses that the increasing evidence results in an increase of $I^*_{max}(S^P, S)$.

Regarding τ^*, its decrease for higher average rates reflects that the modulations caused by the stimuli are local and do not result in significant difference in the total number of spikes elicited by each song. For cells with low rates, the variability of τ^* increases. For these cells the discriminative precision is specially sensitive to the presence of single events corresponding to a reliable pattern. Since the features that produce the reliable patterns are different across cells and have a different distribution across songs, τ^* has a high variability in this range. Of course, for longer recordings this dependence on the distribution of the relevant features would be weaker. This arbitrariness of τ^* is an effect of the arbitrariness of the selection of a discrete set of stimuli S.

The opposite correlations of the mutual information and the optimal time scale with the average number of spikes per song is consistent with their own negative correlation. However, the fact that this correlation is lower reflects that it results indirectly from the dependence on the average number of spikes per song.

3.4.2 The time scale characteristic of reliable patterns

We here use the discrimination analysis to examine an alternative quantity τ_p, related to the time scale of the reliable patterns in the response. As discussed above, the discrimination performance is influenced by the interplay of the properties of the cell with the contingent distribution of

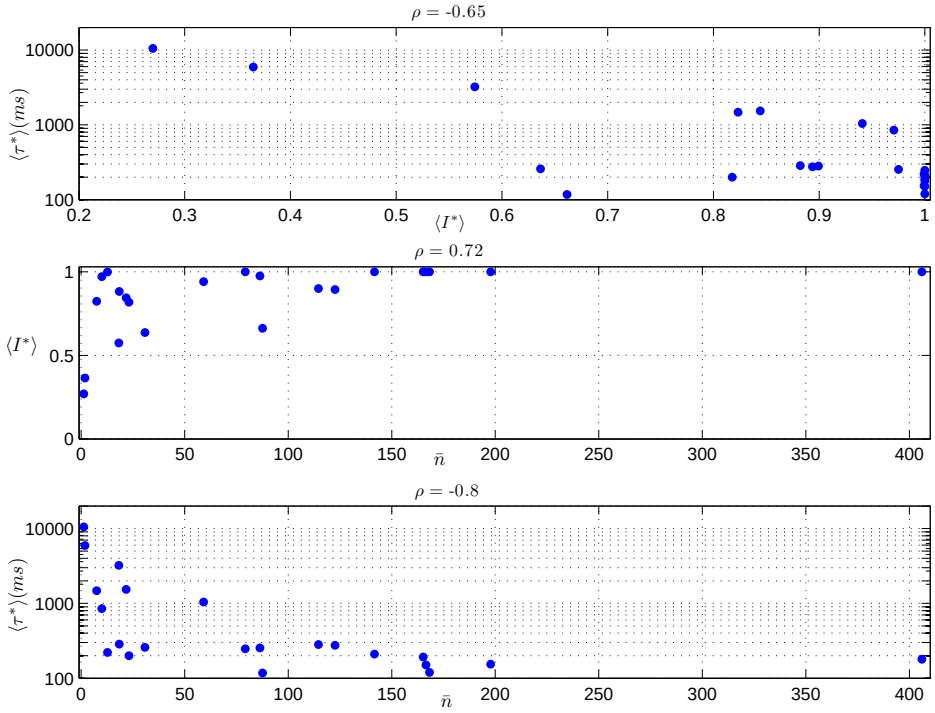

Figure 3.17: Mutual information and optimal time scale in dependence on the average number of spikes per song. Each dot corresponds to the average value obtained for an individual cell when the discrimination analysis is repeated for all combinations of $M' = 5$ trials per song. The Victor distance is used in the discrimination analysis. $I^*_{max}(S^P, S)$ is maximized across different classifiers, with the exponent $z = -4, -2, 2, 4$. ρ indicates the Spearman rank correlation coefficient.

the relevant features across time and stimuli of $\{S\}$. For the k-th cell, the stimulus S_j can cause a reliable pattern in the response at a given time t_i elicited by the feature $\mathbf{S}_{jk}(t_i)$. Therefore, a particular distribution across time and across stimuli of reliable patterns reflects the arbitrary combination of stimuli $\mathbf{S}_j(t)$, $j = 1, ..., N$, as well as the sensitivity and reliability of the cell. A way to partially disentangle the influence of the particular time-dependencies and of the existence of reliable responses is to generate surrogate data that preserve the reliable patterns but randomize their distribution across time. This is done by cutting segments of the spike trains and randomly reordering them. In particular, for a fixed segment size T_s, the segments are shuffled, but keeping together the same segment for all the trials of the same stimulus. Accordingly, any correspondence across songs, which was already arbitrary for the original spike trains, is removed. Furthermore, any reliable pattern with a time scale higher then T_s is destroyed. However, the major part of reliable patterns up to the time scale of the segments size are preserved. Unavoidably, part of the structure at lower time scales is also destroyed when a pattern is divided in two adjacent segments, and the probability that this happens increases for smaller T_s. Neglecting these discontinuities, and assuming that the spectrotemporal features to which the cell is sensitive do not have a temporal extent higher than T_s, the new spike trains can be thought as responses to new songs with an alternative distribution of the same features. At the population level, the mutual information $I^*_{max}(S^P, S)$ should be the same for the surrogate spike train as long as the reliable patterns are not destroyed.

In Figure 3.18 we show the decrease of $I^*_{max}(S^P, S)$ for all the cells in dependence on the segment size T_s. Furthermore, we illustrate how we define the time scale characteristic of the reliable patterns τ_p. Given $K = 19$ surrogate realizations, we look for which value T_s the average mutual information $\langle I^* \rangle$ extracted from the original spike trains is higher than all the values obtained for the surrogates. This indicates that the original spike trains are no longer consistent with the surrogates with a significance level of $\alpha = 0.05$. We take τ_p to be equal to this value T_s.

Given τ_p, we examine if it is correlated with the mutual information,

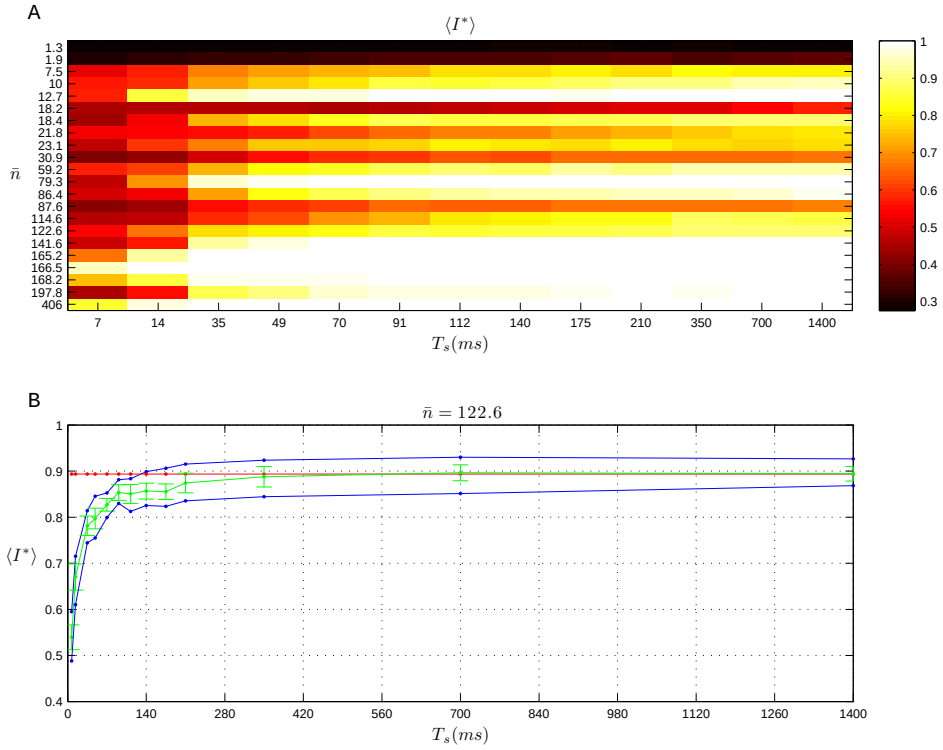

Figure 3.18: The calculation of the characteristic time scale τ_p of the reliable patterns. A. Dependence of the average of $\langle I^* \rangle$ across surrogates on the size T_s of the shuffled segments for all the cells ordered by the average number of spikes per song. B. Exemplary cell for which we show $\langle I^* \rangle$ for the original responses (red) compared with the values obtained for the surrogate data, the average and standard deviation (green), as well as the minimum and maximum across $K = 19$ surrogates (blue). The Victor distance is used in the discrimination analysis. $I^*_{max}(S^P, S)$ is maximized across different classifiers, with the exponent $z = -4, -2, 2, 4$.

the average firing rate or the discrimination precision. We see in Figure 3.19 that none of these correlations is significant. In Figures 3.15, 3.16 we already showed that reliable patterns are observable for all the cells independently of their responsiveness. Comparing τ_p with τ^*, we see that the time scale of the reliable patterns is lower than the time scale of discrimination. While the optimal time scales τ^* takes values in the range $(10 - 10000)$ ms, τ_p is in the range $(5 - 180)$ ms. This means that the structure of the responses across trials of the same stimuli S_j contain sufficient information to achieve the same level of discrimination at lower time scales than the time scale associated to discrimination precision. One may think that this difference in the time scales is due to the way $\langle\tau^*\rangle$ is defined (Section 3.1.2), as the average of all τ for which $I^*_{max}(S^P, S)$ is obtained. However, the same results were obtained (results not shown) when we used an alternative definition, considering the minimum τ instead of the average. For this case, although lower for some cells, $\langle\tau^*\rangle$ was still on average orders of magnitude higher than τ_p across cells.

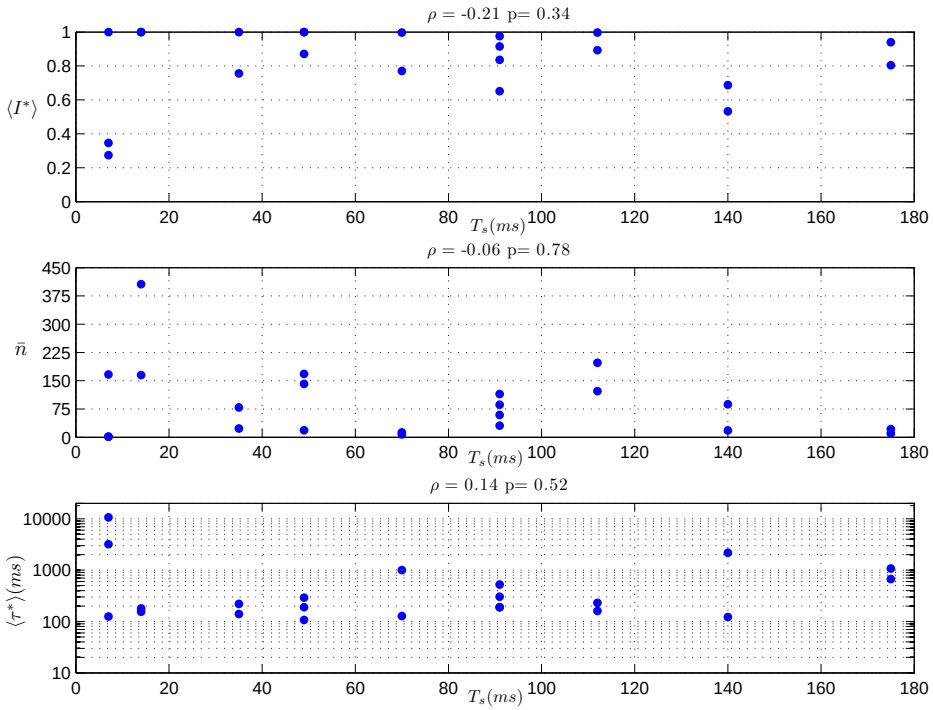

Figure 3.19: Correlation of the average mutual information $\langle I^* \rangle$, the optimal time scale of discrimination $\langle \tau^* \rangle$, and the average number of spikes per song $\bar{n}$ with the characteristic time scale of the reliable patterns τ_p. Each dot corresponds to an individual cell. We indicate the Spearman rank coefficient and the corresponding p value.

3.5 The interpretation of discrimination analysis

In this part of the thesis we dealt with the use of time scale parametric spike train distances to study the information and discriminative precision of neural responses. For this purpose, we studied several measures commonly used to quantify the dissimilarity between spike trains. We started by comparing these measures using time-independent Poisson spike trains to characterize the dependence on the time scale parameter τ (Section 3.2.1). This characterization is necessary to understand the dependence of the mutual information between the stimuli and the predicted stimuli $I^*(S^P, S; \tau)$ (Equation 3.4) on the time scale parameter. We then evaluated the influence of different factors in the discrimination analysis. In particular, we considered the dependence of the mutual information and the discriminative precision on the measure and the selected classifier, as well as on the length of the spike trains used for the discrimination analysis, and the number of trials available for each stimulus.

We showed that the mutual information $I^*_{max}(S^P, S)$, obtained from the maximization across the time scales, is more sensitive to the classifier than to the measure. This result was consistent for simulated time-dependent Poisson spike trains (Section 3.2.2) and for exemplary experimental data comprising recordings from gustatory and visual neurons (Section 3.3.1). In particular, a distance based on binning led to similar results as the more elaborated spike train distances. This suggests that the preeminent ingredient of the discrimination analysis is the use of a classifier to avoid the estimation of the joint probability distribution of stimuli and responses $p(\mathbf{R}, S)$ (Equation 3.1). Once the high-dimensional probabilities do not have to be estimated, binning is not a problem by itself. Regarding the discriminative precision, we observed that the time scale τ^* for which the mutual information is maximal results from a nontrivial interplay of the different sources of temporal structure as well as from the different time scales at which the information is contained. We showed that for simulated time-dependent Poisson spike trains, τ^* does not depend directly on the time scale of the modulation in the rate. It is determined by several factors including the local and global balance between

the rates in response to each stimulus. Furthermore, τ^* and the percentage of information increase Θ (Equation 3.6), which reflect the relevance of temporal coding, are more sensitive to the measure and classifier than the mutual information itself.

3.5.1 The dependence on the classifier and the measure

To our knowledge, the dependence of the mutual information on the classifier, determined by the exponent z in Equation 3.2, has not been addressed in detail before. Victor and Purpura (1997) showed that, for simple Gamma processes, the shape of $I(S^P, S; \tau)$ changes with z. This affects the determination of the optimal time scale τ^* as well as the significance of temporal coding quantified by Θ. However, the particular exponent used is usually not specified in experimental studies (Victor and Purpura, 1998; Machens et al., 2001; Aronov et al., 2003; Di Lorenzo and Victor, 2003; Samonds and Bonds, 2004; Di Lorenzo and Victor, 2007; Roussin et al., 2008; Huetz et al., 2009). When explicitly stated (Mechler et al., 1998; Macleod et al., 1998; Huetz et al., 2006), it is usually not motivated. To our knowledge, only Victor and Purpura (1996), presented the results for $z = -2$ and mention that similar results were obtained for $z = -8, ... - 1$, and Reich et al. (2001b), provided an analysis of the robustness of the median distance for Equation 3.2 when few spikes are present in each spike train. In other studies, to avoid this parametric dependence on the exponent, the discrimination performance was quantified by the percentage of correct classifications obtained as described in Section 3.1.2. Machens et al. (2003) indicated that the use of different supervised and unsupervised classifiers did not affect the results. However, other studies using the percentage of correct classifications to evaluate the discrimination do not report any comparison with other classifiers (Grewe et al., 2003; Narayan et al., 2006; Wohlgemuth and Ronacher, 2007; Wang et al., 2007; Billimoria et al., 2008). Although avoiding the selection of z, the optimality of the classifier based on the random selection of templates may differ for different time scales, cells, or conditions. In general, we showed (Section 3.2.2) that $I^*(S^P, S; \tau)$ should be maximized not only

across time scales τ, but across different classifiers, for example for different values of z. For the comparison of $I^*_{max}(S^P, S)$ and $I^*_{count}(S^P, S)$, both should be maximized separately.

It is also common practice to apply the discrimination analysis for only one particular measure in application to experimental data (e. g. Huetz et al., 2006; Roussin et al., 2008). In other cases various measures are used to verify if the results are independent of the measure (Machens et al., 2003; Narayan et al., 2006), or to compare their performance (Wang et al., 2007). In particular, the two different families of spike train distances based on spikes and intervals (Victor and Purpura, 1996), are usually considered together to examine the nature of the code (e. g. Victor and Purpura, 1996; Di Lorenzo and Victor, 2003; Samonds and Bonds, 2004). We here have not studied the interval-based spike train distances of Victor and Purpura (1997). This is because we were more interested in showing the effects of the particularities of each measure even in the case where they all quantify the similarity of spike times.

3.5.2 The dependence on the length of the spike trains for different codes

In Section 3.2.2, we showed that for pure latency coding of the stimuli, $I^*_{max}(S^P, S)$ is almost independent of the length of the after-transient interval, and τ^* reflects the time scale of the latency difference. However this direct interpretation does not hold in general. When information is extracted at different time scales, the discriminative precision arises from the particular balance between how well for a particular τ the different sources of information can be extracted. This has been confirmed by the examination of the accumulation and distribution of information along the spike trains for the experimental recordings in Sections 3.3.2 and 3.3.3. Choosing a particular L implies some assumptions about which portion of the responses is informative, but also determines which is the best τ to retrieve this information. Different classifiers can be better for different L, leading to counterintuitive results like in Figure 3.5, unless results across several classifiers are scrutinized. This led us to also consider a di-

rect calculation of $I^*(\mathbf{R}, S)$, the normalized mutual information between the stimuli and the responses, calculated directly from Equation 3.1. The calculation is straightforward for simulations for which a large number of data can be generated to avoid estimation biases. Since the mutual information obtained from the classifier $I^*_{max}(S^P, S)$ is a lower bound of $I^*(\mathbf{R}, S)$, this helps to evaluate the optimality of the classifiers. For an optimal classifier, the difference between $I^*(\mathbf{R}, S)$ and $I^*_{max}(S^P, S)$ quantifies the extra information in the responses not related to decoding, but exclusively to the representation and uncertainty of the classification (Quian Quiroga and Panzeri, 2009).

3.5.3 The dependence on the number of trials

In Section 3.2.2, we studied the convergence of the estimated $I^*_{max}(S^P, S)$ and τ^* with the number of trials M of each stimulus. We found that the mutual information generally increases with the number of trials, since the spike trains sample the distribution of responses for each stimulus better. By contrast, the optimal time scale τ^*, had a high variability independently of the number of trials, and a clear trend was not observed. Furthermore, the maximal information retrieved by a rate code, $I^*_{count}(S^P, S)$, converged differently than $I^*_{max}(S^P, S)$.

3.5.4 The application of discrimination analysis to experimental data: transient constant vs. time-dependent stimuli

The measures studied here by construction cannot discriminate between temporal structure caused by correlations with the stimuli, patterns associated to temporal encoding as defined by Theunissen and Miller (1995), or the effect of refractory periods or short-term adaptation. However, depending on the experimental setting, some of these sources are expected to be more relevant and should be taken into account to interpret the results. In general, two main scenarios can be distinguished.

In one case, transient constant stimuli are presented and the responses are recorded including the period of the stimulus presentation and potentially some period after the stimulus presentation (Victor and Purpura,

1996, 1998; Reich et al., 2001b; Aronov et al., 2003; Di Lorenzo and Victor, 2003; Di Lorenzo et al., 2009). In Section 3.2.2 we investigated the influence of the length of the spike trains L used in the discrimination analysis for simulations of different types of responses to transient presentation of constant stimuli. In Section 3.3 we analyzed exemplary cells responding to this type of stimuli. For this scenario, since the stimulus is constant, the temporal structure is not determined by correlations with the stimulus apart from those elicited by the onset and the offset. Therefore, we used the term temporal coding in the broad sense of any contribution of time to determine the information in the responses as in Hallock and Di Lorenzo (2006), in contrast to the more restrictive definition of Theunissen and Miller (1995). The temporal structure is mainly caused by the phasic response and short-term adaptation (see Hallock and Di Lorenzo, 2006, for an overview of the different sources of temporal coding for the cells of the gustatory system). The importance of the transient presentation in the temporal coding was demonstrated by Mechler et al. (1998) for the coding of contrast by $V1$ neurons. After the phasic response the activity depends on the influence of recurrent and feedback circuits (Victor and Purpura, 1998), which contributes to the dependence of the results of the discrimination analysis on L.

Furthermore, for the constant stimuli, the property of the stimulus which is assumed to cause the response is implicitly selected in the construction of the set. For example, for the experimental data analyzed in Section 3.3, the taste quality differentiates the stimuli presented to the gustatory cells, while the spatial phase changes for the stimuli presented to the visual cell. Other properties of the stimulus, for example the contrast for the visual stimuli, are kept constant or assumed to have no causal effect in the responses. In this case, the mutual information between the stimuli and the predicted stimuli is a lower bound of the mutual information between the stimuli and the response. Accordingly, the main problem of interpretation when obtaining a low value for the mutual information is that with $I^*_{max}(S^P, S)$ one evaluates at the same time the capability of the neurons to discriminate and the capability of the spike train distances and the classifier to correctly retrieve any available information. Low

values can be due to an unappropriate measure or classifier, but also because the cell is insensitive to the selected property of the stimulus which distinguish the different stimuli in the set. In a considerable part of the studies applying the discrimination analysis, low values of $I^*_{max}(S^P, S)$ have been obtained. For example, an average $I^*_{max}(S^P, S) < 0.3$ was extracted in Victor and Purpura (1996), Victor and Purpura (1997), Victor and Purpura (1998), Mechler et al. (1998), Reich et al. (2001b), Samonds and Bonds (2004), and Di Lorenzo and Victor (2007). In other cases average values were higher, for example $I^*_{max}(S^P, S) > 0.6$ for Di Lorenzo and Victor (2003), Roussin et al. (2008), and Di Lorenzo et al. (2009), but still low values were found for some cells. In general a better discrimination has been found when using time-dependent stimuli (Huetz et al., 2006, 2009).

For the other scenario time-dependent stimuli are used. Discrimination analysis has mainly been applied in this context to discriminate between naturalistic sounds (Machens et al., 2001, 2003; Narayan et al., 2005, 2006; Huetz et al., 2006; Wohlgemuth and Ronacher, 2007; Huetz et al., 2009). In Section 3.4 we analyzed exemplary cells from the auditory system of birdsongs responding to conspecific songs. In this context, the time-dependent stimuli are regarded as a discrete set of different stimuli rather than simply different realizations from an underlying common probability distribution. Each song presented to a songbird is supposed to be behaviorally relevant and to contain subunits formed by temporal sequences of motifs composed by complex spectrotemporal features that are specific for different singers (Gentner, 2008). Neurons with spectrotemporal receptive fields have been identified in the birds auditory system (Theunissen et al., 2000; Nagel and Doupe, 2008). Accordingly, the temporal structure of the spike train reflects mainly the correlation with the time-dependent stimulus, the particular sequence of spectrotemporal features to which the neuron is sensitive. (Narayan et al., 2005; Billimoria et al., 2008). For neurons with different receptive field the time-dependent stimulus constitutes a different time-dependent profile of relevant features. This results in a highly variable mutual information and discriminative precision across cells (Narayan et al., 2005, 2006; Wang

et al., 2007; Schneider and Woolley, 2010). Therefore, an analysis of the influence of the length L of the spike trains in the discrimination accuracy depends crucially on the specific time-dependent profile resulting for each song and cell. In Sections 3.2 and 3.3, to restrict the number of influencing factors on the estimation of the mutual information and the discriminative precision, we focused on examples reproducing the responses to transient constant stimuli. Oppositely, in Section 3.4 we used responses to time-dependent stimuli to exemplify the influence of the responsiveness of the cells and the statistics of the time-dependent stimuli.

For time-dependent stimuli like the songs in Section 3.4, there is no clear identification of what is causally influencing the responses. The stimuli forming the set $\{S\}$ are not distinguisable by one property, like the spatial phase, but they are behaviorally relevant complex entities, consisting of a sequence of features resulting in a particular time-dependent profile of relevant features for each cell. In this case the mutual information between stimuli and predicted stimuli is not a lower bound of the mutual information between stimuli and responses, and should be consider simply as a measure of discrimination performance, similar to the percentage of correct classifications (Machens et al., 2003; Narayan et al., 2006). The percentage of correct classifications cannot be directly compared to the normalized mutual information. However, qualitatively, higher levels of discrimination performance have been reported when analyzing time-dependent stimuli. This higher levels of performance can be understood given that in the time-dependent stimuli the evidence of the difference between the stimuli increases with time, if the cell is sensitive to some features in the stimuli. For example, percentages higher than 75% were reported in Machens et al. (2001, 2003); Narayan et al. (2005, 2006). The discrimination level depends on the influence of the distribution of these relevant features specific for each cell across time and across stimuli. This means that the statistics of the stimuli, combined with the properties of receptive field and the tuning curve of the cell have to be considered together with the sensitivity of the similarity measures and the appropriateness of the classifier to interpret the discrimination performance.

Discrimination analysis for transient constant stimuli

In Section 3.3 we studied exemplary spike trains elicited by transient constant stimuli. In Section 3.3.1 we showed that the dependence on the measure and classifier used was consistent with the results for the simulated data discussed in Section 3.2.2. We again observed that it is harder to estimate and evaluate the discriminative precision and significance of the temporal coding than the mutual information. We furthermore included in our analysis a direct estimate of $I^*(R,S)$, the mutual information between the stimuli and the total rates in the spike trains. With the number of trials available it was not possible to deal with a space $\mathbf{R}$ of high dimensionality, derived from a finer binning of the spikes. However, calculating $I^*(R,S)$ helps to verify the relevance of the temporal coding. Since $I^*_{count}(S^P,S)$ is only a lower bound of $I^*(R,S)$, one should check if $I^*(R,S)$ is higher than $I^*_{max}(S^P,S)$. To our knowledge, such a control including $I^*(R,S)$ in the analysis has only been carried out by Huetz et al. (2006).

We studied the influence of the length L of the spike trains used for the discrimination analysis. We found some similarities between the simple examples of Section 3.2.2 and the experimental data. In particular, for cell g_2, we showed the effect of the existence of information at different time scales, as in example B of Section 3.2.2, where both the latency and the total spike counts were informative. We argued that one quantity, τ^*, is not enough to reflect the time scales at which information is contained in the code. Its value depends in a nontrivial way on the balance between how the sources of information are distributed in different locations of the spike trains and on the time scale characteristic of these sources.

The dependence of the discrimination analysis on the length of the responses has been studied before (Machens et al., 2001, 2003; Narayan et al., 2005, 2006; Wang et al., 2007; Wohlgemuth and Ronacher, 2007). However, our analysis differed in several ways. First, in all these studies time-dependent stimuli were used, so that increasing L also increases the amount of evidence in the stimuli contributing to their discrimination. Therefore in these cases how the discrimination improves depends more on the statistics of the different time-dependent profiles than on the encod-

ing process. Second, in these studies the dependence on L is shown for a fixed value of τ, usually corresponding to τ^* extracted for a particular sufficiently large length of the spike trains (for which the mutual information is high). Oppositely, in Figure 3.10, the values of $I^*_{max}(S^P, S)$ were maximized independently for each value of L. As indicated by the variability of τ^* on L, using a fixed τ can led to underestimate the available information. Narayan et al. (2006) derived an estimate of the time scale of temporal integration fitting the cumulative curves to an exponential. They discussed the existence of two distinct time scales in the discrimination, one related to the discriminative precision and the other to the time of integration. A longer integration time scale can result from the underestimation of $I^*_{max}(S^P, S)$ when τ is fixed for all L.

To better understand the accumulation of information, we examined which amount of information can be extracted locally in different parts of the spike trains and how redundant it is (Section 3.3.3). We saw that in different segments the best classifier differs. In some cases, even if $I^*_{max}(S^P, S)$ was almost equal, different classifiers indicated different discriminative precision and importance of the temporal coding. It was not possible to relate the discriminative precision to the way information was temporally distributed. For cell g_2 the information was almost homogeneously distributed. For cell v_1 it was concentrated at the times of the onset and offset of the stimulus presentation. The analysis confirmed the observed accumulation of information, indicating that the first $L \simeq 2$ s for the gustatory cells and the first $L \simeq 400$ ms for the visual cell already contains almost all the nonredundant information. Nonetheless, the information contained in the posterior part of the responses varies from cell to cell. For cell g_2, almost the same information could be obtained excluding the first 2 s. By contrast, for cell v_1, almost no information is contained after the first 400 ms.

For transient constant visual stimuli, Reich et al. (2001b) already carried out the discrimination analysis considering separately different parts of the responses. They compared the information extracted in the transient part of the response, on the tonic part, and the off part of the response -after the stimulus has been removed. Although they did not focus on the

redundancy of the information in these distinct parts, they already indicated that the information contained in the time of the first spike and all the other sources of information were redundant, so that the first spike was even more informative than the full response for some cells. In another study, Nelken et al. (2005) used other estimators of the mutual information to identify a minimal ensemble of properties of the responses that contains all the available information. It was shown that the total rates and the mean of the spike times contained together the same information as the full responses. Examining which parts of the responses contribute the most to the mutual information provides more information about the time scales and the nature of the encoding than the discriminative precision. The way information is distributed and combined in different parts of the responses and in different time scales can be too complex to be faithfully reflected by one quantity.

Discrimination analysis for time-dependent stimuli

In Section 3.4 we analyzed the responses of exemplary cells to time-dependent stimuli. In particular, we examined the responses of auditory cells of songbirds when listening to conspecific songs. We have examined the influence of the receptive field and the tuning curve indirectly by studying the correlation of the mutual information and the optimal time scale for discrimination with the average number of spikes per song. We did it in this way due to the difficulty to estimate the spectrotemporal receptive field of these cells and their tuning curves, in particular with only a small number of $M = 6$ repetitions of each song. We considered a preliminary estimation of the receptive fields following the reverse-correlation technique described in Theunissen et al. (2000) too poor to be used for further analysis. We attribute this poor estimation to the rich temporal structure of the responses incompatible with the Poissonian spike generation necessary for the consistency of spike triggered average estimators (Paninski, 2003a; Schwartz et al., 2006). Furthermore, we found, for many of the neurons, a low frequency (~ 0.1 Hz) modulation of the responses unrelated to the spectrotemporal properties of the songs (see for

example the second cell in Figure 3.15).

We found a strong positive rank correlation between the mutual information and the average number of spikes per song (Section 3.4.1). These results are consistent with other studies that found the same positive correlation when calculating the mutual information from the discrimination analysis (Schneider and Woolley, 2010) in the midbrain of zebra finches, or when calculating the mutual information rate between acoustic stimuli and responses using the direct method of Strong et al. (1998), like Escabí et al. (2005) in the cat inferior colliculus. For the mutual information rate, this correlation was explained by Escabí et al. (2005) using a simple integrate-and-fire neuron model. In this model, for the same synaptic drive, the spike threshold determines the firing rate and the selectivity of the neuron to specific stimuli, so that the information per spike also increases with the threshold. With an alternative model, Narayan et al. (2005) showed that the discrimination performance was higher when the receptive field contains an inhibitory component, which also results in a higher selectivity of the neuron. However, despite the positive rank correlation, both in Figure 3.17 and in Figure 3 in Schneider and Woolley (2010), high values of the mutual information are also obtained for some cells with low rates, indicating that discrimination is highly dependent on the distribution of the few spikes across time and stimuli. By contrast, when the information rate between stimuli and responses is calculated (Figure 4 Escabí et al. (2005)), the information rates for the lower rates are never comparable to the ones of higher rates.

We also found a negative rank correlation between the optimal time scale and the average number of spikes per song. This correlation is consistent with the fact that the correlation with the time-dependent stimuli is the main source of temporal structure in the spike trains. When the cell is not very selective, the rate is higher and the local differences in the modulation are blurred when considered in longer windows. For highly selective cells responding reliably and sparsely, the optimal time scale can be higher, but, above all, is principally more sensitive to the distribution of the spikes across time and stimuli, as it occurs for the mutual information.

To scrutinize the relation between the existence of reliable spikes or

firing patterns in the spike trains and the discriminative precision, we constructed surrogate data preserving the structure across trials but not across stimuli inside segments of a given length. These surrogates differ from other surrogates previously used in combination with the discrimination analysis. Victor and Purpura (1996) proposed two resampling techniques to evaluate if the obtained mutual information can be accounted by considering that the spike trains result from Poisson processes with modulated rates. The Poisson hypothesis is tested by creating surrogate data pooling the trials of the same stimulus and then randomly reassigning each spike to the spike trains. This procedure uses the peristimulus time histogram (PSTH) as an estimate of the underlying modulated rate. The discrimination analysis is repeated for independent realizations of the surrogates to compare with the results from the original spike trains (e. g. Di Lorenzo and Victor, 2003; Wang et al., 2007). The other type of resampling proposed by Victor and Purpura (1996), was the exchange resampling. It was motivated by the observation that the Poisson surrogates led to higher values of the mutual information when the variability of the spike counts across trials was higher than Poissonian. The exchange surrogates maintain the number of spikes in each spike train. At the same time temporal correlations are destroyed by repeatedly randomly exchanging two spikes of two trials of the same stimulus. These surrogates are individually not consistent with the Poisson hypothesis, since it is unlikely to get the same number of spikes for two spike trains generated from a Poisson process. By contrast, they represent the null hypothesis that *the time-dependent Poisson process is modulated by a slowly varying influence unrelated to the stimulus* (Victor and Purpura, 1996). The two resampling techniques maintain the original PSTH for each individual stimuli, and alter the reliable patterns across trials, just oppositely to the surrogates we proposed here.

We defined a time scale related to the reliable patterns as an alternative to the optimal time scale for discrimination. We found that this time scale does not depend on the average number of spikes per song and is accordingly also uncorrelated to the mutual information and the discriminative precision. This time scale is not completely independent of the

arbitrary distribution of relevant features across time and across stimuli, since the mutual information obtained for the original data is taken as a reference for the null hypothesis rejection. Therefore, we do not claim that the specific value of the reliable patterns time scale obtained for each cell indicates its degree of reliability. To evaluate properly the reliability it would be necessary to know what the cell is encoding. However, the lower order of magnitude of this time scale compared to the optimal time scale for discrimination indicates that the latter does not correspond to the time scale up to which spike timing is informative.

For transient constant stimuli, the temporal structure is mainly related to the transient response and to short-term adaptation. In this context, despite of the problems of interpretation of the discriminative precision and the importance of temporal coding, it is pertinent to ask wether spike times are relevant or not. When time-dependent stimuli are used, the temporal structure is considered to be mainly caused by the correlation with the time-dependent stimulus. Given that, spike timing necessarily has to play a role.

To our knowledge, the first applications of the discrimination analysis to complex time-dependent stimuli were presented in Machens et al. (2001, 2003), examining the responses of auditory receptors in grasshoppers to conspecific signals. The main conclusion of these studies was that sufficient information to distinguish songs is readily available for single neurons when the spike trains are analyzed on a millisecond time scale. Machens et al. (2001) showed the discrimination performance of a single exemplary cell, and Machens et al. (2003) showed the average and the standard deviation of the performance of an ensemble of cells. Based on the average discrimination performance, the optimal time scale was extracted and associated with the time scale of the features of the song that the neurons encode. From the ensemble statistics of the discrimination performance it is not possible to judge the variability of the optimal time scale for each cell. However, our results indicate that this correspondence between the optimal time scale and a characteristic time scale of the features to which the neuron is sensitive to cannot be concluded from the discrimination analysis.

In other studies, in which the results are considered separately for each cell, a high variability of the discrimination performance (Wang et al., 2007) and the optimal time scale (Narayan et al., 2006) across cells illustrate the difficulty to interpret the results of the discrimination analysis applied to time-dependent stimuli. We do not mean here that proving that single neuron responses suffice to discriminate the songs is not an interesting result on itself. However, that one neuron discriminates better than another or which is the optimal time scale of discrimination should be interpret with caution. Furthermore, average results across cells are only meaningful if the cells are sensitive to similar features in the stimuli. For an heterogeneous group of cells like the one studied here (Section 3.4), and as generally expected in Field L and CM, an average like the one presented in Machens et al. (2003) is not very informative. To study which is the time scale of the features to which the neuron is sensitive, we believe that characterizing the spectrotemporal receptive field is a better alternative. However, this is a difficult task when using naturalistic non-Gaussian stimuli (Sharpee et al., 2004; Schwartz et al., 2006), and when the dimensionality of the stimuli is high, like for the acoustic stimuli. Furthermore, a more realistic spike generation modeling (e. g. Berry and Meister, 1998; Paninski et al., 2004; Pillow et al., 2005) is needed when the spike train differ from the Poissonian statistics.

3.5.5 Spike train distances to study population coding

Extensions for population codes of the Victor distance (Aronov et al., 2003) and of the van Rossum distance (Houghton and Sen, 2008) have been proposed, and our conclusions derived for single neurons are applicable to these measures as well. These extensions have been developed adding a new parameter that quantifies if the interactions between cells in the population contribute to a better discrimination of the stimuli. In one extreme of the parameter range, the spike train distances are insensitive to which neuron fires each spike, while in the other the distances are calculated separately for each cell. To determine the contribution of interactions to discrimination one has to take as a reference the maxi-

mal discrimination achievable without considering the interactions. Since $I^*_{max}(S^P, S)$ is only a lower bound of $I^*(\mathbf{R}, S)$, the evaluation of population codes with spike train distances is even harder, and the parameter associated with the population coding more difficult to interpret than τ^*.

3.6 Conclusions

In this part of the thesis we analyzed the factors influencing a discrimination analysis that combines a measure of spike train similarity with a classifier. This analysis provides the mutual information between stimuli and predicted stimuli, and an optimal time scale of discrimination. We showed that the mutual information obtained from the classification is more robust than the discriminative precision, or any measure quantifying the relevance of temporal coding, to the factors and parameters influencing the discrimination analysis. This mutual information is a lower bound of the mutual information between the stimuli and the responses when each category of the classifier is associated with a constant stimuli. By contrast, it is generally difficult to interpret the meaning of the discriminative precision related to the optimal time scale which, strictly, indicates nothing more than the parameter for which the maximal mutual information is obtained.

Spike train distances are generally useful to quantify the level of similarity between neural responses (e. g. Kreiman et al., 2000), or to evaluate the performance of single neuron models to reproduce the spike times elicited by a stimulus (Jolivet et al., 2008). Furthermore, they have been often applied to characterize the temporal structure in the spike trains relevant for the neural code (e. g. Victor, 2005, and references therein). However, we showed that it is difficult to interpret the results obtained using time scale parametric spike train distances to decipher the neural code or to identify biologically meaningful time scales. According to our results, the time scale parametric nature of these measures is mainly an advantage in the sense that it allows maximizing the mutual information across a whole set of measures with different sensitivities. This is in contrast to the view that the main advantage is the possibility to calculate

the discriminative precision and examine temporal coding. Therefore, to find the maximal mutual information, more elaborated classifiers could be employed. Similarly, for spike trains with a rate profile containing more than one relevant time scale, like in the presence of bursting, time scale adaptive spike train distances (Kreuz et al., 2007a, 2009, 2011) could be more appropriate to obtain the maximal mutual information.

Chapter 4

Conclusions

In this thesis we studied different techniques used to analyze the dependence between brain signals, or between the stimuli and the neuronal responses. In Chapter 2 we focused on the assessment of causal interactions, which is usually carried out from continuous signals like local field potentials (LFP's), or BOLD signals, reflecting the activity neural regions. In Chapter 3 we focused on the discrimination analysis, designed to examine how well a cell discriminates the stimuli given the elicited spike trains, and to study the discriminative precision as well as the role played by spike timing in the neural encoding. The conclusions specific to each type of analysis were already discussed in each part. We here briefly consider what they have in common. We then highlight our main achievements and point to the work in progress related to the content of the thesis.

In both cases, we characterized the specificity of the measures used to study the dependencies or properties of interest. For the causality measures, we identified various sources of bias related to the properties of the different dynamics or to the levels of noise. For the quantities derived from the discrimination analysis, we showed that their depend on multiple factors, including, for example, the length of the spike trains used for the discrimination analysis, or the statistics of the stimuli. The emphasis on the interpretation of the measures was a key aspect in our analysis. We discussed all the obstacles that prevent from assessing causality using the

measures implementing the criterion of the mapping of nearest neighbors or Granger causality. Similarly, we showed that the discrimination precision associated with the optimal time scale for discrimination is a concept tightly related to the statistical analysis carried out, and difficult to interpret in terms of time scales of the stimulus or biological time scales relevant for the neural code.

The causal analysis and the discrimination analysis studied here have in common that they are data-driven approaches. When studying such a complex system as the brain, it is very difficult to find a statistical quantity specific to a neuronal mechanism or dependence. This should be taken into account when studying the effective connectivity and the neural code, respectively, using these techniques. It might be argued that, although too unspecific in general, the statistical quantities used for the study of causality as well as the discrimination performance and the discriminative precision, provide useful information when calculated for the appropriate data. We here did not exclude this possibility, and indeed show, for the case of the discrimination analysis, that it can be useful to examine the temporal distribution of information about the stimuli. However, based on our examples and reviewing previous applications of the techniques, we showed that one has to be cautious when interpreting the statistical quantities in terms of underlying neuronal properties.

Regarding the analysis of causality, our main achievement was the systematic analysis of the sources of bias in dependence on the coupling strength, the levels of noise and the parameters used for the reconstruction (Sections 2.2.3, 2.2.4). We focused on few examples to explain in detail the biases in the different regimes, and show that the appropriate normalization substantially improves the specificity, specially for weak couplings. We also proposed a new measure using rank statistics that further attenuates the remaining biases (Chicharro et al., 2008; Andrzejak et al., 2008; Chicharro and Andrzejak, 2009). Furthermore, we reviewed the literature critically, considering the assumptions required by the criteria to test for causality and the limitations due to the estimation of the measures of causality from finite data sampling (Section 2.3). Although there are several review articles focusing on the study of causal interac-

tions in the brain from time series analysis (Gourevitch et al., 2006; Ding et al., 2006; Bressler and Seth, 2010), they are more focused on the description of the measures than on their interpretation. In particular, to our knowledge, the difference between testing for causality and quantifying the causal effect (Chicharro and Andrzejak, 2010) has not been addressed in the field. Here we only discussed this difference examining the definition of the criterion of Granger causality.

For the discrimination analysis, our main purpose was to illustrate all the influencing factors preventing from a direct interpretation of the measures in terms of the encoding properties of the cell. We showed that the discrimination analysis may be more useful to explore the structure of the responses than to extract a few quantities like the mutual information, the optimal time scale of discrimination or some measure quantifying the relevance of temporal coding. These quantities reflect the interplay of too many factors, and are difficult to interpret in general (Chicharro et al., 2009, 2010a). An exception is the mutual information in the case of a set of constant stimuli. In this case the mutual information between stimuli and predicted stimuli is a lower bound of the mutual information between the stimuli and the responses. We found that, in contrast to the emphasis put on the spike train distances in the literature, for our simulated data and exemplary cells, the classifier was the preeminent ingredient to maximize the mutual information (Section 3.2). Furthermore, we extended the usual application of the discrimination analysis in several ways. We proposed a procedure to generate surrogate data that allows investigating the time scale of the reliable patterns in the spike trains (Chicharro et al., 2010b). We also examined the temporal distribution, accumulation and redundancy of the information contained in the responses elicited by transient constant stimuli (Section 3.3). We illustrated that, applied in this way, the analysis can provide more insights into the nature of the encoding (Chicharro et al., 2011).

Appendix A

The relation between the dimension of an attractor and the number of neighbors

The Box-counting dimension of an attractor is defined by considering the number of cubes $N(\epsilon)$ of size ε in an N-dimensional space needed to cover the attractor (see Chapter 3 in Ott, 2002 for a detail explanation). However, this dimension is only sensitive to the support of the attractor, and does not consider how homogeneously the different cubes are occupied. The natural measure μ of an attractor indicates how much time the trajectories expend in each part of the attractor. Accordingly, generalized dimensions characterizing the geometrical structure of the attractor were defined (Grassberger, 1983) as:

$$D_q = \frac{1}{1-q} \lim_{\varepsilon \to 0} \frac{\ln \sum_{i=1}^{N(\varepsilon)} \mu_i^q}{\ln 1/\varepsilon}. \tag{A.1}$$

In particular, for $q = 2$, the dimension can be estimated most easily from experimental data using the correlation sum (Grassberger and Procaccia, 1983):

$$C(\varepsilon) = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} u(\varepsilon - ||\mathbf{x}_i - \mathbf{x}_j||), \tag{A.2}$$

where u is the Heaviside function such that $u(x) = 0$ if $x \leq 0$ and $u(x) = 1$ for $x > 0$, and N is the number of points. The correlation dimension is

defined as:

$$\mathrm{D} = \lim_{\varepsilon \to 0} \lim_{N \to \infty} \frac{\partial \ln C(\varepsilon, N)}{\partial \ln \varepsilon} \tag{A.3}$$

This definition is based on the expectation that in the limit of $\varepsilon \to 0$ and $N \to \infty$ the correlation sum scales like:

$$C(\varepsilon) \propto \varepsilon^{\mathrm{D}}. \tag{A.4}$$

Therefore, the number of neighbors for $\varepsilon \to 0$ is related to the dimension of the dynamics. This means that even for a fractal dimension, there are an uncountable infinite number of trajectories in a given region of the attractor, so that it is possible to find a neighbor close enough for Equation 2.9 to hold. This is true for the original and also for a proper reconstructed delay coordinates space, since it preserves the dimension of the original space.

Appendix B

Comparison of the discriminative precision and the spike timing precision

We here discuss the relation between the discriminative precision, obtained from the discrimination analysis, and the spike timing precision obtained from the direct calculation of the mutual information $I(\mathbf{R}, S)$. For a time-dependent stimulus, the direct method proposed by Strong et al. (1998) to calculate $I(\mathbf{R}, S)$ makes minimal assumptions about the code. The spike trains are converted in words of length T, and each word contains bins of width dt. The mutual information is calculated as:

$$I(\mathbf{R}, S; T, dt) = H(\mathbf{R}; T, dt) - H(\mathbf{R}|S; T, dt). \qquad \text{(B.1)}$$

To compute the entropy $H(\mathbf{R}; T, dt)$, the responses to a single repetition of a long sequence of the time-dependent stimulus is used, which samples appropriately the distribution $p(S)$. Notice that this implies assuming the stationarity of the response, based on the time-dependent nature of the stimulus and the stationarity of the stimulus distribution. Furthermore, assuming this stationarity also requires that the neuron encodes some property of the stimulus which varies fast enough to avoid short-term adaptation and transient responses. This could occur for example for highly selective neurons responding only to particular complex features of the stimulus, which appear sparsely in the time-dependent profile. The neural noise entropy $H(\mathbf{R}|S; T, dt)$ has to be calculated using a set of repetitions of the same time-dependent stimulus and it is therefore harder to estimate from limited experimental data. $I(\mathbf{R}, S)$ is then

obtained as the extrapolation for $T \to \infty$ and $dt \to 0$. This procedure allows to estimate the spike timing precision of the neural code (e. g. de Ruyter van Steveninck et al., 1997; Buracas et al., 1998; Reinagel and Reid, 2000; Wright et al., 2002). In general, this precision is not equal to the precision of single spikes across trials of the same stimulus (see Borst and Theunissen, 1999, for a discussion of the interpretation of the spike timing precision for linear or nonlinear encoding). When dt is decreased the mutual information is expected to increase until a plateau is reached, indicating that at this precision all the information is already obtained. The data processing inequality guarantees that $I(\mathbf{R}, S)$ does not decrease for $dt \to 0$. By contrast, one has to take into account that in this limit, its estimate will be increasingly upward biased due to limited data.

For constant stimuli, the direct method cannot be applied. In particular, the stationarity in time of the response to a constant stimulus cannot be assumed. In that case both the entropy of the responses and the neural noise are calculated across trials of each stimulus (Buracas et al., 1998). The distributions of $p(\mathbf{R})$ and $p(\mathbf{R}|S)$ are calculated across trials considering $\mathbf{R}$ time-locked to the presentation of each constant stimulus. The spike timing precision can also be studied increasing the number of bins to get $\mathbf{R}$, until a plateau is found for $I(\mathbf{R}, S)$. This precision depends on the set or distribution of stimuli used (Buracas et al., 1998), like they do the discriminative precision and the mutual information in the discrimination analysis (Reich et al., 2001b; Samonds and Bonds, 2004; Roussin et al., 2008). Furthermore, it is conceptually different from discriminative precision. None of them is directly related to the precision of single spikes, but while the first is determined by the saturation of $I(\mathbf{R}, S)$ for decreasing dt, the second is related to the peak of $I^*(S^P, S)$ found for a particular range of τ. No information is lost for $dt \to 0$, when the dimensionality of $\mathbf{R}$ increases due to a finer resolution examining the responses. This is assured by the data processing inequality. By contrast, for $\tau \to 0$ the optimality of the classifier decreases. If the maximum $I^*_{max}(S^P, S)$ is found for a given τ^*, this does not imply that the saturation of $I(\mathbf{R}, S)$ is attained for a dt related to τ^*.

Appendix C

Electrophysiology of the songbird's recordings

This description has been provided by Emily Caporello, from the the Laboratory of Auditory Neuroethology and Communication of Tim Gentner, at the University of California, San Diego.

One day prior to recording, we anesthetized each bird (n=9) with isoflurane (1.5-3.0%) and attached a metal pin to the skull with dental acrylic. We verified full auditory and behavioral recovery in the operant apparatus prior to recording. On the recording day, we anesthetized each bird with 20% urethane (7ml/kg; administered in 3 IM injections over 90 minutes), wrapped in a cloth jacket, then head-fixed to a stereotaxic apparatus inside a sound-attenuating recording chamber (Acoustic Systems). A small craniotomy and durectomy approximately 2400 um rostral and 1000 um lateral to the bifurcation of the saggital sinus permitted electrode access to the target brain region. We coated a 16-channel tetrode array (NeuroNexus Technologies, Ann Arbor MI) with fluorescent diI crystals (Sigma-Aldrich, St. Louis MO) before lowering the electrode into the brain. Extracellular waveforms were amplified by a 16-channel AC amplifier (AM systems Inc, Sequim WA), and digitized and stored (CED, Cambridge UK) for offline analysis. To locate auditory sites, we pseudo-randomly presented Song+Noise, Song+Silence, and Noise+Silence stimuli, and segments of unfamiliar conspecific song as the electrode was slowly advanced. When a putative auditory neuron was isolated, we played all stimuli to the anesthetized animal. At each record-

ing site we presented, randomly without replacement, several repetitions of each stimulus used in the present analysis. Spike2 software (CED) controlled stimulus playback and recorded time aligned extracellular voltage traces at all electrode sites. Offline, we extracted putative action potentials from the waveform using threshold-triggered template matching, and clustered putative action potential waveforms with a combination of PCA and changes in waveform shape over time (Spike2, CED). We confirmed isolation quality using the ISI distribution of each putative single or multi-unit. We considered single units to have well-separated waveform clusters and have less than 2% refractory-period violations.

Bibliography

Amblard, P. O. and Michel, O. (2010). On directed information theory and granger causality graphs. *J. Comput. Neurosci.*, doi:10.1007/s10827–010–0231–x pages 1–10.

Ancona, N., Marinazzo, D., and Stramaglia, S. (2004). Radial basis function approach to nonlinear granger causality of time series. *Phys. Rev. E*, 70(5):056221.

Andrzejak, R. G., Chicharro, D., and Ledberg, A. (2008). A new measure for the detection of directional couplings based on rank statistics. *International Symposium on Nonlinear Theory and its Applications*, Budapest, Hungary, September.

Andrzejak, R. G., Chicharro, D., Lehnertz, K., and Mormann, F. (2011). Using bivariate signal analysis to characterize the epileptic focus: The benefit of surrogates. *Phys. Rev. E*, under review.

Andrzejak, R. G., Kraskov, A., Stögbauer, H., Mormann, F., and Kreuz, T. (2003). Bivariate surrogate techniques: Necessity, strengths, and caveats. *Phys. Rev. E.*, 68:066202.

Andrzejak, R. G., Ledberg, A., and Deco, G. (2006a). Detection of event-related time-dependent directional couplings. *New J. Phys.*, 8:6.

Andrzejak, R. G., Mormann, F., Widman, G., Kreuz, T., Elger, C. E., and Lehnertz, K. (2006b). Improved spatial characterization of the epileptic brain by focusing on nonlinearity. *Epilepsy Res.*, 69(1):30–44.

Andrzejak, R. G., Widman, G., Lehnertz, K., David, P., and Elger, C. E. (2001). The epileptic process as nonlinear deterministic dynamics in a stochastic environment: an evaluation on mesial temporal lobe epilepsy. *Epilepsy Res.*, 44:129–140.

Ansari-Asl, K., Senhadji, L., Bellanger, J. J., and Wendling, F. (2006). Quantitative evaluation of linear and nonlinear methods characterizing interdependencies between brain signals. *Phys. Rev. E*, 74(3):031916.

Arnhold, J., Lehnertz, K., Grassberger, P., and Elger, C. E. (1999). A robust methosd for detecting interdepemdences: application to intracranially recorded EEG. *Physica D*, 134:419–430.

Aronov, D., Reich, D. S., Mechler, F., and Victor, J. D. (2003). Neural coding of spatial phase in v1 of the macaque monkey. *J. Neurophysiol.*, 89:3304–3327.

Barnett, L., Barrett, A. B., and Seth, A. K. (2009). Granger causality and transfer entropy are equivalent for gaussian variables. *Phys. Rev. Lett.*, 103(23):238701.

Barreto, E., Josić, K., Morales, C. J., Sander, E., and So, P. (2003). The geometry of chaos synchronization. *Chaos*, 13:151–164.

Barrett, A. B., Barnett, L., and Seth, A. K. (2010). Multivariate granger causality and generalized variance. *Phys. Rev. E*, 81(4):041907.

Berry, M. J. and Meister, M. (1998). Refractoriness and neural precision. *J. Neurosci.*, 18(6):2200–2211.

Besserve, M., Schoelkopf, B., Logothetis, N. K., and Panzeri, S. (2010). Causal relationships between frequency bands of extracellular signals in visual cortex revealed by an information theoretic analysis. *J. Comput. Neurosci.*, 29(3):547–566.

Bhattacharya, J., Pereda, E., and Petsche, H. (2003). Effective detection of coupling in short and noisy bivariate data. *IEEE T. Syst. Man. Cy. B.*, 33:85–95.

Bhattacharya, J., Petsche, H., and Pereda, E. (2001). Interdependencies in the spontaneous eeg while listening to music. *Int. J. Psychophysiol.*, 42(3):287–301.

Bialek, W., Rieke, F., de Ruyter van Steveninck, R., and Warland, D. (1991). Reading a neural code. *Science*, 252(5014):1854–1857.

Billimoria, C. P., Kraus, B. J., Narayan, R., Maddox, R. K., and Sen, K. (2008). Invariance and sensitivity to intensity in neural discrimination of natural sounds. *J. Neurosci.*, 28(25):6304–6308.

Borst, A. and Theunissen, F. E. (1999). Information theory and neural coding. *Nature Neurosci.*, 2(11):947–957.

Brainard, M. S. and Doupe, A. J. (2000). Auditory feedback in learning and maintanance of vocal behaviour. *Nature Rev. Neurosci.*, 1:31–40.

Bressler, S. L., Richter, C. G., Chen, Y., and Ding, M. (2007). Cortical functional network organization from autoregressive modeling of local field potential oscillations. *Stat. Med.*, 26(21):3875–3885.

Bressler, S. L. and Seth, A. K. (2010). Wiener granger causality: A well established methodology. *Neuroimage*, In Press, Corrected Proof doi:10.1016/j.neuroimage.2010.02.059.

Bressler, S. L., Tang, W., Sylvester, C. M., Shulman, G. L., and Corbetta, M. (2008). Top-down control of human visual cortex by frontal and parietal cortex in anticipatory visual spatial attention. *J. Neurosci.*, 28(40):10056–10061.

Brovelli, A., Ding, M., Ledberg, A., Chen, Y., Nakamura, R., and Bressler, S. L. (2004). Beta oscillations in a large-scale sensorimotor cortical network: Directional influences revealed by granger causality. *P. Natl. Acad. Sci. USA*, 101:9849–9854.

Bullmore, E. and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Rev. Neurosci.*, 10(3):186–198.

Buracas, G. T., Zador, A. M., DeWeese, M. R., and Albright, T. D. (1998). Efficient discrimination of temporal patterns by motion-sensitive neurons in primate visual cortex. *Neuron*, 20(5):959–969.

Cadotte, A. J., DeMarse, T. B., He, P., and Ding, M. (2008). Causal measures of structure and plasticity in simulated and living neural networks. *PLoS ONE*, 3(10):e3355.

Casdagli, M., Eubank, S., Farmer, J. D., and Gibson, J. (1991). State space reconstruction in the presence of noise. *Physica D*, 51:52–98.

Cenys, A., Lasiene, G., and Pyragas, K. (1991). Estimation of interrelation between chaotic observables. *Physica D*, 52:332–337.

Chen, Y., Rangarajan, G., Feng, J., and Ding, M. (2004). Analyzing multiple nonlinear time series with extended granger causality. *Phys. Lett. A*, 324(1):26–35.

Chicharro, D. (2007). Characterization of event-related time-dependent directional couplings. *Master thesis of the Màster en Tecnologies de la Informació, la Comunicació i els Mitjans Audiovisuals.*, Supervisor:R. G. Andrjezak.

Chicharro, D. and Andrzejak, R. G. (2009). Reliable detection of directional couplings using rank statistics. *Phys. Rev. E*, 80(2):026217.

Chicharro, D. and Andrzejak, R. G. (2010). Evaluation and interpretation of causality in complex networks. SYNCLINE 2010: Synchronization in Complex Networks, Bad Honnef.

Chicharro, D., Andrzejak, R. G., and Kreuz, T. (2009). Studying the precision of temporal neural code: Some limitations of the use of spike train distances. In *BMC Neuroscience 10 (Suppl 1)*, page P130. Eighteenth Annual Computational Neuroscience Meeting, Berlin, Germany.

Chicharro, D., Kreuz, T., and Andrzejak, R. G. (2010a). Limitations of time-scale parametric spike train distances to study precision and reliability. STM 2010: Workshop on spike train measures and their application to neural coding, Plymouth.

Chicharro, D., Kreuz, T., and Andrzejak, R. G. (2011). Studying the neural code with spike train distances. *J. Neurosci. Meth.*, under revision.

Chicharro, D., Kreuz, T., Caporello, E., Gentner, T. Q., and Andrzejak, R. G. (2010b). Limitations of spike train distances to study the timescales of natural sound discrimination. Fortyth Annual Meeting, Society for Neuroscience, program number 578.8.

Chicharro, D., Ledberg, A., and Andrzejak, R. G. (2008). A new measure for the detection of directional couplings based on rank statistics. In *BMC Neuroscience 9 (Suppl 1)*, page P148. Seventeenth Annual Computational Neuroscience Meeting, Portland, Oregon, USA.

Cover, T. M. and Thomas, J. A. (2006). *Elements of Information Theory*. John Wiley and Sons, 2nd edition.

Daunizeau, J., David, O., and Stephan, K. E. (2009). Dynamic causal modelling: A critical review of the biophysical and statistical foundations. *NeuroImage*, In Press, Corrected Proof doi:10.1016/j.neuroimage.2009.11.062.

Dauwels, J., Vialatte, F., Weber, T., and Cichocki, A. (2009). Quantifying statistical interdependence by message passing on graphs—part i: One-dimensional point processes. *Neural Comput.*, 21(8):2152–2202.

Dayan, P. and Abbot, L. F., editors (2001). *Theoretical Neuroscience, Computational and Mathematical Modeling of Neural Systems*. The MIT press, Cambridge, Massachusetts.

de Ruyter van Steveninck, R. R., Lewen, G. D., Strong, S. P., Koberle, R., and Bialek, W. (1997). Reproducibility and variability in neural spike trains. *Science*, 275(5307):1805–1808.

Deshpande, G., Sathian, K., and Hu, X. (2010). Effect of hemodynamic variability on granger causality analysis of fmri. *Neuroimage*, 52(3):884–896.

Di Lorenzo, P. M., Chen, J. Y., and Victor, J. D. (2009). Quality Time: Representation of a Multidimensional Sensory Domain through Temporal Coding. *J. Neurosci.*, 29(29):9227–9238.

Di Lorenzo, P. M. and Victor, J. D. (2003). Taste response variability and temporal coding in the nucleus of the solitary tract of the rat. *J. Neurophysiol.*, 90(3):1418–1431.

Di Lorenzo, P. M. and Victor, J. D. (2007). Neural coding mechanisms for flow rate in taste-responsive cells in the nucleus of the solitary tract of the rat. *J. Neurophysiol.*, 97(2):1857–1861.

Ding, M., Chen, Y., and Bressler, S. L. (2006). Granger causality: Basic theory and application to neuroscience. In *Handbook of Time Series Analysis: Recent Theoretical Developments and Applications.*, pages 437–460. Wiley-VCH Verlag.

Duncan, O. D. (1975). *Introduction to structural equation models*. Academic Press, New York.

Escabí, M. A., Nassiri, R., Miller, L. M., Schreiner, C. E., and Read, H. L. (2005). The contribution of spike threshold to acoustic feature selectivity, spike information content, and information throughput. *J. Neurosci.*, 25(41):9524–9534.

Feldmann, U. and Bhattacharya, J. (2004). Predictability improvement as an asymmetrical measure of interdependence in bivariate time series. *Int. J. Bifurcat. Chaos*, 14:504–514.

Fortune, E. and Margoliash, D. (1992). Cytoarchitectonic organization and morphology of cells of the field l complex in male zebra finches. *J. Comp. Neurol.*, 325:388–404.

Fraser, A. M. and Swinney, H. L. (1986). Independent coordinates for strange attractors from mutual information. *Phys. Rev. A*, 33(2):1134–1140.

Friston, K. J. (1994). Functional and effective connectivity in neuroimaging: A synthesis. *Hum. Brain Mapp.*, 2(5):56–78.

Friston, K. J., Harrison, L., and Penny, W. (2003). Dynamic causal modelling. *Neuroimage*, 19(4):1273–1302.

Gardner, D. (2004). Neurodatabase.org: networking the microelectrode. *Nature Neurosci.*, 7(5):486–487.

Gentner, T. Q. (2008). Temporal scales of auditory objects underlying birdsong vocal recognition. *J. Acoust. Soc. Am.*, 124(2):1350–1359.

Geweke, J. F. (1982). Measurement of linear dependence and feedback between multiple time series. *J. Am. Stat. Assoc.*, 77(378):304–313.

Geweke, J. F. (1984). Measures of conditional linear dependence and feedback between time series. *J. Am. Stat. Assoc.*, 79(388):907–915.

Gourevitch, B., Le Bouquin-Jeannes, R., and Faucon, G. (2006). Linear and nonlinear causality between signals: methods, examples and neurophysiological applications. *Biol. Cybern.*, 95(4):349–369.

Granger, C. W. J. (1963). Economic processes involving feedback. *Information and Control*, 6:28–48.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–38.

Granger, C. W. J. (1980). Testing for causality : A personal viewpoint. *J. Econ. Dynamics and Control*, 2(1):329–352.

Grassberger, P. (1983). Generalized dimensions of strange attractors. *Phys. Lett. A*, 97:227–230.

Grassberger, P. and Procaccia, I. (1983). Measuring the strangeness of strange attractors. *Physica D*, 9:189–208.

Greenland, S. and Brumback, B. (2002). An overview of relations among causal modelling methods. *Int. J. Epidemiol.*, 31(5):1030–1037.

Grewe, J., Kretzberg, J., Warzecha, A. K., and Egelhaaf, M. (2003). Impact of photon noise on the reliability of a motion-sensitive neuron in the fly's visual system. *J. Neurosci.*, 23(34):10776–10783.

Guevara, R., Velazquez, J. L. P., Nenadovic, V., Wennberg, R., Senjanovic, G., and Dominguez, L. (2005). Phase synchronization measurements using electroencephalographic recordings - what can we really say about neuronal synchrony? *Neuroinformatics*, 3(4):301–313.

Guo, S., Seth, A. K., Kendrick, K. M., Zhou, C., and Feng, J. (2008). Partial granger causality - eliminating exogenous inputs and latent variables. *J. Neurosci. Meth.*, 172(1):79–93.

Haavelmo, T. (1943). The statistical implications of a system of simultaneous equations. *Econometrica*, 11:1–12.

Hallock, R. M. and Di Lorenzo, P. M. (2006). Temporal coding in the gustatory system. *Neurosci. Behav. Rev.*, 30(8):1145–1160.

He, D., Zheng, Z., and Stone, L. (2003). Detecting generalized synchrony: An improved approach. *Phys. Rev. E*, 67:026223.

Hiemstra, C. and Jones, J. D. (1994). Testing for linear and nonlinear granger causality in the stock price-volume relation. *J. Financ.*, 49(5):1639–64.

Hlaváčkova-Schindler, K., Paluš, M., Vejmelka, M., and Bhattacharya, J. (2007). causality detection based on information-theoretic approaches in time-series analysis. *Phys. Rep.*, 441:1–46.

Houghton, C. and Sen, K. (2008). A new multineuron spike train metric. *Neural Comput.*, 20:1495–1511.

Huetz, C., Del Negro, C., Lebas, N., Tarroux, P., and Edeline, J. M. (2006). Contribution of spike timing to the information transmitted by hvc neurons. *Eur. J. Neurosci.*, 24.

Huetz, C., Philibert, B., and Edeline, J. M. (2009). A spike-timing code for discriminating conspecific vocalizations in the thalamocortical system of anesthetized and awake Guinea pigs. *J. Neurosci.*, 29(2):334–350.

Hunter, J. D., Milton, J. G., Thomas, P. J., and Cowan, J. D. (1998). Resonance effect for neural spike time reliability. *J. Neurophysiol.*, 80(3):1427–1438.

Jarvis, E. D. (2004). Learned birdsongs and the neurobiology of human language. *Ann. N.Y. Acad. Sci.*, 1016:749–777.

Jolivet, R., Kobayashi, R., Rauch, A., Naud, R., Shinomoto, S., and Gerstner, W. (2008). A benchmark test for a quantitative assessment of simple neuron models. *J. Neurosci. Meth.*, 169(2):417–424.

Kaiser, A. and Schreiber, T. (2002). Information transfer in continuous processes. *Physica D*, 166:43–62.

Kantz, H. and Schreiber, T. (2003). *Nonlinear Time Series Analysis*. Cambridge Univ. Press, Cambridge, UK, 2nd edition.

Kayser, C. and Logothetis, N. K. (2009). Directed interactions between auditory and superior temporal cortices and their role in sensory integration. *Frontiers in Integrative Neuroscience*, 3:article 7.

Kocarev, L. and Parlitz, U. (1996). Generalized synchronization, predictability, and equivalence of unidirectionally coupled dynamical systems. *Phys. Rev. Lett.*, 76:1816–1819.

Kraskov, A., Stogbauer, H., and Grassberger, P. (2004). Estimating mutual information. *Phys. Rev. E*, 69(6):066138.

Kreiman, G., Krahe, R., Metzner, W., Koch, C., and Gabbiani, F. (2000). Robustness and variability of neuronal coding by amplitude-sensitive afferents in the weakly electric fish eigenmannia. *J. Neurophysiol.*, 84(1):189–204.

Kreuz, T., Chicharro, D., Andrzejak, R. G., Haas, J. S., and Abarbanel, H. D. (2009). Measuring multiple spike train synchrony. *J. Neurosci. Meth.*, 183(2):287–299.

Kreuz, T., Chicharro, D., Greschner, M., and Andrzejak, R. G. (2011). Time-resolved and time-scale adaptive measures of spike train synchrony. *J. Neurosci. Meth.*, 195:92–106.

Kreuz, T., Haas, J. S., Morelli, A., Abarbanel, H. D. I., and Politi, A. (2007a). Measuring spike train synchrony. *J. Neurosci. Meth.*, 165:151–161.

Kreuz, T., Mormann, F., Andrzejak, R. G., Kraskov, A., Lehnertz, K., and Grassberger, P. (2007b). Measuring synchronization in coupled model systems: A comparison of different approaches. *Physica D*, 225:29–42.

Le Van Quyen, M., Martinerie, J., Adam, C., and Varela, F. (1999). Nonlinear analyses of interictal eeg map the brain interdepences in human focal epilepsy. *Physica D*, 127:250–266.

Lehnertz, K., Bialonski, S., Horstmann, M. T., Krug, D., Rothkegel, A., Staniek, M., and Wagner, T. (2009). Synchronization phenomena in human epileptic brain networks. *J. Neurosci. Meth.*, 183(1):42–48.

Lorenz, E. N. (1963). Deterministic nonperiodic flow. *J. Atmos. Sci.*, 20:130–141.

Lungarella, M., Ishiguro, K., Kuniyoshi, Y., and Otsu, N. (2007). Methods for quantifying the causal structure of bivariate time series. *Int. J. Bifurcat. Chaos*, 17(3):903–921.

Lütkepohl, H. (1993). Testing for causation between to variavles in a higher dimensional var model. In *Studies in applied econometrics*, pages 75–91. Physica, Heidelberg.

Lütkepohl, H. (2006). *New introduction to multiple time series analysis*. Springer-Verlag, Berlin.

Machens, C. K., Prinz, P., Stemmler, M. B., Ronacher, B., and Herz, A. V. M. (2001). Discrimination of behaviorally relevant signals by auditory receptor neurons. *Neurocomputing*, 38-40:263–268.

Machens, C. K., Schütze, H., Franz, A., Kolesnikova, O., Stemmler, M. B., Ronacher, B., and Herz, A. V. M. (2003). Single auditory neurons rapidly discriminate conspecific communication signals. *Nature Neurosci.*, 6:341–342.

Macleod, K., Backer, A., and Laurent, G. (1998). Who reads temporal information contained across synchronized and oscillatory spike trains? *Nature*, 395:693–698.

Mainen, Z. F. and Sejnowski, T. J. (1995). Reliability of spike timing in neocortical neurons. *Science*, 268(5216):1503–1506.

Marko, H. (1973). Bidirectional communication theory - generalization of information-theory. *IEEE T. Commun.*, 12:1345–1351.

Massey, J. (1990). Causality, feedback and directed information. *Proc. Intl. Symp. Info. Th. Appli., Waikiki, Hawai, USA*.

McIntosh, A. R. and Gonzalez-Lima, F. (1991). Structural modeling of functional neural pathways mapped with 2-deoxyglucose -effects of acoustic startle habituation on the auditory system. *Brain Res.*, 547(2):295–302.

McIntosh, A. R. and Gonzalez-Lima, F. (1994). Structural equation modeling and its application to network analysis in functional brain imaging. *Hum. Brain. Mapp.*, 2:2–22.

Mechler, F., Victor, J. D., Purpura, K. P., and Shapley, R. (1998). Robust temporal coding of contrast by v1 neurons for transient but not for steady-state stimuli. *J. Neurosci.*, 18(16):6583–6598.

Mormann, F., Andrzejak, R. G., Elger, C. E., and Lehnertz, K. (2007). Seizure prediction: the long and winding road. *Brain*, 130:314–333.

Nagel, K. I. and Doupe, A. J. (2008). Organizing principles of spectro-temporal encoding in the avian primary auditory area field l. *Neuron*, 58:938–935.

Nalatore, H., Ding, M., and Rangarajan, G. (2007). Mitigating the effects of measurement noise on granger causality. *Phys. Rev. E*, 75(3):031123.

Narayan, R., Ergun, A., and Sen, K. (2005). Delayed inhibition in cortical receptive fields and the discrimination of complex stimuli. *J. Neurophysiol.*, 94(4):2970–2975.

Narayan, R., Graña, G. D., and Sen, K. (2006). Distinct time-scales in cortical discrimination of natural sounds in songbirds. *J. Neurophysiol.*, 96(1):252–258.

Nelken, I., Chechik, G., Mrsic-flogel, T. D., King, A. J., and Schnupp, J. W. H. (2005). Encoding stimulus information by spike numbers and mean response time in primary auditory cortex. *J. Comput. Neurosci.*, 19:199–221.

Nolte, G., Meinecke, F. C., Ziehe, A., and Muller, K. R. (2006). Identifying interactions in mixed and noisy complex systems. *Phys. Rev. E*, 73(5,).

Ott, E. (2002). *Chaos in Dynamical Systems*. Cambridge Univ. Press, Cambridge, UK.

Paiva, A. R. C., Park, I., and Principe, J. C. (2009). A reproducing kernel hilbert space framework for spike train signal processing. *Neural Comput.*, 21(2):424–449.

Paiva, A. R. C., Park, I., and Principe, J. C. (2010). A comparison of binless spike train measures. *Neural Comput. Appl.*, 19(3):405–419.

Palus, M. and Vejmelka, M. (2007). Directionality of coupling from bivariate time series: How to avoid false causalities and missed connections. *Phys. Rev. E*, 75(5):056211.

Paninski, L. (2003a). Convergence properties of three spike-triggered analysis techniques. *Network: Comput. Neural Syst.*, 14:435–464.

Paninski, L. (2003b). Estimation of entropy and mutual information. *Neural Comput.*, 15:1191–1253.

Paninski, L., Pillow, J. W., and Simoncelli, E. P. (2004). Maximum likelihood estimation of a stochastic integrate-and-fire neural encoding model. *Neural Comput.*, 16(12):2533–2561.

Panzeri, S., Petersen, R. S., Schultz, S. R., Lebedev, M., and Diamond, M. E. (2001). The role of spike timing in the coding of stimulus location in rat somatosensory cortex. *Neuron*, 29(3):769–777.

Panzeri, S., Senatore, R., Montemurro, M. A., and Petersen, R. S. (2007). Correcting for the sampling bias problem in spike train information measures. *J. Neurophysiol.*, 98(3):1064–1072.

Pearl, J. (2009). *Causality: Models, Reasoning, Inference*. Cambridge University Press, New York, 2nd edition.

Pearl, J. (2010). An introduction to causal inference. *Int. J. Biostat.*, 6(2):article 7.

Pecora, L. M. and Carroll, T. L. (1990). Synchronization in chaotic systems. *Phys. Rev. Lett.*, 64:821–825.

Pecora, L. M., Moniz, L., Nichols, J., and Carroll, T. L. (2007). A unified approach to attractor reconstruction. *Chaos*, 17(1):013110.

Pereda, E., Quian Quiroga, R., and Bhattacharya, J. (2005). Nonlinear multivariate analysis of neurophysiological signals. *Prog. Neurobiol.*, 77:1–37.

Pereda, E., Rial, R., Gamundi, A., and González, J. (2001). Assesment of changing interdependencies between human electroencephalograms using nonlinear methods. *Physica D*, 148:147–158.

Pillow, J. W., Paninski, L., Uzzell, V. J., Simoncelli, E. P., and Chichilnisky, E. J. (2005). Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model. *J. Neurosci.*, 25(47):11003–11013.

Quian Quiroga, R., Arnhold, J., and Grassberger, P. (2000). Learning driver-response relationships from synchronization patterns. *Phys. Rev. E*, 61:5142.

Quian Quiroga, R., Kraskov, A., Kreuz, T., and Grassberger, P. (2002). Performance of different synchronization measures in real data: A case study on electroencephalographic signals. *Phys. Rev. E*, 65:041903.

Quian Quiroga, R. and Panzeri, S. (2009). Extracting information from neuronal populations: information theory and decoding approaches. *Nature Rev. Neurosci.*, 10(3):173–185.

Reich, D. S., Mechler, F., and Victor, J. D. (2001a). Formal and attribute-specific information in primary visual cortex. *J. Neurophysiol.*, 85(1):305–318.

Reich, D. S., Mechler, F., and Victor, J. D. (2001b). Temporal coding of contrast in primary visual cortex: When, what, and why. *J. Neurophysiol.*, 85(3):1039–1050.

Reinagel, P. and Reid, R. C. (2000). Temporal coding of visual information in the thalamus. *J. Neurosci.*, 20(14):5392–5400.

Rieke, F., Warland, D., de Ruyter van Steveninck, R. R., and Bialek, W. (1997). *Spikes: Exploring the neural code*. MIT Press, Cambridge, Massachusetts.

Roebroeck, A., Formisano, E., and Goebel, R. (2005). Mapping directed influence over the brain using granger causality and fmri. *Neuroimage*, 25(1):230–242.

Roebroeck, A., Formisano, E., and Goebel, R. (2009). The identification of interacting networks in the brain using fmri: Model selection, causality and deconvolution. *NeuroImage*, In Press, Corrected Proof doi:10.1016/j.neuroimage.2009.09.036.

Rolls, E. T., Treves, A., and Tovee, M. J. (1997). The representational capacity of the distributed encoding of information provided by populations of neurons in primate temporal visual cortex. *Exp. Brain Res.*, 114(1):149–162.

Rosenblum, M. G. and Pikovsky, A. S. (2001). Detecting direction of coupling in interacting oscillators. *Phys. Rev. E*, 64:045202(R).

Rosenblum, M. G., Pikovsky, A. S., and Kurths, J. (1996). Phase synchronization of chaotic oscillators. *Phys. Rev. Lett.*, 76(11):1804–1807.

Roussin, A. T., Victor, J. D., Chen, J. Y., and Di Lorenzo, P. M. (2008). Variability in responses and temporal coding of tastants of similar quality in the nucleus of the solitary tract of the rat. *J. Neurophysiol.*, 99(2):644–655.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.*, 66:688–701.

Rubinov, M. and Sporns, O. (2010). Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, 52(3):1059–1069.

Rulkov, N. F. and Afraimovich, V. S. (2003). Detectability of nondifferentiable generalized synchrony. *Phys. Rev. E*, 67:066218.

Rulkov, N. F., Sushchik, M. M., Tsimring, L. S., and Abarbanel, H. D. I. (1995). Generalized synchronization of chaos in directionally coupled chaotic systems. *Phys. Rev. E*, 51(2):980–994.

Samonds, J. M. and Bonds, A. B. (2004). From another angle: Differences in cortical coding between fine and coarse discrimination of orientation. *J. Neurophysiol.*, 91(3):1193–1202.

Sauer, T., Yorke, J. A., and Casdagli, M. (1991). Embedology. *J. Stat. Phys.*, 65(3-4):579–616.

Schiff, S. J. (2005). Dangerous phase. *Neuroinformatics*, 3:315–317.

Schiff, S. J., So, P., Chang, T., Burke, R. E., and Sauer, T. (1996). Detecting dynamical interdependence and generalized synchrony through mutual prediction in a neural ensemble. *Phys. Rev. E*, 54:6708–6724.

Schmitz, A. (2000). Measuring statistical dependence and coupling of subsystems. *Phys. Rev. E*, 62(5):7508–7511.

Schneider, D. M. and Woolley, S. M. N. (2010). Discrimination of communication vocalizations by single neurons and groups of neurons in the auditory midbrain. *J. Neurophysiol.*, 103(6):3842–3265.

Schnupp, J. W. H., Hall, T. M., Kokelaar, R. F., and Ahmed, B. (2006). Plasticity of temporal pattern codes for vocalization stimuli in primary auditory cortex. *J. Neurosci.*, 26(18):4785–4795.

Schreiber, S., Fellous, J. M., Whitmer, D., Tiesinga, P., and Sejnowski, T. J. (2003). A new correlation-based measure of spike timing reliability. *Neurocomputing*, 52-54:925–931.

Schreiber, T. (2000). Measuring information transfer. *Phys. Rev. Lett.*, 85:461–464.

Schreiber, T. and Schmitz, A. (2000). Surrogate time series. *Physica D*, 142:346–382.

Schwartz, O., Pillow, J. W., Rust, N. C., and Simoncelli, E. P. (2006). Spike-triggered neural characterization. *J. Vision*, 6(4):484–507.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell. Sys. Tech. J.*, 27:379–423, 623–656.

Sharpee, T., Rust, N. C., and Bialek, W. (2004). Analyzing neural responses to natural signals: Maximally informative dimensions. *Neural Comput.*, 16(2):223–250.

Smirnov, D. A. and Andrzejak, R. G. (2005). Detection of weak directional coupling: Phase-dynamics approach versus state-space approach. *Phys. Rev. E*, 71:036207.

Smirnov, D. A. and Bezruchko, B. P. (2003). Estimation of interaction strength and direction from short and noisy time series. *Phys. Rev. E*, 68:046209.

So, P., Barreto, E., Josić, K., Sander, E., and Schiff, S. J. (2002). Limits to the experimental detection of nonlinear synchrony. *Phys. Rev. E*, 65:046225.

Stam, C. J. (2005). Nonlinear dynamical analysis of eeg and meg: Review of an emerging field. *Clin. Neurophysiol.*, 116:2266–2301.

Stam, C. J. and van Dijk, B. W. (2002). Synchronization likelihood: an unbiased measure of generalized synchronization in multivariate data sets. *Physica D*, 163:236–251.

Stark, J., Broomhead, D. S., Davies, M. E., and Huke, J. (1997). Takens embedding theorems for forced and stochastic systems. *Nonlinear Analysis- Theory Methods and Applications*, 30(8):5303–5314.

Strong, S. P., Koberle, R., de Ruyter van Steveninck, R. R., and Bialek, W. (1998). Entropy and information in neural spike trains. *Phys. Rev. Lett.*, 80(1):197–200.

Takens, F. (1981). Detecting strange attractors in turbulence. In Rand, D. A. and Young, L., editors, *Dynamical Systems and Turbulence*, volume 898 of *Lecture Notes in Mathematics*, pages 366–381. Springer-Verlag, Berlin.

Tass, P. A., Fieseler, T., Dammers, J., Dolan, K., Morosan, P., Majtanik, M., Boers, F., Muren, A., Zilles, K., and Fink, G. R. (2003). Synchronization tomography: A method for three-dimensional localization of phase synchronized neuronal populations in the human brain using magnetoencephalography. *Phys. Rev. Lett.*, 90:088101.

Theiler, J. (1986). Spurious dimensions from correlation algorithms applied to limited time-series data. *Phys. Rev. A*, 34:2427–2432.

Theunissen, F. and Miller, J. P. (1995). Temporal encoding in nervous systems: a rigorous definition. *J. Comput. Neurosci.*, 2:149–62.

Theunissen, F. E., Sen, K., and Doupe, A. J. (2000). Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *J. Neurosci.*, 20:2315–2331.

Theunissen, F. E. and Shaevitz, S. S. (2006). Auditory processing of vocal sounds in birds. *Curr. Opin. Neurobiol.*, 16(4):400 – 407.

Thiel, M., Romano, M. C., Kurths, J., Rolfs, M., and Kliegl, R. (2006). Twin surrogates to test for complex synchronisation. *Europhys. Lett.*, 75(4):535.

Thomson, E. E. and Kristan, W. B. (2005). Quantifying stimulus discriminability: A comparison of information theory and ideal observer analysis. *Neural Comput.*, 17(4):741–778.

Treves, A. and Panzeri, S. (1995). The upward bias in measures of information derived from limited data samples. *Neural Comput.*, 7(2):399–407.

van Rossum, M. C. W. (2001). A novel spike distance. *Neural Comput.*, 13:751–763.

Vicente, R., Wibral, M., Lindner, M., and Pipa, G. (2010). Transfer entropy: a model-free measure of effective connectivity for the neurosciences. *J. Comput. Neurosci.*, doi:10.1007/s10827–010–0262–3 pages 1–23.

Victor, J. D. (2005). Spike train metrics. *Curr. Opin. Neurobiol.*, 15(5):585–592.

Victor, J. D. (2006). Approaches to information-theoretic analysis of neural activity. *Biological Theory*, 1(3):302–316.

Victor, J. D. and Purpura, K. P. (1996). Nature and precision of temporal coding in visual cortex: A metric-space analysis. *J. Neurophysiol.*, 76:1310–1326.

Victor, J. D. and Purpura, K. P. (1997). Metric-space analysis of spike trains: theory, algorithms and application. *Network: Comput. Neural Syst.*, 8:127–164.

Victor, J. D. and Purpura, K. P. (1998). Spatial phase and the temporal structure of the response to gratings in v1. *J. Neurophysiol.*, 80(2):554–571.

Wang, L., Narayan, R., Graña, G., Shamir, M., and Sen, K. (2007). Cortical discrimination of complex natural stimuli: Can single neurons match behavior? *J. Neurosci.*, 27(3):582–589.

Wiener, N. (1956). The theory of prediction. In *Modern Mathematics for Engineers*. McGraw-Hill, New York.

Wohlgemuth, S. and Ronacher, B. (2007). Auditory discrimination of amplitude modulations based on metric distances of spike trains. *J. Neurophysiol.*, 97(4):3082–3092.

Wright, B. D., Sen, K., Bialek, W., and Doupe, A. J. (2002). Spike timing and the coding of naturalistic sounds in a central auditory area of songbirds. In *Advances in Neural Information Processing, Cambridge, MA: MIT Press*.

Wright, S. (1921). Correlation and causation. *J. Agric. Res.*, 20:557–585.

Zhou, Z., Ding, M., Chen, Y., Wright, P., Lu, Z., and Liu, Y. (2009). Detecting directional influence in fmri connectivity analysis using pca based granger causality. *Brain Res.*, 1289:22–29.